Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 14, No. 3, 2026, 367-379.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Categorizing Rules from the Expurgation Point of View using Large Language Models

Hassan Deldar[1*] and Mohammad Mehdi Homayounpour[2]

1. Shahab Danesh University, Qom, Pardisan, Iran.
2. Department of Computer Engineering, Amirkabir University of Technology, Tehran, Iran.

Article Info

Article History:
Received 18 June 2025
Revised 26 February 2026
Accepted 13 May 2026

DOI:10.22044/jadm.2026.16400.2763

Keywords:
Law Expurgation, Textual Similarity, Large Language Models, Artificial Intelligence, Natural Language Processing.

**Corresponding author: hdeldar@shdu.ac.ir (H. Deldar).*

Abstract

In most countries, the legislative process has a long history, which leads to increasing diversity and multiplicity of laws and, consequently, the proliferation of a large volume of regulations, bylaws, guidelines, resolutions, and circulars. This has made it difficult to access laws that are valid in both time and place. The focus of this article is on the application of artificial intelligence in the domain of legal statutes to assist in identifying the need for amendments to laws or specific provisions. The general framework of the proposed process consists of two key components. First, the texts of legal clauses or articles are enriched through the generation of enriched data using large language models, which involves producing embedding vectors, thematic classification, and extracting the provisions of each law. Second, a retrieval-augmented text generation (RAG) system is developed with the aid of large language models to determine conflicts or the need for expurgation in the output, utilizing the enriched data, predefined prompts, and the Chain of Thought (CoT) technique. The proposed method was evaluated on two benchmark datasets. On the COLIEE 2025 dataset, our approach outperformed the 2024 winners in legal implication tasks, achieving an F1 score of 0.6521 with minimal prompting. The second evaluation used over 1,000 legal clauses covering abrogation and neutral rules, yielding an impressive F1 score exceeding 73.41%. The findings of the proposed methodology demonstrate that, even with limited expertise in the legal domain, it is possible to identify conflicts and the necessity for refining legal texts to an acceptable degree within a reasonable timeframe for legal experts, leveraging the capabilities of large language models.

1. Introduction

In Iran, legislation in its current form has a history of more than one hundred and twenty years. While this has resulted in substantial legislative experience, it has also led to increasing diversity and proliferation of laws, along with the creation of a significant volume of regulations, bylaws, guidelines, resolutions, circulars, judicial procedures, interpretative and advisory opinions, and legal scholars' theories. This has made it difficult to identify the governing law in a timely manner and across different contexts, as among the existing laws and regulations—which, according to the latest data indexed in the National System of Laws and Regulations of the Islamic Republic of Iran, exceed 12,000 laws, 111,000 regulations, and 22,000 rulings [1]—it remains unclear which provisions are valid and which have become obsolete [2].

The advances in artificial intelligence across various fields, particularly in natural language processing and text analysis, are undeniable. The primary focus of legal artificial intelligence is on

harnessing AI capabilities to assist with legal matters and subjects. Since most legal sources exist as texts and documents—including court records, contracts, and legal opinions—the majority of legal AI applications rely on natural language processing technologies [5]. The application of artificial intelligence in the compilation and expurgation processes enables machines to perform tasks intelligently and accurately, offering multiple capabilities such as thematic classification of laws and regulations, establishing interconnections between them, and facilitating expurgation diagnoses, including identifying repealed, obsolete, valid, and invalid provisions. The purpose of codification and expurgation of laws is to clarify and elucidate legal provisions. While codification and expurgation constitute highly valuable undertakings, they are simultaneously precise and challenging endeavors [4,7]. Artificial intelligence, particularly natural language processing, offers promising solutions to these challenges in legal expurgation and formulation.

By leveraging the capabilities of natural language models—including their precise adjustment and training during prompt execution with numerous examples—we have developed a comprehensive tool to assist legal experts in law expurgation.

The structure of the article is as follows: first, we review the related literature, followed by a description of the data and evaluation criteria. We then present the proposed methodology, the experimental setup, and the results obtained. Finally, we offer a discussion of the findings and concluding remarks. Before proceeding, the following glossary provides definitions of key terms frequently used in legal research and statutory interpretation:

Glossary:

****Rule expurgation:** refers to the systematic collection of rules and regulations organized thematically, with subsequent identification and removal of invalid or obsolete provisions, culminating in a validated thematic compilation. This process encompasses three key actions: collection, classification, and the refining of laws, which is called expurgation. Literally, expurgation entails both reviewing existing laws and making necessary amendments, corrections, and organizational improvements to address ambiguities. The process also identifies and eliminates abrogated and obsolete laws that may create contradictions or ambiguities, thereby preventing legal inconsistencies. Expurgation provides the authority to amend existing laws.**

****Abrogation:** constitutes the replacement of previous provisions with new ones, making coexistence impossible. This legislative action explicitly or implicitly removes a law's legal validity. Through the enactment of new legislation, the previous law loses its force and is superseded by the new provisions. In legal terminology, the newly enacted law is termed the "abrogating" or "repealed" law, while the superseded legislation is referred to as the "abrogated" or "obsolete" law.**

2. Related works

Legal scholars traditionally emphasize rule-based and symbolic approaches utilizing knowledge graphs and ontologies, while natural language processing researchers focus primarily on embedding techniques and data mining methods [10,5]. Initial research efforts concentrated on thematic and semantic classification of legal texts, employing either traditional text-based methods (including HAL, LSA, and LDA) or deep neural network architectures (such as CNN, LSTM, and BiLSTM) alongside transformer-based models like BERT. These approaches convert words and sentences into embedding vector representations, enabling measurement of semantic similarity between textual elements through cosine similarity metrics [9]. Such vector representations facilitate both the extraction of semantic categories from legal corpora and subsequent semantic analysis through named entity recognition and relationship extraction techniques, thereby identifying laws requiring expurgation [11,12,13].

Current methodologies for legal document understanding and retrieval predominantly fall into two categories. The first employs symbolic artificial intelligence to enhance the interpretability of legal analyses, while the second utilizes embedding-based approaches and deep neural networks for task execution. Symbolic methods benefit from explicit representation of events and relationships that improve decision interpretability, while also demonstrating potential for performance enhancement when combined with deep neural network architectures [5].

Legal documents typically contain an average of 1,260 tokens. Research has demonstrated the effectiveness of Lawformer, a Longformer-based pre-trained language model specifically adapted for processing lengthy Chinese legal documents. To address document length challenges, Lawformer employs a sliding attention mechanism

with limited scope rather than the conventional global attention approach that processes all tokens simultaneously [14].

Most publicly available language models are trained on general-domain corpora, including news articles, books, and Wikipedia content. For specialized domains like legal texts, fine-tuning becomes essential. Two critical considerations emerge in legal text fine-tuning: first, supervised training requires labeled legal data, which demands substantial resources for annotation; second, legal texts exhibit domain-specific terminology and syntactic structures distinct from general language. Transfer learning methods partially address the first challenge by reducing the required training data volume. The second challenge necessitates domain-specific fine-tuning for optimal performance [15].

Empirical studies demonstrate significant improvements when pre-trained BERT models undergo fine-tuning with extensive legal datasets. Such domain adaptation enhances both classification accuracy and training efficiency. Larger datasets prove particularly beneficial across all application domains, even without extensive labeling [16].

Further research employing LEGAL-BERT for multi-class legal classification tasks revealed that while fine-tuned BERT models achieve strong performance, they exhibit bias toward documents overrepresented in training data [17]. The specialized nature of legal texts, with their formal syntax and semantics rooted in comprehensive legal knowledge, sometimes qualifies them as a distinct sublanguage. Comparative analyses indicate that the original BERT architecture underperforms in certain legal applications without domain-specific adaptation. Notably, smaller pre-trained models achieve substantially better results when fine-tuned with legal domain occurrences [18].

Legal texts are notably lengthy and complex. The average word count in the COLIEE 2019 dataset is approximately 3,000 words per document. This dataset originates from the Legal Information Extraction/Implication competition, an annual event hosted by Northwestern University in Chicago to advance the automated processing of legal texts. The competition comprises two datasets: legal cases and legal statutes. For each dataset, two tasks are included: legal information retrieval and legal information inference.

Representation learning methods face significant challenges in encoding such lengthy documents into semantic space. On the one hand, defining inter-document relationships in the legal domain extends beyond the conventional notion of topical relevance. On the other hand, compiling a sufficiently large dataset for this task is highly challenging, if not unfeasible. The scarcity of suitable data further complicates the training of deep neural models. To mitigate these challenges, BERT-PLI has been proposed—a modified version of BERT that extracts semantic relationships at the paragraph level and subsequently infers document-level relationships by aggregating paragraph-level interactions [19].

Multilingual models are typically designed for languages that share morphological and syntactic similarities, such as those derived from Latin. Non-Latin languages, however, derive limited benefit from such shared representations. Convolutional models like BERT have gained prominence due to their strong performance. Nevertheless, these models predominantly focus on English, relegating other languages to resource-constrained multilingual alternatives. In one study, a pre-trained monolingual BERT for Persian, ParsBERT, was introduced, demonstrating superior performance compared to multilingual architectures and models [20].

Named entity recognition (NER) is a natural language processing task that involves identifying and extracting text spans associated with entities and classifying them into predefined semantic categories. These categories include persons, organizations, locations, dates, times, monetary values, as well as facilities, products, and events. In this study, a conditional random field (CRF)—a probabilistic model akin to a hidden Markov model—is employed to capture structural dependencies among labels in the data. This approach incorporates a fully connected layer following the BERT output to extract named entities. The proposed architecture was evaluated on two Persian datasets, [21] Peyma and [22] Arman, which are annotated with named entity labels. The results indicate strong performance in named entity identification [11].

Legal texts are typically lengthy, so a critical challenge in processing them lies in reducing their length while retaining essential information. This task is not accomplished through summarization but rather requires the extraction of all key phrases from the text. By identifying these key phrases—which comprehensively cover the text's topics—similar texts can be retrieved efficiently. Due to the inherent complexity of legal texts, the automatic extraction of key phrases is highly challenging. In one study, a two-stage method for key phrase extraction has been proposed, consisting of scoring and selection. For scoring, a convolutional neural

network (CNN) analyzes phrases along with their contextual neighborhoods to extract relevant features. The resulting outputs are then processed by a multilayer perceptron (MLP) to classify phrases as either key or non-key phrases. Finally, the selection stage identifies key phrases based on the highest assigned scores [23].
An alternative approach involves segmenting legal documents into smaller sections and categorizing them using Named Entity Recognition (NER). In [24], Conditional Random Field (CRF) models are employed for this purpose.

Jha et al. introduced the Contrastive Long Document Encoding (CoLDE) framework [25], a transformer-based model that utilizes distinctive positional embeddings and a multi-head piecewise attention layer combined with a supervised contrastive learning objective for similarity assessment. Empirical results demonstrate that the proposed CoLDE method surpasses existing state-of-the-art models in long-document matching tasks.
Zhang et al. applied a pre-trained BERT model to Chinese named entity extraction [26]. Documents are first partitioned into paragraphs, each of which is tokenized using the BERT tokenizer. After processing the tokens through the BERT model, the hidden layer outputs are extracted. A fully connected neural network layer followed by a softmax layer then predicts the entity labels. Despite minimal training data, the fine-tuned model achieves competitive performance.

In one study, it was demonstrated that focusing solely on specific regions of long documents suffices for classification tasks. This approach circumvents the 512-token input limitation of BERT models while maintaining satisfactory results during fine-tuning and deployment. The present paper proposes methods for identifying critical document segments, utilizing both statistical approaches and self-attention layers integrated with neural networks [12].
A review of [27] indicates that judicial outcomes in the Turkish legal system can be accurately predicted through deep learning methodologies, particularly LSTM and GRU networks.
One case study illustrates a process where competency questions are initially generated by a large language model (LLM). These questions undergo expert review and subsequent expurgation through the same LLM using the Retrieval-Augmented Generation (RAG) approach to produce answers. Subsequently, for knowledge graph construction, the processed competency questions along with their answers and an LLM-generated ontology are re-input into the LLM. Through specific prompts, the model extracts key entities, relations, and concepts from these responses, mapping them into an ontological structure to generate the knowledge graph [28].
A related study explores ontology generation using large language models across three specific tasks: term type determination, class discovery, and non-classification relationship extraction, employing task-specific prompt templates [29]. Recent advances in prompt engineering research demonstrate that techniques such as Chain-of-Thought (CoT) and Tree-of-Thought (ToT) reasoning yield superior results for complex tasks without requiring model fine-tuning [30]. These methods enable models to articulate their reasoning processes step-by-step or decompose problems into smaller components for individual analysis. If a particular reasoning path proves unproductive, it can be discarded in favor of more promising avenues, ultimately leading to effective task resolution [30,31,32].

3. Datasets

Two distinct datasets were utilized in this investigation. The primary dataset was acquired through web crawling of authoritative government sources, including parliamentary websites and official gazettes, supplemented by data extracted from legal repositories. This collection process was implemented using the Python programming language and the Elasticsearch repository and engine. The dataset's accuracy, validity, and comprehensiveness are inherently dependent on the reliability of these official sources. All programming scripts and collected data have been made publicly accessible in a GitHub repository to facilitate future research endeavors. Following data acquisition, legal texts were systematically segmented into discrete articles and clauses. This structural division was implemented to enhance processing efficiency and address inherent limitations of linguistic models in handling extensive textual content. Subsequent preprocessing operations standardized and sanitized the legal text sections using specialized digest and parser libraries combined with regular expressions. This normalization process included: (1) unification of shared Persian-Arabic characters (e.g., "ی" and "ک") into standardized forms, and (2) systematic removal or standardization of control characters, including alignment markers, extraneous spacing, and half-spaces.
This dataset, consisting of over 1,000 legal clauses, was annotated based on the presence of explicit or

implicit legal abrogation. The classification labels are defined as follows:
Label 0 (Neutral Rule): Clauses that simply state a regulation without implying the removal or invalidation of any prior rule.
Label 1 (Abrogation Present): Clauses that explicitly or implicitly nullify, supersede, or repeal a previously stated legal provision.
For instance, a sample entry demonstrating Label 1 might look like this:

Text1: ‘ماده ۵ - کلیه مؤسسات موضوع این قانون می توانند در جهت اجراء و تکمیل پروژه ها و برنامه های آموزشی و تحقیقاتی و عمرانی از هدایا و کمکهای مردمی استفاده نمایند و نحوه هزینه کردن اینگونه کمکهای مردمی نامه ای خواهد بود که به پیشنهاد هیأت امناء مؤسسات ذیربط به تابع آیین تصویب وزارتین مربوطه برسد.’

Text2: ‘ماده هشتم - تبصره - هدایایی که اشخاص یا مؤسسات بعنوان دانشگاه با آنها اداره میکنند تقدیم بدانشگاه خاص امر جهت آن امثال و وقف تبدیل و شود صرف کنندگان هدیه میل مطابق باید عایدات قبیل این است بوزارت ساله همه مخارج و عایدات صورت نیست جایز دیگر بمصرف آن است آزاد مذکور هدایای قبول و رد در دانشگاه شد خواهد تقدیم اوقاف.’

Annotation: Label 1 (Explicit Abrogation).
The secondary dataset originates from the COLIEE 2025 competition, a preeminent international event in legal informatics. This study specifically employed the legal case data component of the competition, concentrating on the task of legal information entailment.
This dataset consists of pairs of legal clauses extracted from benchmark collections such as the COLIEE 2025 competition. Each pair was annotated to indicate whether a legal entailment exists between them. The classification labels are defined as follows:
Label 0 (No Entailment): The hypothesis clause does not logically or legally follow from the premise clause.
Label 1 (Entailment Present): The hypothesis clause can be inferred from, or is entailed by, the premise clause under legal reasoning.
For illustration, a sample annotated entry is shown below:
Premise: “The case law holds that proof of intention is not required since the system is one of voluntary reporting and because strict liability attaches to those who fail to report”
Hypothesis: “That jurisprudence makes it abundantly clear that a traveler’s subjective intention when failing to report is irrelevant. It has been unequivocally held that such intention is not required since the system is one of voluntary reporting and because strict liability attaches to those who fail to report.”
Annotation: Label 1 (Entailment Present)

4. Evaluation Criteria

The performance evaluation of the proposed method incorporates four key metrics: accuracy, precision, recall, and F1 score. In the evaluation formulas, TP (true positive) represents correctly identified positive instances, while FN (false negative) denotes positive instances incorrectly classified as negative. TN (true negative) indicates correctly identified negative instances, and FP (false positive) refers to negative instances erroneously classified as positive.
Accuracy measures the proportion of correct predictions relative to total predictions, representing the overall correctness of the model's outputs. Precision quantifies the exactness of positive predictions, reflecting the model's ability to avoid false positives. Recall assesses the model's capacity to identify all relevant positive instances, measuring the completeness of positive predictions. The F1 score represents the harmonic mean of precision and recall, providing a balanced measure of the model's performance. For selecting optimal embedding vector extraction models, we employ the perplexity metric. Perplexity is defined as the mean negative log-likelihood of a probabilistic sequence of tokens $X = \{x0.x1.\ldots.xt\}$ generating a sentence. This metric evaluates a model's predictive capability by measuring how effectively it anticipates subsequent words or characters given the preceding context. Lower perplexity values indicate superior predictive performance, demonstrating the model's enhanced ability to forecast sequential elements.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\Pr ecision = \frac{TP}{TP + FP} \tag{2}$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = \frac{2 \times Percision \times \mathrm{Re}call}{Percision + \mathrm{Re}call} \tag{4}$$

$$PPL(X) = e^{-\frac{1}{t}\sum_{i}^{t}\log(p\theta(xi|x<i))} \tag{5}$$

5. Proposed Method

The purpose of the codification and expurgation of laws is to clarify and systematize existing legal provisions. Law expurgation entails collecting laws and regulations thematically, distinguishing between repealed and obsolete cases, and ultimately presenting a coherent, updated version of valid laws and regulations.

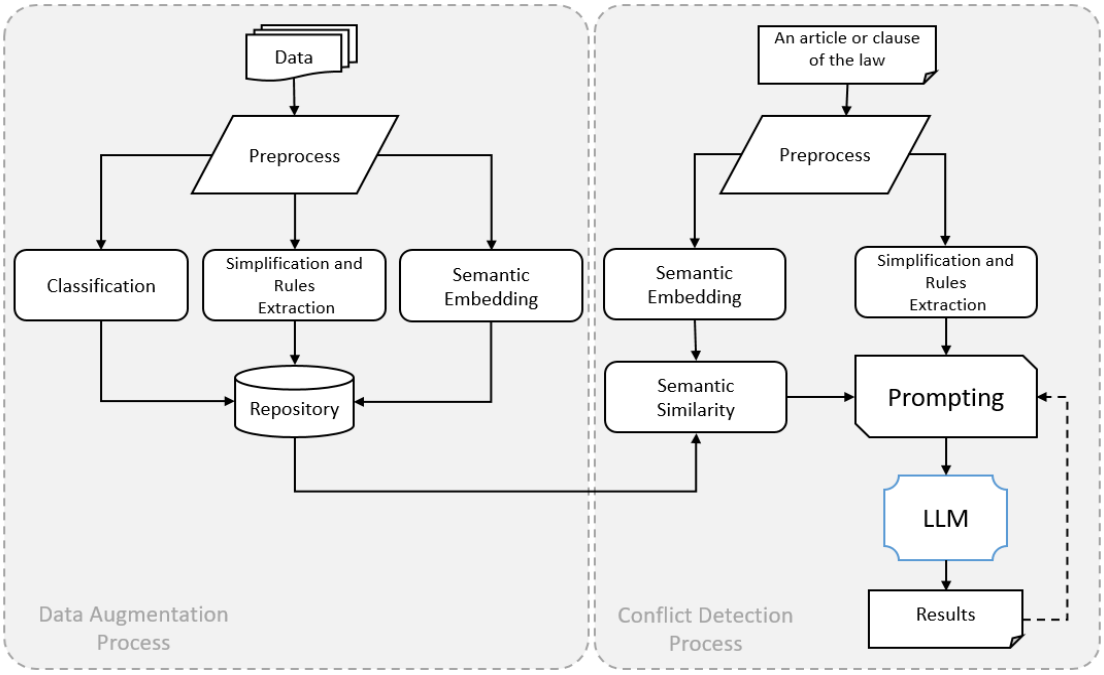

Figure 1. General framework of the proposed method.

This process grants relevant authorities the power to amend existing legislation. It should be noted that abrogating a law constitutes an action whereby the legislator explicitly or implicitly nullifies its legal validity. Codification and expurgation of laws is a highly useful yet precise and challenging task. The complexity arises from the abundance and extraordinary proliferation of approvals and regulations across most fields, often resulting in confusion regarding legal materials and topics. Determining the validity of laws, regulations, or specific articles—as well as decrees or regulations that sometimes attain significance equivalent to laws—poses considerable difficulty. Identifying which provisions have been superseded within a vast volume of decrees and regulations is particularly ambiguous, requiring extensive research and effort. This ambiguity is often implicit rather than explicit in legal texts, as portions of laws—or even parts of articles—may have been repealed, amended, expanded, or restricted over time, whether in short or long intervals. The overlap of laws further compounds these challenges. In some cases, an article from an unrelated law or decree may be altered, amended, or invalidated alongside the primary legislation in question. Additionally, judicial interpretations and opinions issued by the Supreme Court have broadened the scope of study, potentially increasing the complexity of the task.

This article proposes a method that, with minimal prior knowledge of the legal domain and leveraging the capabilities of large language models, can assist experts in identifying ambiguities, abrogations, and areas requiring expurgation in specific legal texts. This approach achieves an acceptable level of accuracy in significantly less time than conventional methods.

Figure 1 illustrates the general framework of the conflict detection process, which consists of two key components. The first component generates auxiliary data using language models, while the second component functions as a Retrieval-Augmented Generation (RAG) system. This RAG module receives the user's input text—such as a legal article or clause—and compares it with all existing data on the same or similar subjects. If a conflict or implicit/explicit abrogation is detected, the system presents it to the user in the output. A mechanism for re-evaluating the generated response is also incorporated. Identifying conflicts or abrogation between legal texts requires that the texts address a single subject, exhibit semantic or even verbal similarity, and contain provisions that cannot be logically reconciled. An example of such a conflict is provided in Table 1.

Table 1. Examples of repealed and obsolete laws.

عنوان: قانون نحوه انجام امور مالی و ها و مؤسسات آموزش معاملاتی دانشگاه عالی و تحقیقاتی تاریخ تصویب: ۱۳۶۹/۱۰/۱۸ مرجع تصویب: مجلس شورای اسلامی	عنوان: قانون اجازه تاسیس دانشگاه در طهران تاریخ تصویب: ۱۳۱۳/۰۳/۰۸ مرجع تصویب: مجلس شورای ملی
Title: Law on the Procedure for Conducting Financial and Transactional Affairs of Universities and Higher Education and Research Institutions Date of Approval: ۱۳۶۹/۱۰/۱۸ Authority of Approval: Islamic Consultative Assembly	Title: Law on the Establishment of a University in Tehran Date of Approval: ۱۳۱۳/۰۳/۰۸ Authority of Approval: National Assembly
ماده ۵ - کلیه مؤسسات موضوع این قانون می توانند در جهت اجراء و تکمیل پروژه ها و برنامه های آموزشی و تحقیقاتی و عمرانی از هدایا و کمکهای مردمی استفاده نمایند و نحوه هزینه کردن اینگونه کمکهای مردمی تابع آیین‌نامه ای خواهد بود که به پیشنهاد هیأت امناء مؤسسات ذیربط به تصویب وزارتین مربوطه برسد.	ماده هشتم - تبصره - هدایایی که اشخاص یا مؤسسات بعنوان وقف و امثال آن جهت امر خاص بدانشگاه تقدیم میکنند اداره آنها با دانشگاه است این قبیل عایدات باید مطابق میل هدیه کنندگان صرف شود و تبدیل آن بمصرف دیگر جایز نیست صورت عایدات و مخارج همه ساله بوزارت اوقاف تقدیم خواهد شد دانشگاه در رد و قبول هدایای مذکور آزاد است.
Article 5 - All institutions subject to this law may use public donations and grants to implement and complete educational, research, and development projects and programs, and the manner in which such public donations are spent will be subject to regulations that are proposed by the board of trustees of the relevant institutions and approved by the relevant ministries.	Article Eight - Note - Gifts that individuals or institutions present as endowments and the like for a specific purpose to the university are administered by the university. Such income must be spent according to the wishes of the donors and it is not permissible to convert it to other uses. The income and expenditure statement will be submitted to the Ministry of Endowments every year. The university is free to reject or accept the aforementioned gifts.

In this case, one law delegates the discretion over the allocation of gifts to universities to the donors' will, while another law assigns this authority to the board of trustees. (It should be noted that the selection of examples was informed by the majority opinion of available experts.)

In the initial step, all national legal texts were collected from official government websites and relevant authorities, as depicted in Figure 2.

These texts then underwent preprocessing via natural language processing methods and regular expressions to establish a unified and standardized corpus. These processed texts were then subjected to three parallel natural language modeling processes: (1) thematic classification, (2) simplification of complex legal texts, and (3) extraction of embedding vectors for each legal clause or article. The generated embedding vectors enable the identification of semantically similar articles or clauses across different laws. By combining thematic classification with semantic similarity analysis using embedding vectors, we significantly narrow the search scope for each clause or article. This method confines searches to materials that share both subject matter and high semantic similarity, substantially improving response speed and dramatically reducing the number of prompts required for natural language models.

Based on this foundation, linguistic models were developed to identify thematic classification and semantic similarity. The collection and preprocessing process yielded 170,000 legal sections, which were subsequently classified using linguistic models. For this purpose, all Persian models available in the Huggingface repository were evaluated based on their text tokenizers' ability to process Persian text at the word level. Through this evaluation, the HooshvareLab/bert-fa-base-uncased model was selected. The model was trained with a dataset generated by legal experts, consisting of 120,000 sentences classified into 50 legal subject categories. The training process was conducted on a GPU 3090 with 12 epochs, a learning rate of 5e-2, and 12 batches. The Pearson correlation coefficient for the test dataset, representing 20% of the total dataset, ultimately reached 0.86. Utilizing this model, all legal sections were successfully classified into 50 categories, enabling more efficient search processes in subsequent stages.

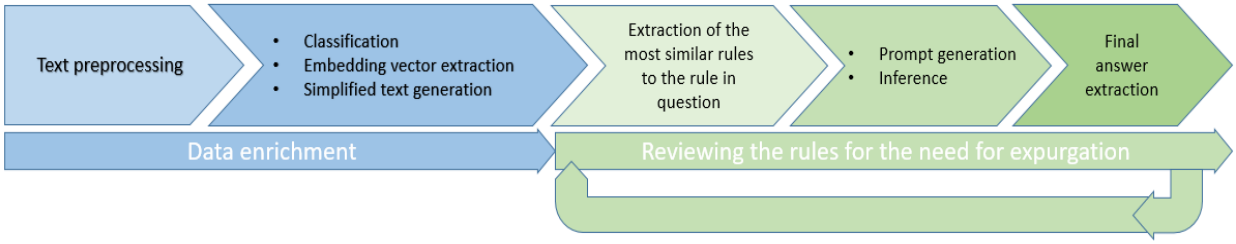

Figure 2. Steps to implement the proposed method.

Additionally, establishing semantic similarities between sections was required. For this process, multiple language models were fine-tuned according to Table 2 using all segmented rule data through the Masked Language Modeling (MLM) method. Notably, 2,500 legal sentences were extracted from the rules' content to serve as benchmark data for model comparison. These sentences were randomly selected from the 50 subject categories identified in the previous stage, ensuring 50 sentences were included from each category. By comparing the perplexity metrics of the fine-tuned language models against the rule texts, the HooshvareLab/bert-fa-base-uncased model was again selected for implementation in the similarity detection process.

Table 2. Comparing perplexity values of fine-tuned models with rules data.

Model Name	Perplexity
roberta-fa-zwnj-base	**14.78**
distilbert-base-multilingual-cased	**39.55**
HooshvareLab-bert-fa-base-uncased	**7.29**

In the next step, the embedding vectors of all legal section texts were generated and stored using the recently fine-tuned model. Whenever similarity between a given text and one of the legal sections needs to be determined, it suffices to generate the embedding vector of the target text and compute the nearest vector distances between this vector and those of the legal sections using methods such as cosine similarity. If the text in question is itself a legal section, the cosine similarity computation can be restricted to only those laws pertaining to the subject of the section to improve processing speed. Another step involved converting the legal sections into sentences, which were transformed into if-then structures through prompt engineering and the use of large language models such as Llama-3.1 70B. This step reduces complexity and facilitates processing in subsequent stages. For instance, using a prompt, the content of the legal sections in Table 1 was converted into rules with the assistance of the Llama-3.1 70B model, as illustrated in Table 3.

An example of a system prompt to extract rules from legal section texts is provided to the Llama-3.1 70B model:

“You receive a legal text and convert it into precise logical sentences in the if-then format, and you should not produce any explanations or additional text.”

After extensive testing, we observed that in some cases, using an English prompt yields better results with models such as Llama-3.1 70B.

Table 3. An example of simplifying and extracting rules from sections of laws.

عنوان: قانون اجازه تاسیس دانشگاه در طهران ماده ۸ - تبصره	عنوان: قانون نحوه انجام امور مالی و معاملاتی دانشگاهها و مؤسسات آموزش عالی و تحقیقاتی ماده ۵
Title: Law on the Establishment of a University in Tehran Article 8 - Note	Title: Law on the Procedure for Conducting Financial and Transactional Affairs of Universities and Higher Education and Research Institutions Article 5
اگر هدایایی به دانشگاه اهدا شود، آنگاه دانشگاه اداره آنها را بر عهده دارد. اگر هدایایی برای امر خاصی اهدا شود، آنگاه آن هدایا باید صرف همان امر خاص شود. اگر هدایایی به دانشگاه اهدا شود، آنگاه دانشگاه موظف است هر ساله گزارش عایدات و مخارج را به وزارت اوقاف ارائه کند. اگر هدایایی به دانشگاه اهدا شود، آنگاه دانشگاه در رد و قبول آن هدایا آزاد است.	اگر مؤسسه‌ای موضوع این قانون باشد، آنگاه می‌تواند از هدایا و کمک‌های مردمی برای پروژه‌ها، برنامه‌های آموزشی، تحقیقاتی و عمرانی استفاده کند. اگر مؤسسه‌ای از هدایا و کمک‌های مردمی استفاده کند، آنگاه نحوه هزینه کردن آن تابع آیین‌نامه‌ای خواهد بود که به پیشنهاد هیأت امناء مؤسسه به تصویب وزارتین مربوطه برسد.
If gifts are donated to the university, then the university is responsible for their administration. If gifts are donated for a specific purpose, then those gifts must be spent on that specific purpose. If gifts are donated to the university, then the university is required to submit an annual report of income and expenditure to the Ministry of Endowments. If gifts are donated to the university, then the university is free to reject or accept those gifts.	If an institution is subject to this law, then it can use public gifts and donations for educational, research and development projects and programs. If an institution uses public gifts and donations, then the way they are spent will be subject to a regulation that is proposed by the board of trustees of the institution and approved by the relevant ministries.

Consequently, the final tasks were executed using the following English prompt:

PROMPT:

```
SYS_PROMPT = """You receive a Persian legal text and convert it into precise logical rules in the if-then format in Persian, and you do not need to produce any additional explanations or text.
Each rule should be in one sentence."
"""
PROMPT = f"Persian legal text: {SENTENCE}."
```

Through these processes, legal texts become enriched, with the generated data stored in a repository featuring rapid search capabilities.

In another step, a RAG was created with the help of large natural language models such as Llama-3.1 70B. In this step, input texts such as legal sections are processed using the rules extraction method to convert them into logical sentences, while simultaneously identifying similar legal sections through similarity vector analysis. For each section, the extracted rules—combined with the appropriate prompt—are evaluated using the Chain-of-Thought (CoT) technique to detect potential conflicts or abrogates between the input

text's rules and those of each similar section, generating preliminary results for legal expert assessment. This process involved transforming legal sections into more comprehensible and fluent if–then statements, while simultaneously aligning semantically similar legal articles with each corresponding section. The contextual depth of each clause was enhanced through this enrichment step. Following this, we employed precise prompts and a step-by-step reasoning procedure to guide the model in identifying any contradictions between the statement of a given legal section and its semantically related articles. The English prompt used is as follows:

PROMPT:

```
SYS_PROMPT = """
You receive two Persian legal rule texts and analyze them carefully, explain your answer step by step, to see whether these two rules logically conflict or abrogate with each other.
Finally, state the final conclusion for the presence or absence of conflict or abrogate with the words "yes" or "no".
"""
PROMPT = f"Rule 1: {SENTENCE1}. Rule 2: {SENTENCE2}."
```

Table 4. Examples of obsolete and obsolete law sections for use as few-shot.

Abrogating text	Obsolete text
صدور شناسنامه و سایر خدمات سجلی در مقابل دریافت وجه انجام خواهد شدکه نحوه اخذ وجوه و میزان و نوع این خدمات بموجب آئین**نامه اجرائی این قانون** خواهد بود.	سازمان ثبت احوال کشور مکلف است از اول سال ۱۳۶۳ در قبال ارائه خدمات سجلی وجوهی **به شرح زیر دریافت و به حساب درآمد عمومی واریز** کند.
Issuance of birth certificates and other registration services will be carried out in exchange for payment. The method of receiving the funds, the amount, and the type of these services will be in accordance with the executive regulations of this law.	The National Civil Registration Organization is required to receive and deposit the following amounts into the general revenue account for providing registration services, starting from the beginning of 1984.
مرکز آمار ایران **وابسته بسازمان برنامه و** باشد و زیر نظر وزیر مشاور و **بودجه** می شود رئیس سازمان برنامه و بودجه اداره می و رئیس آن سمت معاونت وزیر مشاور و رئیس سازمان برنامه و بودجه را دارد.	السابق تحت نظر فی‌مرکز آمار ایران کما سازمان انجام وظیفه خواهد کرد. دولت مکلف است ظرف شش ماه از تاریخ تصویب این قانون تاسیس **مرکز آمار ایران** را تقدیم مجلسین نماید.
The Statistical Center of Iran is affiliated with the Planning and Budget Organization and is managed under the supervision of the Minister of State and the Head of the Planning and Budget Organization, and its head holds the position of Deputy Minister of State and Head of the Planning and Budget Organization.	The Statistical Center of Iran will continue to operate under the supervision of the organization as before. The government is required to submit the establishment of the Statistical Center of Iran to the two chambers of parliament within six months from the date of approval of this law.
ماده ۶۹۰ - هر کس به وسیله صحنه‌سازی از قبیل پی‌کنی، دیوارکشی، تغییر حد فاصل، امحای مرز، کرت‌بندی، نهرکشی، حفر چاه، غرس اشجار وزارعت و امثال آن به تهیه آثار تصرف در اراضی مزروعی اعم از کشت شده یا در آیش زراعی، جنگلها و مراتع ملی شده، کوهستانها، باغها، قلمستانها،منابع آب، چشمه‌سارها، انهار طبیعی و پارکهای ملی، تأسیسات کشاورزی و دامداری و دامپروری و کشت و صنعت و اراضی موات و بایر و سایراراضی و املاک متعلق به دولت یا شرکتهای وابسته به دولت یا شهرداریها یا اوقاف و همچنین اراضی و املاک و موقوفات و محبوسات و اثلاث باقیه که برای مصارف عام‌المنفعه اختصاص یافته یا اشخاص حقیقی یا حقوقی به منظور تصرف یا ذیحق معرفی کردن خود یا دیگری، مبادرت نماید یا بدون اجازه سازمان حفاظت محیط زیست یا مراجع ذیصلاح دیگر مبادرت به عملیاتی نماید که موجب تخریب محیط زیست و منابع طبیعی گردد یااقدام به هر گونه تجاوز و تصرف عدوانی یا ایجاد مزاحمت یا ممانعت از حق در موارد مذکور نماید به **مجازات یک ماه تا یک سال حبس محکوم‌می‌شود**.	ماده ۵۵ به شرح زیر اصلاح و یک تبصره به آن اضافه می‌شود: ماده ۵۵ - هر کس به قصد تصرف به منابع ملی مذکور در ماده ۱ قانون ملی شدن جنگل‌های‌کشور تجاوز کند به **یک سال تا سه سال حبس تأدیبی‌محکوم خواهد شد**. تبصره - وزارت منابع طبیعی مکلف است به وسیله گارد جنگل و مأموران خود به محض‌اطلاع رفع تجاوز کند و مراتب را برای رسیدگی به موضوع وتعقیب کیفری کتباً به‌دادسرای محل اعلام دارد. اعیانی که در عرصه مورد تجاوز احداث شود به حکم دادگاه به‌نفع دولت ضبط خواهد شد.
Article 690 - Anyone who, by means of staging, such as erecting a fence, changing the boundary, erasing the border, dividing the land, digging a canal, digging a well, planting trees and crops, and the like, attempts to create signs of occupation in cultivated lands, whether cultivated or fallow, nationalized forests and pastures, mountains, gardens, pens, water sources, springs, natural rivers and national parks, agricultural and livestock facilities, livestock breeding and agro-industrial facilities, fallow and barren lands, and other lands and properties belonging to the government or companies affiliated with the government or municipalities or endowments, as well as lands, properties, endowments, enclosures and remaining thirds that are allocated for public benefit or to natural or legal persons for the purpose of occupying or presenting themselves or others as the rightful owners, or who attempts an operation without the permission of the Environmental Protection Organization or other competent authorities that causes environmental destruction and natural resources or commits any aggression or aggressive seizure or creates a nuisance or prevents the right in the aforementioned cases, shall be sentenced to one month to one year in prison.	Article 55 is amended as follows and a note is added to it: Article 55 - Anyone who trespasses with the intention of seizing the national resources mentioned in Article 1 of the Law on Nationalization of the National Forests of the Country shall be sentenced to one to three years of disciplinary imprisonment. Note - The Ministry of Natural Resources is obliged to remove the encroachment through the forest guard and its officers upon being informed and shall notify the local court in writing for consideration of the matter and criminal prosecution. Any property built in the encroached area shall be confiscated by court order for the benefit of the state.

For example, rules from the examples listed in Table 3, such as "If gifts are donated for a specific cause, then those gifts must be spent on that specific cause" and "If an institution uses public gifts and donations, then the way they are spent will be subject to a regulation that is proposed by the

board of trustees of the institution and approved by the relevant ministries," were provided to the model along with the prompt: "Is there a conflict between the given legal provisions or not? Explain the steps step by step and give the final answer. If there is a conflict, just say 'yes' and if there is no conflict, say 'no' and do not produce any other answer." The model correctly identified the conflict. In addition to these sample prompts, multiple examples (such as those in Table 4) related to each task were created for each prompt and provided to the model alongside the prompts to guide its responses as few-shot learning.

The generated answers are ultimately presented to the model for evaluation, accompanied by both the original prompts and the full texts of the relevant sections. The model is then prompted to assess whether each generated answer (regarding the presence or absence of conflict/abrogation) constitutes an acceptable determination, with the final judgment being delivered to the user. Studies show that better results are achieved if a larger model other than the large language model used in the previous stages is used in the judgment section.

6. Experiments

Two tests were conducted to examine and validate the proposed frame-work.

6.1. Experiment 1 (Entailment Identification in Legal Cases Using Similarity and LLMs):

In the first experiment, the COLIEE 2025 competition dataset was utilized for the competition's second task, which involved identifying entailments within legal case texts. This dataset comprises 4,297 legal cases, with an average of approximately 34 rules provided per case for entailment analysis. For this test, in Stage 1, the framework measured all similar court rules, employing three language models for this process. As shown in Table 5, the investigation revealed that most of the correct cases of entailment are in the four similar previous relevant cases. Specifically, the findings indicate that using the BGE-M3 model to extract similarity vectors between a new case and relevant previous cases yielded an 83% probability that the correct required entailment cases were contained within the four most similar cases. In the enrichment stage, for each court case text, the four vector-represented legal provisions with the highest semantic similarity were retrieved. Rather than conducting exhaustive pairwise comparisons between every legal article and the court text, the reasoning pipeline was limited to these four most relevant provisions. This design choice significantly optimized computational efficiency by reducing the number of inference operations required for large language model reasoning.

Table 5. Tested models for extracting embedding vectors to find the most similar entailment previous relevant cases required for each a new case.

Model Name	Recall
COLIEE_2025_FineTuend_Bert	0.696
Jina-embedding-v3	0.739
BGE-M3	**0.829**

Following the proposed framework, in the second stage, large natural language models were employed with appropriate prompts to identify the correct relevant previous cases for entailment in a new case. During this process, numerous models were evaluated using different prompts, with results presented in Table 6. The mixtral-small-28b_fewshot model, utilizing few-shot examples in its command structure, demonstrated optimal performance. As evidenced in Table 7, this model's performance (0.6512 for the second task)[33] surpassed the best result of the 2024 competition. In other words, our model achieved a higher F1-score of 0.6521 compared to the competition result of 0.6512. It also outperformed the other models in terms of precision and accuracy, indicating greater overall reliability. Regarding recall, the LLaMA-70B model obtained the best score, reflecting a higher proportion of correctly retrieved positive instances.

Table 6. Benchmark results for the models tested for the second task on the COLIEE 2025 dataset.

Models	F1	Precision	Recall	Accuracy
deepSeek_lama8	0.4431	0.3109	0.7708	0.5884
lama8	0.4590	0.3783	0.5833	0.7079
gpt-4	0.5333	0.4137	0.7500	0.7212
lama70	0.5734	0.4315	0.8541	0.7300
gpt-4o	0.5777	0.4482	0.8125	0.7477
gpt-4_fewshot	0.5785	0.4794	0.7291	0.7743
mistral-2407	0.6000	0.5769	0.6250	0.8230
gpt-4o_fewshot	0.6115	0.5068	0.7708	0.7920
mixtral-small-28b	0.635	0.5762	0.7083	0.8274
mixtral-small-28b_fewshot	**0.6521**	0.6818	0.625	0.8584

Table 7. Best COLIEE 2024 Competition Results for the Second Task [33].

Team name	F1	Precision	Recall
MIG	0.4701	0.5673	0.4014
OVGU	0.5962	0.5636	0.6327
NOWJ	0.6117	0.6181	0.6054
JNLP	0.6320	0.6967	0.7500
CAPTAIN	0.6360	0.7381	0.5782
AMHR	**0.6512**	0.6364	0.6667

The optimal prompt template yielding superior results was structured as follows:

PROMPT:
Given a query (which is delimited with triple backticks) and the related articles (which is also delimited with triple backticks). Is the query entailed by the related articles? To answer, please use the following format: Step-by-step reasoning: <your step-by-step reasoning> Answer: <a clear "Yes" or "No" response> Query: ```{query}``` Related articles: ```{related_articles}```

6.2. Experiment 2 (Abrogation Detection in Legal Articles Using Similarity Analysis and LLMs):

In the second experiment, the initial phase employed a dataset comprising 1,045 pairs of legal articles collected from the Majles and Datic websites, randomly selected from 50 distinct legal categories. Among these, 530 article pairs contained explicit and implicit abrogations, while 515 pairs were randomly selected based on the highest vector similarity. For similarity measurement, three transformer models were utilized: a fine-tuned BERT model adapted to legal content, BGE_M3, and Jina_embedding_v3. The fine-tuned BERT and BGE_M3 models demonstrated comparable and statistically acceptable performance. In the subsequent enrichment step, in addition to collecting vector-represented legal sections similar to each target section, the texts were converted into an explicit if–then logical structure. This transformation was intended to clearly articulate the scope, conditions, and consequences contained in each legal section. Furthermore, pronoun references and entity names—particularly corporate names—were standardized across all laws to facilitate more consistent and comparable analysis between different legal texts.

In the subsequent phase, the dataset was evaluated using multiple prompt variations and numerous few-shots across different large language models to identify laws containing both implicit and explicit abrogations. As presented in Table 8, the command-r model achieved optimal performance with an F1-score of 0.7341.

Table 8. Results of models tested with different criteria for detecting explicit and implicit abrogation.

Models	F1	Precision	Recall	Accuracy
gpt-4	0.3561	0.8239	0.2271	0.5952
qwen-2.5	0.5893	0.9404	0.4291	0.7052
gpt-4o	0.6298	0.9315	0.4757	0.7244
mistral-nemo	0.6356	0.8580	0.5048	0.7148
deepseek-v3	0.6700	**0.9670**	0.5126	0.7511
llama-3.1-8b	0.6996	0.7299	0.6718	0.7157
llama-3.1-8b_fewshot	0.7003	0.9208	0.5650	0.7617
command-r_fewshot	0.7193	0.6628	**0.7864**	0.6976
command-r	**0.7341**	0.8036	0.6757	**0.7588**

The experimental results demonstrate three key findings: First, the deepseek-v3 model achieved optimal performance in terms of precision. Second, the command-r_fewshot model yielded superior results for the recovery criterion. Finally, the highest accuracy and F1-score metric values were attained by the command-r model.

7. Discussion and Conclusion

The results demonstrate that the proposed method, requiring only limited legal domain knowledge while leveraging the capabilities of large language models, can effectively identify ambiguities, omissions, and expurgation needs in legal texts. Since legal texts are very difficult and complex for individuals who are not experts in this field, understanding the conditions and rules applicable to each law is often intricate and sometimes intertwined. This approach achieves acceptable accuracy levels significantly faster than traditional methods employing ontologies or knowledge graphs. It requires an initial data enrichment phase that involves compiling semantically related legal texts, simplifying their content into logical patterns, and standardizing anaphora references and entity names. Semantic and thematic classification increases the speed of locating laws with similar concepts, while enrichment facilitates easier understanding of complex legal concepts within a logical structure. Ultimately, the examination of contradictory provisions is conducted in a cycle using a RAG-based approach, applying the CoT technique within each semantically similar category and utilizing the enriched legal texts to identify abrogations and conflicting rules. The effectiveness of the proposed method is validated through two evaluation experiments. In the first experiment, conducted on the COLIEE 2025 dataset for the legal entailment task, our method achieved an F1 score of 0.6521, outperforming the competition's baseline score of 0.6512. In the second experiment, using a dataset of real statutory provisions, the highest performance was obtained with the CommandR model, yielding an F1 score of 0.7341. A critical challenge in prompt-based large language model approaches is their pronounced sensitivity to prompt phrasing and the inclusion of few-shot examples. Minor prompt modifications frequently produce substantially divergent outputs. The proposed method addresses this limitation through prompt enrichment and the incorporation of limited few-shot training during prompt execution. The enrichment section requires further development, including the addition of the degree of connection between sentences and the type of connection to the

prompt. Additionally, various simplification processes can be explored for future work. By detecting and adding connections between extracted rules to the prompt, enrichment will move closer to a rules knowledge graph, potentially yielding better results than the current enrichment. The proposed method for judging the final answer involves returning the results and the prompt to the model only once for re-examination. Future work could explore increasing the number of evaluations and employing larger models. If the proposed method is developed with the assistance of approaches such as Tree of Thought (ToT) and Graph of Thought (GoT), more accurate results are likely to be obtained.

References

[1] M. Alipour and A. Bahadori Jahromi, "Legal Feasibility Study of Crowdsourcing the Expurgation of Laws and Regulations, " In Persian p. https://www.sid.ir/paper/412080, 1401, [Online]. Available: https://www.sid.ir/paper/412080/fa.

[2] A. S. H. M. Mr. Mohammad Amin Keykha Farzaneh, "A Study of the Approach to the Expurgation of Laws and Regulations in the Iranian Legal System, " In Persian p . https://rc.majlis.ir/fa/report/show/1630975, 2010.

[3] M. M. Shora, "The Codification and Expurgation of Laws and Regulations of the Country, " In Persian p. https://rc.majlis.ir/fa/law/show/782398, 2010.

[4] S. Soltani, "Reasons for the Failure of the Codification and Expurgation of Laws in Iran, " In Persian 2014, [Online]. Available: https://civilica.com/doc/706648.

[5] H. Zhong, C. Xiao, C. Tu, T. Zhang, Z. Liu, and M. Sun, "How does NLP benefit legal system: A summary of legal artificial intelligence, " *Proc. Annu. Meet. Assoc. Comput. Linguist.*, pp. 5218–5230, Apr. 2020, doi: 10.18653/v1/2020.acl-main.466.

[6] D. J. Langroodi, Essay on Legal Terminology. Ganj Danesh, In Persian, 1401.

[7] W. Hu *et al.*, "BERT_LF: A Similar Case Retrieval Method Based on Legal Facts, " *Wirel. Commun. Mob. Comput.*, vol. 2022, pp. 1–9, Apr. 2022, doi: 10.1155/2022/2511147.

[8] F. A. Naseri, A Collection of Criminal Laws and Regulations. Amir Kabir, In Persian, 1350.

[9] D. Chandrasekaran and V. Mago, "Evolution of Semantic Similarity—A Survey, " *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–37, Mar. 2022, doi: 10.1145/3440755.

[10] A. Das and P. Rad, "Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey, " Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.11371.

[11] E. Taher, S. A. Hoseini, and M. Shamsfard, "Beheshti-NER: Persian Named Entity Recognition Using BERT, " Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.08875.

[12] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, and I. Androutsopoulos, "Large-Scale Multi-Label Text Classification on EU Legislation, " *arXiv Prepr. arXiv1906.02192*, Jun. 2019, [Online]. Available: http://arxiv.org/abs/1906.02192.

[13] F. X. B. da Silva *et al.*, "Named Entity Recognition Approaches Applied to Legal Document Segmentation, " in *Anais do X Symposium on Knowledge Discovery, Mining and Learning (KDMiLe 2022)*, Sociedade Brasileira de Computação - SBC, Nov. 2022, pp. 210–217. doi: 10.5753/kdmile.2022.227949.

[14] C. Xiao, X. Hu, Z. Liu, C. Tu, and M. Sun, "Lawformer: A pre-trained language model for Chinese legal long documents, " *AI Open*, vol. 2, pp. 79–84, 2021, doi: 10.1016/j.aiopen.2021.06.003.

[15] S. Shaghaghian, Luna, Feng, B. Jafarpour, and N. Pogrebnyakov, "Customizing Contextualized Language Models forLegal Document Reviews, " Feb. 2021, [Online]. Available: http://arxiv.org/abs/2102.05757.

[16] E. Elwany, D. Moore, and G. Oberoi, "BERT Goes to Law School: Quantifying the Competitive Advantage of Access to Large Legal Corpora in Contract Understanding, " Nov. 2019, [Online]. Available: http://arxiv.org/abs/1911.00473.

[17] R. Z. Mahari, "AutoLAW: Augmented Legal Reasoning through Legal Precedent Prediction, " *arXiv Prepr. arXiv2106.16034*, Jun. 2021, doi: 10.48550/arXiv.2106.16034.

[18] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, "EGAL-BERT: The Muppets straight out of Law School, "" *arXiv Prepr. arXiv2010.02559*, Oct. 2020, [Online]. Available: http://arxiv.org/abs/2010.02559.

[19] Y. Shao *et al.*, "BERT-PLI: Modeling Paragraph-Level Interactions for Legal Case Retrieval, " in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, California: International Joint Conferences on Artificial Intelligence Organization, Jul. 2020, pp. 3501–3507. doi: 10.24963/ijcai.2020/484.

[20] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, "ParsBERT: Transformer-based Model for Persian Language Understanding, " *Neural Process. Lett.*, vol. 53, no. 6, pp. 3831–3847, Dec. 2021, doi: 10.1007/s11063-021-10528-4.

[21] M. S. Shahshahani, M. Mohseni, A. Shakery, and H. Faili, "PEYMA: A Tagged Corpus for Persian Named Entities, " Jan. 2018, [Online]. Available: http://arxiv.org/abs/1801.09936.

[22] H. Poostchi, E. Zare Borzeshi, and M. Piccardi, "BiLSTM-CRF for Persian Named-Entity Recognition ArmanPersoNERCorpus: the First Entity-Annotated

Persian Dataset, " *Eur. Lang. Resour. Assoc.*, 2018, [Online]. Available: https://aclanthology.org/L18-1701.

[23] V. Tran, M. Le Nguyen, and K. Satoh, "Automatic Catchphrase Extraction from Legal Case Documents via Scoring using Deep Neural Networks, " Sep. 2018, [Online]. Available: http://arxiv.org/abs/1809.05219.

[24] L. A. B. M. Gabriel M. C. Guimarães, Felipe X. B. da Silva, "Legal Document Segmentation and Labeling Through Named Entity Recognition Approaches, " *J. Inf. Data Manag.*, 2024, doi: 10.5753/jidm.2024.3368.

[25] A. Jha, V. Rakesh, J. Chandrashekar, A. Samavedhi, and C. K. Reddy, "Supervised Contrastive Learning for Interpretable Long-Form Document Matching, " *ACM Trans. Knowl. Discov. Data*, vol. 17, no. 2, pp. 1–17, Apr. 2023, doi: 10.1145/3542822.

[26] R. Zhang *et al.*, "Rapid Adaptation of BERT for Information Extraction on Domain-Specific Business Documents, " Feb. 2020, [Online]. Available: http://arxiv.org/abs/2002.01861.

[27] E. Mumcuoğlu, C. E. Öztürk, H. M. Ozaktas, and A. Koç, "Natural language processing in law: Prediction of outcomes in the higher courts of Turkey, " *Inf. Process. Manag.*, vol. 58, no. 5, p. 102684, Sep. 2021, doi: 10.1016/j.ipm.2021.102684.

[28] V. K. Kommineni, B. König-Ries, and S. Samuel, "From human experts to machines: An LLM supported approach to ontology and knowledge graph construction, " Mar. 2024, [Online]. Available: http://arxiv.org/abs/2403.08345.

[29] H. B. Giglou, J. D'Souza, and S. Auer, "LLMs4OL: Large Language Models for Ontology Learning, " Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.16648.

[30] J. Wei *et al.*, "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, " Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.11903.

[31] S. Yao *et al.*, "Tree of Thoughts: Deliberate Problem Solving with Large Language Models, " May 2023, [Online]. Available: http://arxiv.org/abs/2305.10601.

[32] J.Long, "Large Language Model Guided Tree-of-Thought" May 2023, [Online]. Available: http://arxiv.org/abs/2305.08291.

[33] R.Goebel, Y. Kano, M.-Y. Kim, J. Rabelo, K. Satoh, ad M. Yoshioka, "Overview of Benchmark Datasets and Methods for the Legal Information Extraction/Entailment Competition (COLIEE) 2024, " 2024, pp. 109–124. doi: 10.1007/978-981-97-3076-6_8.

دسته‌بندی قوانین از دیدگاه تنقیح با استفاده از مدل‌های زبانی بزرگ

حسن دلدار[1,*] و محمدمهدی همایون‌پور[2]

[1] دانشگاه شهاب دانش، قم، ایران.

[2] دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران.

ارسال ۲۰۲۵/۰۶/۱۸؛ بازنگری ۲۰۲۶/۰۲/۲۶؛ پذیرش ۲۰۲۶/۰۵/۱۳

چکیده:

در بیشتر کشورها، سابقه طولانی قانون‌گذاری موجب افزایش تنوع و تعدد قوانین و در نتیجه گسترش حجم عظیمی از مقررات، آیین‌نامه‌ها، دستورالعمل‌ها، مصوبات و بخشنامه‌ها شده است. این امر دسترسی به قوانین معتبر و لازم‌الاجرا را از نظر زمانی و مکانی دشوار ساخته است. این مقاله به‌کارگیری هوش مصنوعی را برای شناسایی نیاز به اصلاح قوانین یا برخی از مواد و احکام آن‌ها بررسی می‌کند. چارچوب پیشنهادی شامل دو بخش است. در بخش نخست، متون بندهای قانونی با استفاده از مدل‌های زبانی بزرگ غنی‌سازی می‌شوند که شامل تولید بردارهای تعبیه، دسته‌بندی موضوعی و استخراج احکام هر قانون است. در بخش دوم، سامانه تولید متن مبتنی بر بازیابی (RAG) با بهره‌گیری از مدل‌های زبانی بزرگ، داده‌های غنی‌شده، دستورهای از پیش تعریف‌شده (Prompts) و تکنیک زنجیره تفکر(CoT)، برای شناسایی تعارض‌ها یا نیاز به تنقیح قوانین توسعه می‌یابد. روش پیشنهادی بر روی دو مجموعه‌داده معیار ارزیابی شد. در مجموعه‌داده رقابت COLIEE 2025، این روش در مسئله استلزام حقوقی با استفاده از تعداد بسیار کمی دستور، عملکردی بهتر از برندگان رقابت ۲۰۲۴ داشته و به امتیاز F1 برابر با ۰.۶۵۲۱ دست یافت. همچنین، در ارزیابی بیش از ۱۰۰۰ بند قانونی شامل مقررات ناسخ و منسوخ و قواعد خنثی، امتیاز F1 بیش از ۷۳.۴۱ درصد حاصل شد. نتایج نشان می‌دهد این روش حتی با دانش تخصصی محدود حقوقی، قادر به شناسایی تعارض‌ها و ضرورت اصلاح و تنقیح متون قانونی با دقت قابل قبول و در زمان مناسب است.

کلمات کلیدی: تنقیح قوانین، شباهت متنی، مدل‌های زبانی بزرگ، هوش مصنوعی، پردازش زبان طبیعی.