Записки научных
семинаров ПОМИ
Том 540, 2024 г.

A. Sabieva, A. Zhamankhan, N. Zhetessov, A. Kubayeva, I. Akhmetov, A. Pak, D. Akhmetova, A. Zhaxylykova, A. Yelenov

SURVEY ON THE LEGAL QUESTION ANSWERING PROBLEM

Abstract. Recent advances in multi-document summarization in the legal domain have demonstrated significant progress in the extraction and compression of information from legal texts. Current methods utilize a combination of natural language processing, machine learning, and data mining techniques to identify and distill key elements and themes from a multitude of legal documents. This process creates structured, concise, and relevant summaries based on specific legal queries or topics, often referred to as multi-document abstracts. These abstracts facilitate a more efficient review by capturing the essence of complex and voluminous legal materials without losing the necessary detail. The focus of recent research has been on enhancing the accuracy of information retrieval, improving the coherence of the generated summaries, and ensuring the relevacy of the content to the specific legal issue at hand. Although challenges remain, particularly in the nuances of legal language and the diversity of document types, the trajectory of the field is toward more sophisticated and user-friendly systems that promise to transform the landscape of legal research and information accessibility.

§1. Introduction

Information retrieval is a key problem in computer science, especially in locating pertinent objects from extensive datasets. This is especially true for web search engines, whose primary focus is document retrieval. However, within the domain of legal documents, understanding and interpreting statutory provisions pose unique challenges for legal professionals [22]. The intricate nature of statutory language and the need to apply these provisions to diverse and unforeseen circumstances necessitate a more efficient and comprehensive approach to legal document comprehension. Current

Key words and phrases: multi-document summarization, legal documents, natural language processing.

This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, grant no. AP14871214.

approaches range from extractive summarization to deep learning models and graph-based methods [9]. The following literature review aims to provide an overview of the current state-of-the-art methods and their effectiveness in the legal domain.

Several studies have utilized extractive summarization techniques to summarize legal documents. As generative models have advanced significantly, Large Language Models (LLMs) have become indispensable tools for diverse applications. Despite their capabilities, these models face well-known challenges, including the struggle to stay abreast of up-to-date knowledge, the incorporation of long-tail expertise, and the potential risk of compromising private training data. Another area of focus has been the fine-tuning of language models like BERT on legal corpora. This approach has shown promising results in legal question-answering systems. In Vietnamese legal question-answering systems, the BERT language model is fine-tuned on a legal corpus, achieving an 87 F1 score, while pre-training BERT on a legal domain-specific corpus yielded a higher F1-Score of 90.6, showcasing its potential for legal applications [23].

Another study presented a method for summarizing scientific articles using a greedy extractive summarization algorithm. This technique achieved comparable ROUGE scores to state-of-the-art models but used a straightforward statistical inference methodology. This approach demonstrated the potential for efficient summarization of lengthy legal documents.

In response to these challenges, Retrieval-Augmented Generation (RAG) emerges as a promising solution, leveraging an adaptable data repository as a non-parametric memory. The core concept behind RAG is to enhance the generation process by combining the strengths of information retrieval and generative models. This methodology is particularly relevant for legal documents, where the need for precise understanding and application of statutory provisions is paramount. By employing RAG in the legal domain, the aim is to provide legal professionals with a powerful tool that not only aids in comprehending complex statutory language but also assists in efficiently applying legal principles to real-world scenarios.

The RAG process involves an input query, retrieval of relevant legal information from case law, and integration with the LLM to enhance the understanding of statutory provisions. This interaction between retrieval results and the generative process offers multiple possibilities, such as serving as augmented input, contributing as latent representations, supporting lengthy contexts, or influencing specific steps in the generation process.

The adaptable nature of non-parametric memory in RAG proves crucial in accommodating the nuanced and evolving nature of legal language. While RAG has seen extensive application in various domains, including text, codes, audio, images, videos, and more, its potential in legal document assistance still needs to be explored. A comprehensive survey on RAG applications, particularly in the legal context, is necessary for researchers and practitioners to understand its capabilities and potential impact fully.

This work aims to bridge this gap by presenting a systematic exploration of the application of RAG, facilitated by LLMs, in legal document comprehension. We consider the foundational aspects of RAG, examining how it can revolutionize interpreting statutory provisions by combining the strengths of retrieval and generation. Through a comprehensive survey, we seek to address the challenges faced by legal professionals in understanding legal language and pave the way for an innovative tool that could significantly enhance legal document comprehension and application.

§2. Bibliometry

Question answering (QA) has been a significant focus in recent decades, with studies exploring various domains, including law [2–4]. The field has mainly concentrated on long-form question answering (LFQA), which involves the retrieval of information from external documents to generate paragraph-length answers [5–7]. Large language models (LLMs), such as those developed by OpenAI [8], have significantly advanced the field of QA. However, inherent limitations, such as the generation of hallucinated text, have prompted the exploration of innovative approaches like Retrieval Augmented Generation (RAG) [9, 10].

RAG, integrating retrieval and text generation in a unified framework, has demonstrated strong performance in QA [11–13]. Several surveys [14–16] have focused on RAG methodologies, primarily in text-related tasks facilitated by LLMs. However, these surveys often overlook the broader applicability of RAG across multiple modalities within the broader context of Artificial Intelligence and Generative Capabilities (AIGC).

The interpretation of open-textured legal terms has been a recurring theme in the legal domain. Previous efforts involved rule-based reasoning tools and providing case law extracts to aid user understanding. Additionally, automated summarization of open-textured concepts and using

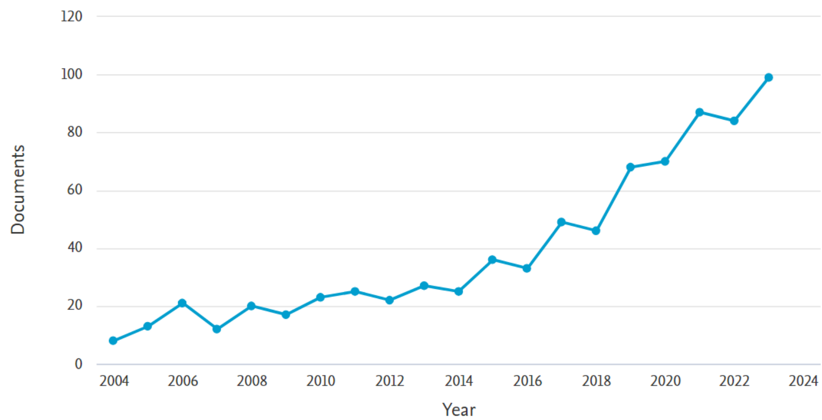

Figure 1. The number of documents on the topic of legal question answering by years.

pre-trained language models to rank sentences in case law have been explored. Building upon this, the application of an augmented LLM, specifically GPT-4, to explain the interpretation of legal terms by incorporating retrieved passages from court cases has been proposed.

Numerous investigations have examined language model augmentation for more accurate and well-reasoned answers, introducing prompting methods and injecting external information into model inputs. In the context of legal documents, a proposed pipeline aims to discover and insert relevant context into prompts based on user queries, transforming identified sections into concise explanations using GPT-4.

Figure 1 displays the number of documents from 2004 to 2024. A general upward trend indicates an increase in the number of documents over time. Notable points of increase are around 2014 and 2018, and a sharper rise after 2020, reaching a peak in 2024. It also shows a few years of plateau or minor declines, particularly around 2008 and 2012. After 2012, there is a noticeable upward trend, with an acceleration post-2020, which could suggest a growing interest or need for research in this area. The steady increase might also indicate advancements in the field, increased funding, or more researchers contributing to the literature.

Figure 2 shows the number of documents associated with various authors. The names are abbreviated, and the bars represent the quantity of

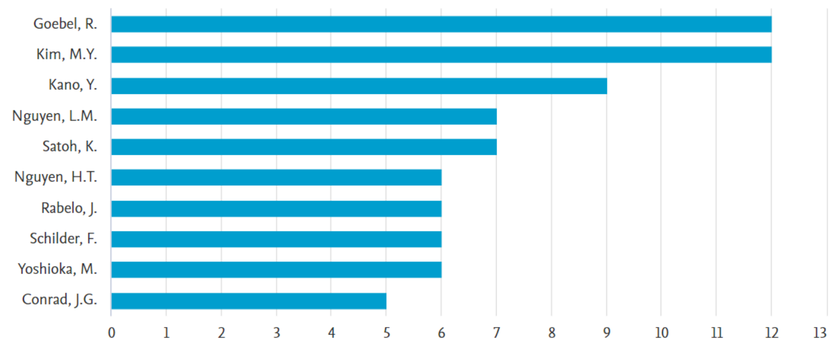

Figure 2. The number of documents on the topic of legal question answering by authors.

documents attributed to each author. R. Goebel, M. Y. Kim, and Y. Kano emerge as the most prolific authors, each having published a comparable and significant number of documents. Other authors, such as L. M. Nguyen and K. Satoh, have also contributed notably, though to a lesser extent. The chart displays a wide distribution of documents among authors, and the distribution shows that a few authors have produced many documents, which is common in academic publications where certain researchers dominate the discourse within specific fields.

Figure 3 represents the distribution of documents across various fields or disciplines. Social Sciences has the largest portion, representing 28.5% of the total, followed by Computer Science with 20.5%. This suggests that these two fields are the most studied or have the most published documents among the categories shown. The significant difference between these and other fields like Medicine, Engineering, Arts and Humanities, or Mathematics could reflect the current research focus or societal priorities. The 'Other' category, which makes up 9.5%, indicates a diversity of lesser-represented fields, which could be niche or emerging areas of study.

Figure 4 illustrates the productivity of various institutions, with the Japan Advanced Institute of Science and Technology and the University of Alberta standing out as highly productive, which could suggest a strong research culture or significant funding at these institutions. The presence of institutions like the University of Alberta and Thomson Reuters indicates a geographic and sectoral diversity in research output. The spread

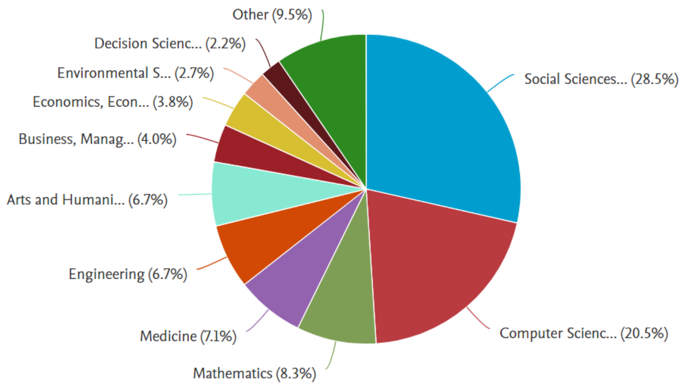

Figure 3. The share of documents on the topic of legal question answering by domain.

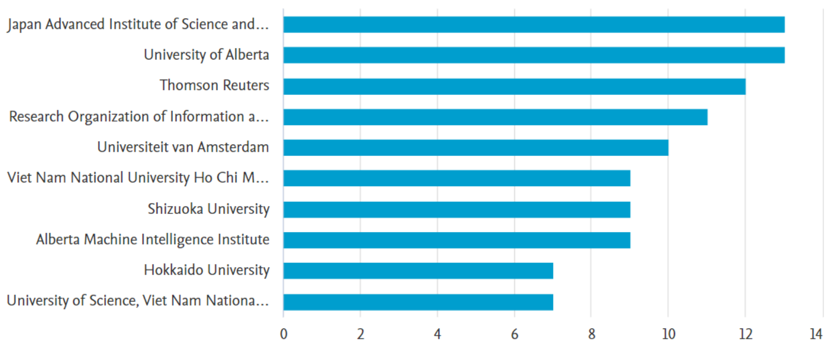

Figure 4. The number of documents on the topic of legal question answering by affiliation.

among institutions suggests that while some are clear leaders, there is still a significant contribution from a range of universities and institutes worldwide.

Figure 5 suggests that certain funding bodies are more active or provide more support for research, with the National Key Research and Development Program leading. This may reflect the strategic priorities of different nations or regions, especially if these organizations are governmental. It

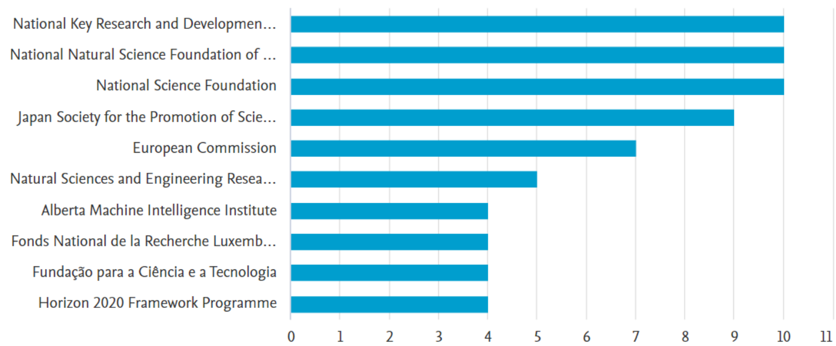

Figure 5. The number of documents on legal question answering by funds.

also may indicate the size and scope of funding available in these regions for research endeavors.

While automatic summarization has been extensively studied in various domains, including news articles and court proceedings, legislative text has yet to be a primary focus. Previous works have predicted bill passage and legislators' voting behavior but often treated documents as "bags of words" without considering the importance of individual sentences [17, 18]. The BillSum corpus, explicitly designed for summarizing legislation, represents a notable advancement in this area [19].

§3. Datasets

There are many existing legal databases. We will review five major databases, two minor, and one Kazakh legal databases: Westlaw, LexisNexis, HeinOnline, Bloomberg Law, EUR-Lex-Sum, and AdiletZan.

(1) *Westlaw:* Westlaw is a comprehensive online legal research platform that offers access to a vast repository of legal documents, including case law, statutes, and academic journal articles.
(2) *HeinOnline:* HeinOnline specializes in offering historical and academic legal materials. Its collections are precious for those conducting thorough legal historical research or seeking scholarly legal analyses. The database includes a wealth of law journal articles, historical legal documents, and more, making it an essential resource for academic research.

Table 1. Comparison of Dataset Sizes Across Legal Databases

Database	**Size of Database**
LexisNexis	293 million
Westlaw	600 million
HeinOnline	208 million
Bloomberg	1.75 million
AdiletZan	0.38 million
EUR-Lex-Sum	3.9 million

(3) *Bloomberg Law:* Bloomberg Law combines legal, business, and news information, providing users with a dynamic resource covering the latest legal developments, financial data, and market analysis.
(4) *LexisNexis:* LexisNexis is a versatile platform providing a wide range of legal, news, and business information.
(5) *AdiletZan:* the information system “Adilet” is a comprehensive web-based search system and electronic collection of Kazakhstan’s normative legal acts. It is notable for being free of charge and offering all national laws in Kazakh and Russian, sourced exclusively from official documents. The database is also regularly updated, ensuring access to the most current legal information.
(6) *EUR-Lex-Sum:* manually curated document summaries of legal acts from the European Union law platform. Documents and their respective summaries are presented as cross-lingual paragraph-aligned data in several of the 24 official European languages, as shown in Figure 6, enabling access to various cross-lingual and lower-resourced summarization setups. The dataset contains up to 1,500 document/summary pairs per language, including a subset of 375 cross-lingually aligned legal acts with texts available in all 24 languages.

§4. Methods

In this section, we consider algorithms designed to provide better functionality and quality of semantic analysis compared to known methods and describe a methodology for quantitatively evaluating their accuracy will be created for each type of algorithm. We survey the following types of algorithms.

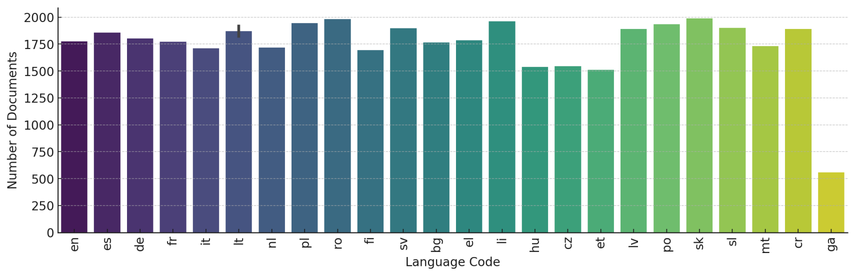

Figure 6. The number of all EUR-Lex-Sum pairs across the 24 official EU languages before filtering.

4.1. Text Summarization Methods.

(1) *Linguistic Structure Analysis*
- *Morphological Analysis* is the study of the internal structure of words, including affixes, roots, and endings. It is used to extract the primary forms of words and determine their grammatical characteristics. It is applied in machine translation, search systems, syntactic analysis, and other areas to improve the understanding of text structure.

(2) *Semantic Analysis*

(a) *Distributional Semantics* explores the semantics of words based on their distribution in the text. Methods include text corpus analysis, identifying word usage contexts, and constructing word vector representations. Word vector representations are used for tasks such as finding similar words, text classification, sentiment analysis, etc. The basic idea of the skip Word vector representations are trained as

$$\max_{\theta} \prod_{t=1}^{T} \prod_{-c \leqslant j \leqslant c, j \neq 0} p(w_{t+j} | w_t; \theta), \tag{1}$$

where w_t is the target word at position t, w_{t+j} is a context word, c is the size of the context window, and θ represents the model parameters. Word vectors are often constructed using methods such as *word2vec* or GloVe. One common approach is the skip-gram model, which predicts context words given a target word, as shown in Algorithm 1.

Algorithm 1: Skip-gram Model

```
Data: Corpus
Result: Predict context words
foreach word w in the corpus do
    foreach context word c within a window around w do
        Predict c given w;
    end
end
```

Algorithm 2: Latent Dirichlet Allocation (LDA)

```
Data: A set of documents
Result: Assigning each word to a topic
foreach document d do
    foreach word w in d do
        Assign w to a topic t with probability P(t|d);
    end
end
```

(b) *Topic Modeling* is the analysis of textual data to identify thematic structure. Methods include probabilistic models such as Latent Dirichlet Allocation (LDA) and matrix decomposition. It is used to extract themes from texts, classify thematic documents, and analyze large corpora's structure. In particular, LDA assumes that each document is a mixture of topics, and each word's presence is attributable to one of the document's topics; pseudocode for the LDA generative model is shown in Algorithm 2.

(c) *Text Processing and Summarization.* The web scraping approach and summarization approach for legal documents uses Python scripts to extract data from legal documents accurately. A comparative study for techniques such as LUHN, LSA, LEXRANK, and SUMBASIC based on ROUGE scores concluded that LUHN and LSA are the most effective algorithms for summarization.

4.2. Question Answering (QA) Methods.

(3) *Knowledge Representation*
- *Knowledge Graphs (KG)* organize information as nodes (entities) and relationships between them. They facilitate intelligent data processing by representing knowledge in a structured graphical form. The task of the Knowledge Graph is to describe and organize information so that computers can understand the relationships between entities and use this knowledge for more intelligent data processing.

(4) *Optimization and Efficiency*
- *Optimization Heuristics* include heuristics and optimization methods to improve the performance of natural language processing models. They are applied to optimize weights, parameters, and model structures. Heuristics are widely used in machine learning and NLP to improve the accuracy and efficiency of algorithms.

(5) *Advanced Language Models*

(a) *BERT* is a pre-trained language representation method and model. These models are used to extract high-quality language features from textual data or use their data to fine-tune these models for specific tasks (classification, entity recognition, question answering, etc.), in particular, for extracting word and sentence embedding vectors. These embeddings are helpful for keyword expansion/search, semantic search, and information retrieval.

(b) *ALBERT* is a Transformer architecture based on BERT but with significantly fewer parameters, achieved through two parameter reduction methods. First, it is the factorized parameterization of embeddings by decomposing a large vocabulary embedding matrix into two small embedding matrices; the hidden layer size is separated from the vocabulary embedding size. This simplifies increasing the former size without significantly expanding the vocabulary embedding parameter size. The second method includes shared parameterization between layers. This method prevents parameter growth with network depth.

Text summarization methods include techniques such as topic modeling, which identifies thematic structures for summarizing texts, and web scraping coupled with summarization algorithms like LUHN, LSA, LexRank, and SUMBASIC, which is particularly effective for translating legal documents based on thematic content.

For question answering, methods involve distributional semantics, exploring word semantics based on text distribution, and leveraging advanced language models like BERT and ALBERT. These models excel in understanding context and extracting information to answer questions effectively.

The Computing with Words methodology based on fuzzy logic can be effectively used in question-answering systems to handle the imprecision and vagueness of natural language queries. It allows systems to interpret subjective and vague terms (e.g., "often" or "a few") by defining them within a range of values [20, 21].

These summarization and question answering methods improve efficiency and accuracy in understanding and extracting insights from textual data.

We strictly adhere to informed consent when working with data from users or research participants during our research. During our research activities, we strictly comply with all applicable laws and regulations regarding data protection and research ethics. We commit to openness and transparency regarding the research methodology and results. The code used in the research will be published, and access to the data will be provided according to privacy policies. We adhere to high standards of responsibility and ethics in natural language processing. We strictly adhere to the principles of fairness, impartiality, and non-discrimination in all aspects of the research.

§5. Evaluation metrics

Evaluating the effectiveness of text summarization of legal documents requires a nuanced approach due to the complexity and specificity of legal language and concepts. Here are some key evaluation metrics and methods that can be used.

(1) **Relevance**: this metric measures how well the summary captures the key points and arguments of the original legal document. It can be evaluated using human judges or automated methods such

as precision, recall, and F1-score,

$$\text{Precision} = \frac{\text{Number of relevant items retrieved}}{\text{Total number of items retrieved}},$$

$$\text{Recall} = \frac{\text{Number of relevant items retrieved}}{\text{Total number of relevant items}},$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

(2) **Completeness**: this metric measures how much of the original legal document's content is included in the summary; it can be assessed using automated methods such as the percentage of sentences or words from the original document included in the summary,

$$\text{Completeness} = \frac{\text{Original summary number of sentences or words}}{\text{Total number of sentences or words in origin}}.$$

(3) **Conciseness**: This metric measures how efficiently the summary conveys the information from the original legal document. It can be evaluated using automated methods like the average length of sentences or the ratio of summary length to original document length,

$$\text{Avg. length} = \frac{\text{Total number of words in the summary}}{\text{Total number of sentences in the summary}},$$

$$\text{Ratio} = \frac{\text{Length of summary}}{\text{Length of original document}}.$$

(4) **Clarity**: This metric measures how the summary communicates the information from the original legal document. It can be assessed using human judges or automated methods, such as readability scores or the percentage of sentences that are easy to understand:

$$\begin{aligned}&\text{Readability(Flesch-Kincaid)}\\&= 206.835 - 1.015 \times \left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6 \times \left(\frac{\text{total syllables}}{\text{total words}}\right).\end{aligned}$$

(5) **Consistency**: This metric measures the consistency of the information presented in the summary. It assesses whether the summary is internally consistent and accurately represents the original legal document. Consistency can be evaluated using human judges or automated methods like cosine similarity or Jaccard similarity; in the following definition, A and B are term frequency vectors of the original document and the summary:

$$\text{Cos. Similarity} = \frac{A \cdot B}{\|A\| \cdot \|B\|}.$$

(6) **Timeliness**: This metric measures the actuality of the information in the summary. It assesses whether the summary includes recent developments or changes in the law. Timeliness can be evaluated using automated methods such as the publication date of the documents included in the summary.

(7) **Bias**: This metric measures the presence of bias in the summary. It assesses whether the summary is biased toward specific sources or viewpoints. Bias can be evaluated using automated methods, such as determining the summary's diversity of sources or perspectives.

(8) **Usability**: This metric measures how easily users can utilize the summary. It assesses whether the summary is user-friendly and intuitive. Usability can be evaluated using human judges or automated methods, such as when users find information in the summary.

(9) **Scalability**: This metric measures how well the summary scales to handle many legal documents. It assesses whether the summary can handle a high volume of documents without sacrificing performance. Scalability can be evaluated using automated methods, such as the time it takes for the summary to generate a response to a query:

$$\text{Response time} = \frac{\text{Time taken to generate a summary}}{\text{Number of documents processed}}.$$

(10) **Robustness**: This metric measures how well the summary performs under different conditions. It assesses whether the summary is robust to changes in the legal documents. Robustness can be evaluated using automated methods, such as the stability of the summary's performance over time.

§6. Systems

How is AI used in jurisprudence today?

In recent years, there has been a notable integration of artificial intelligence within jurisprudence. Legal firms and organizations leverage AI to streamline processes, alleviate employees' workload, and save time and labor costs. This trend reflects a growing acknowledgment of AI's potential to enhance efficiency within the legal profession. We will consider all available tools for a lawyer in Kazakhstan:

- designers of documents working based on standard and unified templates, in which any deviation from the form requires manual correction;
- counterparty verification services that aggregate publicly available information from public registers, which rarely allow finding valuable information;
- judicial practice selection systems and legal reference systems that perform basic searches by keywords, phrases, and tags in the open database of court decisions, and NPAs, which provide all documents containing the searched word without considering the context, etc.

This list may include the following:

- *E-otinish* – for processing requests for information from government agencies
- *Zakon.kz* – Latest news from Kazakhstan and the world, changes in the legislation of the Republic of Kazakhstan
- *IS "Paragraph"* is a unique reference system that combines several blocks of information, from specific information for lawyers, accountants, and medical workers to reference information everyone needs daily.
- Database of judicial acts is a service that contains legal acts (decisions, rulings, sentences) that entered into legal force or were issued by courts in 2009–2024.

These tools do not bring complete automation of creative and expert jurisprudence; they certainly facilitate a lawyer's work, but only in searching for information.

However, in the world market, the automation of the listed functions has already allowed significant relief to the employees of courts, lawyers, notaries, legal companies, and organizational departments. Artificial intelligence copes with most tasks since a person's creative abilities are not required to perform such functions.

In the USA, the penetration of AI technologies into jurisprudence is even more comprehensive. For example, Dashboard Legal, a project management system, helps lawyers and attorneys track documents and deadlines and collaborate remotely with colleagues, making "good old" e-mails unnecessary.

The innovative service Bankrotech, which Sberbank presented, can be considered an excellent example of how AI benefits a lawyer's work. This is a single aggregator system for conducting bankruptcy procedures, which collects information about participants in court procedures (bankruptcies) from all available sources and allows you to structure the storage of data and documents, forecast and form strategies for working with assets, and interact with other creditors.

ChatGPT. Another idea is to use the neural network as a junior lawyer, for example, to draw up simple contracts or search for various information. In short, it is possible to use ChatGPT in legal work, but its results must be carefully checked. Still, LLMs can be very helpful for lawyers now.

Technologies in jurisprudence can also be referred to as LegalTech and Legal AI. A few products that are used and popular in different countries, including USA, United Kingdom, Canada, and Australia:

- *Jurispect* is a platform that uses artificial intelligence to analyze legal issues and automatically generate legal reports, providing clients with quick and accurate answers to legal questions.
- *LegalSifter* is a product that uses machine learning technology to scan contracts and other legal documents to identify important terms and provide recommendations for action;
- *Neota Logic* automates legal processes, including generating legal documents and reports. It allows you to create intelligent applications for solving various legal problems;
- *Doxly* provides opportunities for joint work on legal documents and automating document preparation processes, including generating reports and reports.

These products help lawyers and other legal professionals efficiently process large volumes of information and quickly obtain the necessary legal documents and documentation.

Table 2. Comparison Table.

Name	Rel	Comp	Con	Cla	Con	Tim	Bias	Usa	Sca	Rob	Free	Rank
LexisNexis	1	1	1	1	1	1	1	1	1	1		10
LegalSifter	1	1	1	1	0	1	0	1	1	1		8
Neota Logic	1	1	0	1	1	0	1	0	1	0		6
ChatGPT	1	0	0	1	0	1	0	1	1	1		6
Jurispect	1	0	0	0	0	1	1	0	0	0		3
Doxly	1	0	1	0	1	0	0	0	0	0		3
Bankrotech	0	0	0	0	0	0	0	0	0	1		1

§7. System comparison

We selected Bankrotech, ChatGPT, Jurispect, LegalSifter, Neota Logic, Doxly, and LexisNexis technologies as the most popular tools and compared them according to the following characteristics:

- **Relevance:** Shows how well the summary reflects the key points and arguments of the source document;
- **Completeness:** This indicator measures how much of the content of the original legal document is included in the summary;
- **Conciseness:** This metric measures how efficiently the summary conveys the information from the original legal document;
- **Clarity:** This metric measures how the summary communicates the information from the original legal document;
- **Consistency:** It assesses whether the summary is internally consistent and whether it accurately represents the original legal document;
- **Timeliness:** This metric measures how up-to-date the information in the summary is;
- **Bias:** It assesses whether the summary is biased toward specific sources or viewpoints;
- **Usability:** This metric measures how easy it is for users to use the summary;
- **Scalability:** This metric measures how well the summary scales to handle many legal documents;
- **Robustness:** It assesses whether the summary is robust to changes in the legal documents.

The result are shown in Table 2. The table ranks LexisNexis highest, excelling across all assessed areas and providing comprehensive and precise legal summaries. LegalSifter and ChatGPT also perform well, with some limitations noted for LegalSifter. Neota Logic, Jurispect, and Doxly show moderate performance, with potential for enhancement through fine-tuning. Bankrotech ranks lowest, suggesting room for improvement in legal information services. These insights highlight the benefit of integrating multiple systems for a well-rounded legal information solution, supplemented by plagiarism checks as needed.

§8. Conclusion

In this work, we have presented a survey on legal question answering, providing a comprehensive overview of the current state-of-the-art methods and technologies in the field. It covers various aspects, including the challenges in legal document comprehension, methods and algorithms for text summarization and question answering, evaluation metrics for assessing summarization effectiveness, the utilization of AI in jurisprudence, comparison of popular legal information systems, and implications for the legal profession.

Overall, our research contributes to advancing the understanding of how natural language processing and artificial intelligence can be applied to address the unique challenges faced in the legal domain. It sheds light on the potential of innovative approaches like Retrieval-Augmented Generation (RAG) and advanced language models such as BERT and ALBERT in improving legal document comprehension and information retrieval.

Furthermore, the research underscores the importance of ethical considerations, transparency, and adherence to legal and ethical standards in developing and deploying AI technologies in jurisprudence. By providing insights into the strengths and limitations of existing systems and methodologies, the research is a valuable resource for legal professionals, researchers, and practitioners seeking to leverage AI for enhanced legal document analysis and decision-making.

References

1. I. Akhmetov and A. Pak, *TSP review: performance comparison of the well-known methods on a standardized dataset*, Proc. 2023 19th Int. Asian Sch.-Semin. Optim. Probl. Complex Syst. (OPCS) (2023), pp. 4–9.

2. A. Ravichander, A.W. Black, S. Wilson, T. Norton, and N. Sadeh, *Question answering for privacy policies: Combining computational and legal perspectives*, Proc. 2019 Conf. Empir. Methods Nat. Lang. Process. Int. Joint Conf. Nat. Lang. Process. (EMNLP-IJCNLP) (2019), pp. 4947–4958.
3. P. Wang, L. Li, L. Chen, Z. Cai, D. Zhu, B. Lin, Y. Cao, Q. Liu, T. Liu, and Z. Sui, *Large Language Models are not Fair Evaluators*, ArXiv preprint arXiv:2305.17926 (2023).
4. A. Louis, G. van Dijck, and G. Spanakis, *Interpretable Long-Form Legal Question Answering with Retrieval-Augmented Large Language Models*, ArXiv preprint arXiv:2309.17050 (2023).
5. K. Krishna, A. Roy, and M. Iyyer, *Hurdles to Progress in Long-form Question Answering*, ArXiv preprint arXiv:2103.06332 (2021).
6. S. Wang, F. Xu, L. Thompson, E. Choi, and M. Iyyer, *Modeling Exemplification in Long-form Question Answering via Retrieval*, Proc. 2022 Conf. North Am. Chapter Assoc. Comput. Linguist.: Hum. Lang. Technol. (2022), pp. 2079–2092.
7. F. Xu, Y. Song, M. Iyyer, and E. Choi, *A Critical Evaluation of Evaluations for Long-form Question Answering*, ArXiv preprint arXiv:2305.18201 (2023).
8. J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F.L. Aleman, et al., *GPT-4 Technical Report*, ArXiv preprint arXiv:2303.08774 (2023).
9. Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y.J. Bang, A. Madotto, and P. Fung, *Survey of Hallucination in Natural Language Generation.* — ACM Comput. Surv. **55**(12) (2023), article 248, 38 pages.
10. P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*, ArXiv preprint arXiv:2005.11401 (2021).
11. G. Mialon, R. Dessì, M. Lomeli, C. Nalmpantis, R. Pasunuru, R. Raileanu, B. Rozière, T. Schick, J. Dwivedi-Yu, A. Celikyilmaz, E. Grave, Y. LeCun, and T. Scialom, *Augmented Language Models: a Survey*, ArXiv preprint arXiv:2302.07842 (2023).
12. A. Lazaridou, E. Gribovskaya, W. Stokowiec, and N. Grigorev, *Internet-augmented language models through few-shot prompting for open-domain question answering*, ArXiv preprint arXiv:2203.05115 (2022).
13. J. Saad-Falcon, O. Khattab, C. Potts, and M. Zaharia, *ARES: An Automated Evaluation Framework for Retrieval-Augmented Generation Systems*, ArXiv preprint arXiv:2311.09476 (2023).
14. H. Li, Y. Su, D. Cai, Y. Wang, and L. Liu, *A Survey on Retrieval-Augmented Text Generation*, ArXiv preprint arXiv:2202.01110 (2022).
15. A. Asai, S. Min, Z. Zhong, and D. Chen, *Retrieval-based Language Models and Applications*, Proc. 61st Annu. Meet. Assoc. Comput. Linguist. (2023), pp. 41–46.
16. Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, Q. Guo, M. Wang, and H. Wang, *Retrieval-Augmented Generation for Large Language Models: A Survey*, ArXiv preprint arXiv:2312.10997 (2024).
17. S.M. Gerrish and D.M. Blei, *Predicting legislative roll calls from text*, Proc. 28th Int. Conf. Mach. Learn. (2011), pp. 489–496.

18. T. Yano, N.A. Smith, and J.D. Wilkerson, *Textual Predictors of Bill Survival in Congressional Committees*, Proc. 2012 Conf. North Am. Chapter Assoc. Comput. Linguist.: Hum. Lang. Technol. (2012), pp. 793–802.
19. M. Kim, Y. Xu, and R. Goebel, *Summarization of Legal Texts with High Cohesion and Automatic Compression Rate.* — New Front. Artif. Intell. (2013), pp. 190–204.
20. A. Kali, P. Shamoi, Y. Zhangbyrbayev, and A. Zhandaulet, *Computing with Words for Industrial Applications.* — Intell. Syst. Appl. (2023), pp. 257–271.
21. P. Shamoi and A. Inoue, *Computing with Words for Direct Marketing Support System*, Proc. Midwest Artif. Intell. Cogn. Sci. Conf. (2012).
22. J.B. Ruhl, D.M. Katz, and M.J. Bommarito, *Harnessing legal complexity.* — Science, **355**(6332) (2017), pp. 1377–1378.
23. C.-N. Chau, T.-S. Nguyen, and L.-M. Nguyen, *VNLawBERT: A Vietnamese Legal Answer Selection Approach Using BERT Language Model*, Proc. 2020 7th NAFOSTED Conf. Inf. Comput. Sci. (2020), pp. 298–301.

Поступило 15 ноября 2024 г.

Kazakh-British Technical University,
Almaty, Kazakhstan
E-mail: ak_sabiyeva@kbtu.kz

Kazakh-British Technical University, Almaty, Kazakhstan
E-mail: al_zhamankhan@kbtu.kz

Kazakh-British Technical University, Almaty, Kazakhstan
E-mail: n_zhetesov@kbtu.kz

Kazakh-British Technical University, Almaty, Kazakhstan
E-mail: a_kubaeva@kbtu.kz

Kazakh-British Technical University, Almaty, Kazakhstan
Institute of Information and Computational Technologies, Almaty, Kazakhstan
E-mail: i.akhmetov@kbtu.kz,i.akhmetov@ipic.kz

Kazakh-British Technical University, Almaty, Kazakhstan
Institute of Information and Computational Technologies, Almaty, Kazakhstan
E-mail: a.pak@kbtu.kz,a.pak@ipic.kz

Kazakh-British Technical University, Almaty, Kazakhstan
E-mail: di_akhmetova@kbtu.kz

Kazakh-British Technical University, Almaty, Kazakhstan
Institute of Information and Computational Technologies, Almaty, Kazakhstan

Institute of Information and Computational Technologies,
Almaty, Kazakhstan