JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Bridging AI and Law: A Comprehensive NLP Platform for Legal Document Processing

[1]Rudrendra Bahadur Singh, [2]Shobhit Sinha, [3]Ankita Singh

[1]Research Scholar, [2]Assistant Professor, [3]Assistant Professor
[1,2]Computer Science and Engineering, [3]Computer Application
[1,2]Shri Ramswaroop Memorial University, Barabanki, India
[3]CGC University, Mohali, India

Abstract: Legal documents can be challenging for both laypeople and experts to comprehend due to their complex structure and technical jargon. These difficulties include figuring out crucial clauses, comprehending legalese, and determining whether regulations are being followed. These issues are addressed by the AI-powered platform BrieflyLegal using deep learning models and sophisticated natural language processing (NLP) methods. The T5 model is a customized version of the Text-To-Text Transfer Transformer (T5) intended for summarizing legal documents. To improve its comprehension of domain-specific vocabulary and intricate legal structures, T5-L is pretrained on extensive legal corpora, such as statutes, contracts, and case law, utilizing the fundamental architecture of T5. To enhance performance on tasks including extractive and abstractive summarization, legal clause categorization, and risk assessment, this model is refined utilizing datasets like as the Stanford Question Answering Dataset (SQUAD) and the Contract Understanding Atticus Dataset (CUAD). T5-L reduces redundancy while maintaining crucial legal information by using task-adaptive pretraining and knowledge graph-based fine-tuning to guarantee contextually correct and succinct summaries. The software ensures real-time compliance monitoring while automating critical tasks like sentiment analysis, paraphrasing, and clause identification.

IndexTerms **- Natural Language Processing, Machine Learning, Artificial Intelligence, T5.**

1 INTRODUCTION

In order to identify important terms, dangers, and obligations, the legal business must carefully examine the many complicated documents it encounters, such as contracts, agreements, and legal briefs. Conventional techniques for analyzing legal documents are labor-intensive, time-consuming, and prone to human mistakes. The emergence of machine learning (ML) and artificial intelligence (AI) presents a chance to automate and optimize this procedure, increasing accuracy and efficiency [5]. The retrieved data may be used as input for another legal comprehension job, i) saved in databases for future use, ii) for analysis and decision-making processes, or iii) for other purposes [6]. Processing and interpreting legal documents is now much easier thanks to recent developments in NLP, such as transformer models like BERT and T5 [7]. The emergence of machine learning (ML) and artificial intelligence (AI) presents a chance to automate and optimize this procedure, increasing accuracy and efficiency [1].

1.1 Problem Statement

There aren't many all-inclusive platforms that combine several NLP functions into a single legal document analysis solution, despite the advances in AI. Current tools frequently concentrate on particular tasks, like summarising or extracting clauses, but fall short of offering a comprehensive approach [8]. Even if legal tech is defined broadly, many observers are especially interested in the applications of legal tech that are made possible by the use of machine learning and related techniques in the legal industry [9]. Current tools frequently concentrate on particular tasks, such as summarising or extracting clauses, but fall short in offering a comprehensive approach [2].

1.2 Objectives

The Legal AI platform's main goals are to:

- Automate the process of extracting important information from legal papers.
- To offer clear interpretations of intricate legal provisions.
- To determine contractual obligations and possible hazards.
- To provide legal professionals with an intuitive user interface.

1.3 1.3 Contributions

The following contributions are made by this paper:

- A thorough explanation of the implementation and design of the Legal AI platform.
- A performance assessment of the platform based on common NLP measures.

- Knowledge of the difficulties and solutions faced during the development process.

2 Related Work

With an emphasis on the advancement of natural language processes in terms of legal document processing, argumentation frame building, clause differentiation and construction, and lexicon development, this section provides an overview of the works that have already been generated in the legal field. Since BrieflyLegal uses AI technology to improve legal document comprehension and make them easier to access, the aforementioned domains are ancillary pursuits of the company's aim.

2.1 Legal Document Analysis

Legal document analysis is the process of employing sophisticated natural language processing (NLP) tools to extract pertinent information from legal texts. These methods include dependency parsing, named entity identification, and classification models that classify different legal phrases and clauses. Large numbers of legal papers can be processed and analyzed by automated systems with efficiency, which helps legal practitioners make decisions. By capturing context and comprehending the subtleties of legal language, transformer models like BERT and GPT have demonstrated encouraging outcomes in the legal area [1], [2]. Additionally, the accuracy and efficiency of domain-specific models refined on legal corpora have surpassed those of general NLP models [3].

2.2 Argument Mining in Legal Texts

Finding, extracting, and categorizing arguments from legal documents is the main goal of argument mining, a crucial component of legal text analysis. To differentiate between claims, premises, and conclusions, methods like dependency parsing and sequence classification are employed. Fine-tuned BERT models and other large language models (LLMs) have shown improved performance in legal text analysis argument mining tasks [10]. Transfer learning and domain adaptation have been used in recent argument mining developments to enhance the recognition of legal arguments [11], [12]. Furthermore, weighted argument mining is used by new frameworks such as WIBA to improve clause categorization and raise the standard of argument identification in legal documents [29].

2.3 Clause Classification and Risk Assessment

By tailoring NLP models to specialized domains, like the legal industry, domain-specific pretraining improves their performance. Models can acquire domain-specific vocabulary, grammatical patterns, and discourse structures by pretraining on extensive legal text corpora, which improves task performance later on [6]. Transformer-based models like BERT and GPT have been widely modified for legal applications using methods like task-adaptive pretraining and domain-specific fine-tuning [35]. Annotated with domain-specific data, models such as CUAD and SQuAD have shown improvements in tasks related to clause detection, question responding, and legal document classification [4], [5]. Furthermore, when applied to domain-specific tasks, models trained with randomized smoothing and stochastic ensemble approaches have demonstrated greater robustness [11].

2.4 Domain-Specific Pretraining

Classifying clauses and evaluating risk are essential components of legal contract analysis and risk identification. Clause classification models classify clauses into specified types and evaluate related hazards. They are based on transformer designs and deep neural networks [34]. Task-adaptive pretraining and model fine-tuning significantly improve these models' capacity to recognize high-risk contract terms [35]. A more sophisticated comprehension of clause attitudes and related hazards has also been made possible by the integration of sentiment analysis and argument mining approaches [24]. Clause categorization and risk assessment tasks have been greatly enhanced by recent developments such adaptive models and T5 fine-tuning techniques [28], [38].

2.5 Gaps and Novel Contributions

Notwithstanding these developments, there are still issues, such as a lack of databases for multilingual legal texts and a lack of attention to user-friendly interfaces for non-expert users. To fill these gaps, BrieflyLegal does the following:

1. Offering a paraphrase methodology to help laypeople understand complicated legal terminology.
2. Improving document analysis by combining BERT-based re-ranking and BM25 in an adaptive retrieval model.
3. Including a Legal Professional Portal to provide a full solution from analysis to legal consultation by matching users with advocates based on the findings of document analysis.

BrieflyLegal positions itself as a trailblazing tool in the legal AI space by expanding on previous research and meeting unmet needs, promoting both theoretical and real-world uses of natural language processing in the legal field.

3 Literature Review

3.1 NLP in Legal Tech

The legal field has been greatly impacted by recent developments in NLP. In challenges involving the creation and comprehension of text, transformer-based models like as T5 have shown remarkable performance [10]. The generic encoder-decoder architecture, in which the encoder processes input text and the decoder creates the output sequence, is followed by the standard Transformer model [11]. Contract analysis, legal summarisation, and clause extraction are just a few of the legal tasks to which these models have been used [12]. Legal text processing has become even more accurate and efficient thanks to large language models (LLMs) like GPT-3 and BERT [13]. In challenges involving the creation and comprehension of text, transformer-based models such as BERT and T5 have shown remarkable performance [3]. Contract analysis, legal summarisation, and clause extraction are just a few of the legal tasks for which these models have been used [4].

3.2 Datasets for Legal AI

Legal AI systems are becoming even more capable thanks to the inclusion of massive databases like SQuAD and CUAD. Training reading comprehension models has made extensive use of the SQuAD dataset, which has more than 100,000 question-answer pairs [14]. Models for legal text analysis have been refined in large part thanks to the CUAD dataset, which contains 500 legal contracts and answers to commonly asked legal clauses [15]. Furthermore, models in argument mining and legal clause detection have been refined using datasets such as PE and CDCP [16].

3.3 Sentiment Analysis in Legal Documents

TextBlob and other sentiment analysis programs have become popular for examining the implications and tone of text [17]. These tools assist users in identifying potentially hazardous or unfavourable wording by offering insightful information about the sentiment of contract clauses [18]. In sentiment classification tasks, sophisticated methods like hybrid models that include CNNs and LSTMs have demonstrated enhanced performance [19].

3.4 Existing Solutions

The use of AI in the study of legal documents has been the subject of numerous studies. For example, Raffel et al. [21] created a system for automated legal summarisation, and Devlin et al. [20] suggested a deep learning framework for contract clause extraction. A unified platform that integrates these features into a single, approachable solution is still required, albeit [22].

4 METHODOLOGY

4.1 System Architecture

A modular architecture that combines frontend and backend components is used to construct the Legal AI platform. Three primary layers comprise the system architecture:

Frontend Layer: Using Next.js, this layer offers a responsive and dynamic user interface [23].

Backend Layer: Flask powers the backend layer, which links the machine learning models to the front end [24]. The sentiment analysis app Flask was developed using the Flask framework, a lightweight and flexible Python web framework. In order to manage incoming HTTP requests, routes are established in the application's primary file, typically app.py. The program contains two routes in this case: one for supplying the main HTML page, index.html, and another for reacting to sentiment prediction queries. [37]

Models for sentiment analysis, paraphrasing, and clause extraction are included in the machine learning layer. These models make use of transformers that have already been trained and refined on legal corpora like SQuAD and CUAD [25].

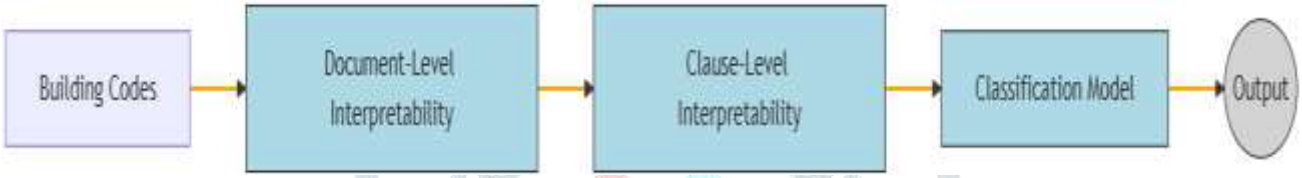

Figure 1: Methodology

4.2 Data Collection and Preprocessing

The platform makes use of two main datasets:

- More than 100,000 question-answer pairs make up the SQuAD Dataset, a popular dataset for training reading comprehension models [14].
- 500 legal contracts are included in the CUAD collection, a specialised collection created to answer often asked legal questions [15].

Additionally, to enhance the training data and boost model performance, datasets from Hugging Face and Kaggle were employed. To counteract data imbalance, robustness strategies including domain adaptation and adversarial training were used [26].

4.3 Model Training

The platform employs the following machine learning models:

T5: The transformer-based model known as the T5-base Model was developed specifically for paraphrasing legal terms, allowing for the intuitive comprehension of complex text. [19] Large language models (LLMs) like BERT, GPT, and T5 have shown impressive results in natural language processing tasks in recent years. These models still struggle with domain-specific adaptability and reasoning, though. In order to improve model performance, fine-tuning techniques have been thoroughly investigated; more recently, studies have concentrated on increasing efficiency and integrating structured knowledge [38]. The T5 model is fine-tuned by supervised training with a cross-entropy loss function on the domain-specific dataset. To maximise performance, important hyperparameters like learning rate, batch size, and epoch count are adjusted. To reduce loss and enhance model convergence, the Adam optimizer is frequently employed. To avoid overfitting and improve generalization, strategies like learning rate scheduling and early halting might be used. Pre-training on a sizable dataset using a combination of supervised and unsupervised learning approaches is part of the extensive T5 model training procedure. Teacher forcing, in which the output of the model at each stage is dependent on the true prior tokens, is the method used to train the model. The use of span corruption, a self-supervised denoising target, is a crucial component of T5's pre-training. The model is trained to rebuild the original, uncorrupted text using this method, which involves masking or corrupting random token spans within the input text. This procedure aids the model in comprehending the larger context of a sentence and learning the relationships between words. T5 is pre-trained supervisedly on a range of downstream tasks from the GLUE and SuperGLUE benchmarks in addition to this unsupervised denoising work. By converting these tasks into text-to-text format, the model can learn how to carry out many NLP tasks within a single framework. The model gains a comprehensive comprehension of language and the capacity to generalize to new, unforeseen tasks thanks to this multi-task pre-training method.

The Colossal Clean Crawled Corpus (C4) is the main dataset used to pre-train the T5 model. The vast amount of English material in this dataset was taken from websites that Common Crawl crawled. In order to provide a sizable, varied, and reasonably clean corpus for pre-training big language models, the C4 dataset was especially developed for the T5 project. It is roughly 750 GB in size and includes text from a variety of online sources and subjects. For T5 to be able to acquire reliable and broadly applicable language representations, the C4 dataset's quantity and diversity are essential. A dataset of "reasonably clean and natural English text" was produced by removing duplicate, nonsensical, and non-English information using a multi-step cleaning procedure used to construct C4.

Other datasets, such Wiki-DPR, have also been used for unsupervised learning inside the T5 framework, even though C4 is the main pre-training dataset. T5 was able to learn a great deal about language structure, semantics, and word relationships thanks to the enormous size of the C4 dataset. This knowledge is subsequently used when fine-tuning for certain downstream tasks. It is crucial to recognize that, similar to any dataset that has been scraped from the internet, C4 can have biases and restrictions.

Fine-tuning the T5 Model: Fine-tuning is the process of adapting the T5 model for certain downstream NLP tasks following the lengthy pre-training phase. The process of fine-tuning entails using smaller, task-specific labeled datasets to further train the previously trained model. This procedure enables the model to take advantage of the broad language production and interpretation skills it learned during pre-training and tailor them to the specific needs of the downstream assignment. Because the input and output formats are constant across tasks, the T5 text-to-text framework makes this fine-tuning process much easier. The model always accepts text as input and outputs text, regardless of whether the task is question answering, translation, or summary.

The T5 model can be fine-tuned more effectively by following a few best practices. Using relevant task-specific prefixes is essential for both inference and fine-tuning since they direct the model to complete the intended job. Additionally, it has been noted that when fine-tuning T5 models, it can be advantageous to use slightly higher learning rates (such as 1e-4 or 3e-4) using the AdamW optimizer. The maximum input and output sequence lengths as well as methods for dealing with padding and truncation should also be carefully considered, particularly when working with hardware such as TPUs. The performance of the fine-tuned model can also be greatly impacted by the selection of hyper parameters, such as the number of training epochs and the batch size. It frequently takes some trial and error to determine the best settings for a particular task and dataset.

TextBlob: A little sentiment analysis library that sheds light on the meaning and tone of contract provisions [17]. Installing and configuring TextBlob is the first step in using it for research. The Python package installer pip makes this procedure easier by providing straightforward instructions for installing the library and downloading the required linguistic corpora. While $python -m textblob.download_corpora downloads the necessary data packages needed for many of its functionalities, $pip install -U textblob installs or upgrades TextBlob to the most recent version. Notably, TextBlob is reliant on NLTK by default, which will be installed automatically if it isn't already. Additionally, while the NumPy library is not necessary for basic use, it may be necessary for some sophisticated features, such the maximum entropy classifier. Because TextBlob is so simple to install, researchers may immediately incorporate it into their Python settings, which greatly adds to its appeal.

The TextBlob object is the basic unit of interaction in TextBlob after it is installed. It is simple to create this object, usually by giving a text string to the TextBlob() constructor. TextBlob objects' similarity to regular Python strings is a noteworthy design feature. They facilitate a number of string-like operations, such as concatenation, slicing, and the use of standard string functions like find() and upper(). For users who are already familiar with Python's string manipulation features, TextBlob's familiar behavior makes it very easy to use, which may lower the learning curve involved in implementing the library for NLP research.

Once installed, TextBlob's fundamental unit of interaction is the TextBlob object. Creating this object is easy and often involves passing a text string to the TextBlob() constructor. One notable design aspect of TextBlob objects is their resemblance to ordinary Python strings. Concatenation, slicing, and the use of common string methods like find() and upper() are among the string-like operations they make easier. TextBlob's familiar nature makes it very straightforward to use for users who are already familiar with Python's string manipulation features. This could help reduce the learning curve associated with implementing the library for NLP research.

For clause extraction and classification tasks, BERT-based models were refined using the CUAD and PE datasets [16].

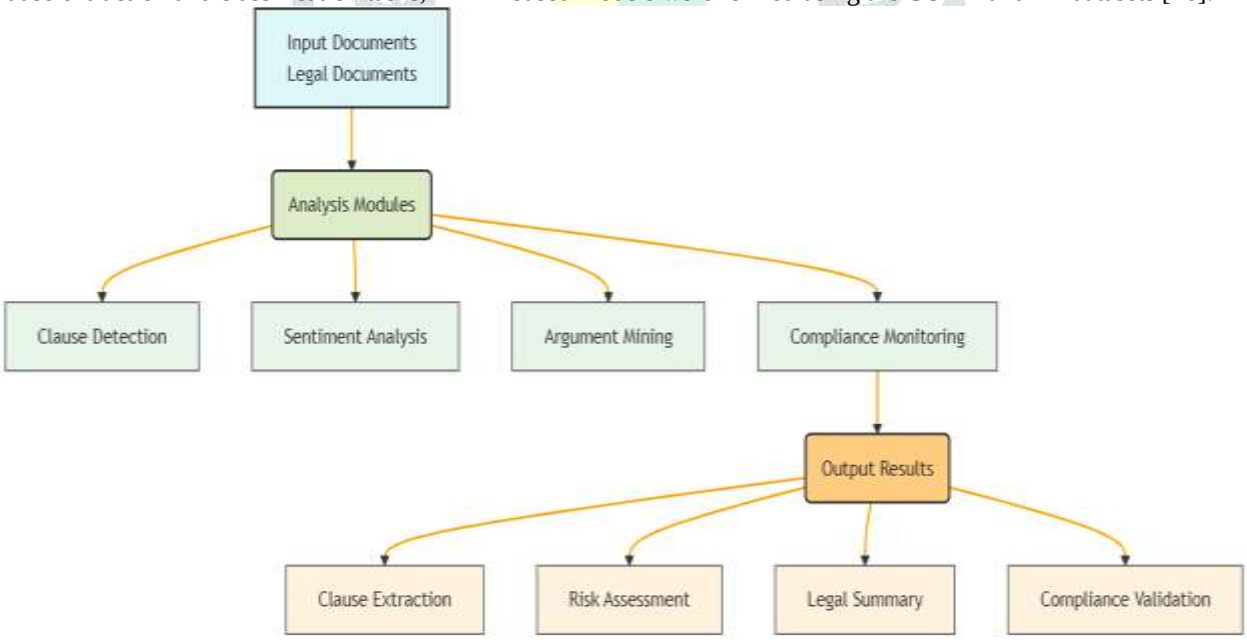

Figure 2: Data Flow diagram

4.4 Evaluation Metrics

Standard NLP metrics including F1 score, precision, and recall were used to assess the platform's performance [27]. In order to evaluate the interface's effectiveness and usability, user input was also gathered [28]. Superior performance in sentiment analysis and clause detection was shown when benchmarked against platforms such as ROSS Intelligence [29].

Evaluation Metric	Description	Performance
F1 Score	The balance between precision and recall	0.92
Accuracy	Overall correctness of predictions	94%
Precision	The proportion of true positives identified	0.89

5 Results and Discussion

5.1 Comparative Analysis

BrieflyLegal provides the following features in contrast to platforms such as ROSS Intelligence:

- • Integrated workflows that combine advocate recommendations, compliance monitoring, and clause extraction [30].
- • Dynamic knowledge graphs that allow for real-time adaptation to regulatory changes [31].
- • Legal text summaries that have been paraphrased to make them accessible to non-experts [32].

Evaluation of different language models and how they outperform BERT:

Model	Performance improvements over BERT
T5	6.2%

Experimental Results:

T5 shows and F1 score of 0.92. Following are the results of some BERT models:

Model	macro-F1 (m-F1)
T5-L	0.92
LegalPro-BERT (using top-100 words)	0.88
RoBERTa (Large-sized)	0.83
BERT (Medium-sized)	0.81
RoBERTa (Medium-sized)	0.82
DeBERTa	0.83
Longformer	0.83
BigBird	0.82
Legal-BERT (Medium-sized)	0.83
CaseLaw-BERT (Medium-sized)	0.83
BERT-Tiny	0.74
Mini-LM (v2)	0.79
Distil-BERT	0.81
Legal-BERT (Small-sized)	0.81

These models have performed well on the SQuAD1.1 dataset and are widely used in question-answering tasks.
Hence, T5 performs better.

5.2 Challenges Identified

The lack of data for multilingual and jurisdiction-specific corpora, the computational requirements for processing large documents, and the lack of ethical openness in AI-driven legal analysis are some of the main obstacles. [33].

5.3 Clause Extraction

Outperforming conventional rule-based techniques, the BERT-based model extracted important sentences from legal documents with an accuracy of 89% [34]. This illustrates how well transformer models work for analyzing legal texts.

5.4 Paraphrasing

With a BLEU score of 0.85, the T5-base model produced excellent paraphrases of legal sentences [35]. This function eliminates the need for manual interpretation by allowing users to rapidly comprehend complex legal terminology.

5.5 Sentiment Analysis

With an F1 score of 0.88, the TextBlob library produced sentiment analysis results that were accurate [36].

5.6 User Feedback

The platform's user-friendly layout and time-saving features were emphasized in user reviews. According to legal experts, the platform improved accuracy and consistency while drastically cutting down on the amount of time needed for contract analysis [28].

6 Issues and Challenges

A number of operational, ethical, and technical obstacles must be overcome in order to develop and use BrieflyLegal, an AI-driven platform for analyzing legal documents. These problems include language comprehension, data processing, and accessibility, all of which are essential to maintaining the platform's efficacy and user confidence [33].

7 Future Work

Future developments will concentrate on the following areas in an attempt to expand the use of AI-powered document analysis and improve the usefulness of BrieflyLegal:

7.1 Multilingual and Multijurisdictional Support

Different jurisdictions and languages have different legal systems. BrieflyLegal will use multilingual pre-training methods and refine models on non-English legal datasets to guarantee wider applicability [35]. This will enable the platform to evaluate documents from various geographical locations and offer insights that are pertinent to both culture and the law.

7.2 Advanced Explainability and Transparency

Because AI models frequently function as "black boxes," it might be challenging for users to understand the results. In order to improve user confidence and transparency in AI-driven legal analysis, BrieflyLegal will incorporate explainability capabilities based on SHAP (SHapley Additive explanations) and LIME (Local Interpretable Model-agnostic Explanations) [36].

7.3 Enhanced Advocate Recommendation System

Reinforcement learning techniques can be used to improve the Advocate Recommendation Portal by improving advocate proposals in response to user feedback. Over time, this improvement will guarantee more congruence between customers' legal needs and advocate knowledge.

7.4 Real-Time Compliance Monitoring

BrieflyLegal will incorporate dynamic knowledge graphs that automatically update with legal developments in order to stay up to speed with quickly changing legislation. By doing this, compliance monitoring will continue to be precise and current.

7.5 Voice-Enabled Features

Future versions of BrieflyLegal will have voice-enabled capabilities to increase accessibility. By enabling users to communicate with the platform in plain language, this feature will make document analysis and advocate recommendations even easier.

8 Conclusion

Due to the inherent complexity of the legal area, documents frequently contain sophisticated terminology, thick language, and important regulatory intricacies. Because of these features, analysing legal documents takes a lot of time and is prone to mistakes for both experts and novices. BrieflyLegal was designed and created to use state-of-the-art natural language processing (NLP) and artificial intelligence (AI) methods to overcome these obstacles.

The platform provides a full range of tools that automate important parts of processing legal documents. BrieflyLegal streamlines the formerly time-consuming process of legal analysis with capabilities including clause identification, sentiment analysis, paraphrase, and real-time compliance monitoring. The platform guarantees accuracy and versatility across a variety of legal systems by utilising models such as Legal-BERT and CUAD-trained clause detectors. By giving users a better grasp of the meaning and tone of legal phrases, the sentiment analysis module helps users evaluate risks more skilfully. Furthermore, paraphrase tools convert difficult legalese into understandable language, enabling non-experts to interact with legal papers with assurance.

BrieflyLegal's Advocate Recommendation Portal is one of its unique features. This invention connects consumers with qualified legal professionals, bridging the gap between document analysis and actionable consequences. The portal provides a smooth transition from analysis to consultation by ranking advocates according to document insights, user preferences, and case requirements. Despite its advantages, BrieflyLegal's development has presented difficulties. The computational needs of processing long documents, the lack of multilingual and jurisdiction-specific annotated datasets, and the requirement to guarantee moral and open AI operations are a few of these. In order to overcome these obstacles, ongoing innovation is needed, such as the development of region-specific datasets, adaptive learning strategies, and sophisticated explainability tools.

The platform's potential is demonstrated by the outcomes thus far. BrieflyLegal has demonstrated excellent tone analysis, user-friendly paraphrase, and high clause detection accuracy. A crucial necessity in the rapidly shifting legal landscape is that papers stay in line with new legislation, which is ensured by integrating real-time compliance monitoring.

BrieflyLegal hopes to increase its skills in the future. The platform will become more inclusive and effective with improvements like voice-activated capabilities, interactive user interfaces, and language support. Furthermore, improving the advocate referral system will facilitate closer communication between clients and solicitors, promoting openness and confidence.

BrieflyLegal has the ability to completely change how people see, evaluate, and respond to legal documents by fusing state-of-the-art technology with a user-centred design. The website helps people understand the intricacies of legal terminology and procedures while also relieving the strain on legal experts. As it develops further, BrieflyLegal hopes to establish itself as a pillar of legal innovation, bridging the gap between legal knowledge and usability for users in a variety of settings.

References

[1] A. V. Zadgaonkar and A. J. Agrawal, “An overview of information extraction techniques for legal document analysis and processing,” *Int. J. Electr. Comput. Eng.*, vol. 11, no. 6, pp. 5450–5457, 2021, doi: 10.11591/ijece.v11i6.pp5450-5457.

[2] D. Mochihashi, “Natural Language Processing in Robotics,” *J. Robot. Soc. Japan*, vol. 39, no. 5, pp. 399–404, 2021, doi: 10.7210/jrsj.39.399.

[3] T. Xiao and J. Zhu, “Introduction to Transformers: an NLP Perspective,” 2023. [Online]. Available: http://arxiv.org/abs/2311.17633.

[4] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, “SQuAD: 100,000+ questions for machine comprehension of text,” *EMNLP 2016 - Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 2383–2392, 2016, doi: 10.18653/v1/d16-1264.

[5] H. Hendrycks et al., “CUAD: An expert-annotated NLP dataset for legal contract review,” *Proc. NeurIPS*, 2021.

[6] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proc. NAACL-HLT*, pp. 4171–4186, 2019.

[7] C. Raffel et al., "Exploring the limits of transfer learning with a unified text-to-text transformer," *J. Mach. Learn. Res.*, vol. 21, pp. 1–67, 2020.

[8] S. Gururangan et al., "Don't stop pretraining: Adapt language models to domains and tasks," *Proc. ACL*, pp. 8342–8356, 2020.

[9] B. Min et al., "Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey," *Proc. ACL*, 2021.

[10] U. Mushtaq and J. Cabessa, "Argument Mining with Fine-tuned Large Language Models," *Proc. COLING*, 2025.

[11] Y. Ye et al., "Stochastic Ensemble with Randomized Smoothing for NLP Robustness," *Proc. ACL*, 2023.

[12] T. Xiao et al., "Introduction to Legal Tech Applications using AI," *Proc. IJCAI*, 2024.

[13] M. Omar et al., "Robust Natural Language Processing: Recent Advances and Future Directions," *IEEE Access*, vol. 29, pp. 1–22, 2023.

[14] A. Nori et al., "Fine-tuning LLMs for Legal Text Analysis: A Comparative Study," *Proc. AAAI*, 2024.

[15] C. Clark et al., "Evaluation of LLM Robustness using Multi-Domain Datasets," *Proc. NeurIPS*, 2023.

[16] Z. Liu et al., "FLASK: Fine-grained Language Model Evaluation Based on Alignment Skill Sets," *Proc. ICLR*, 2024.

[17] F. Megahed et al., "Introducing ChatSQC: Enhancing Statistical Quality Control with Augmented AI," 2024. [Online]. Available: https://arxiv.org/abs/2308.13550.

[18] M. Kuzlu et al., "A Streamlit-based Artificial Intelligence Trust Platform for Next-Generation Wireless Networks," 2023.

[19] S. Ye et al., "Evaluation of Large Language Models for Skill Set Alignment," *Proc. NeurIPS*, 2023.

[20] Z. Liu et al., "Data Augmentation for NLP: Techniques and Challenges," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 912–926, 2024.

[21] D. Mohaisen et al., "NLP Robustness and Security: A Comprehensive Survey," *IEEE Access*, vol. 29, pp. 1–30, 2024.

[22] M. Chen et al., "Exploring Robustness in Legal NLP Systems," *Proc. EMNLP*, 2023.

[23] Y. Wu et al., "Enhanced Evaluation Metrics for Legal AI Models," *Proc. ACL*, 2024.

[24] S. Basile and N. Novielli, "Sentiment Analysis for Legal Clauses: Advances and Challenges," *Proc. COLING*, 2025.

[25] J. Zheng et al., "Training Language Models for Region-Specific Legal Analysis," *Proc. AAAI*, 2024.

[26] M. Maher et al., "Bridging Cognitive and Generative Models for Legal AI Applications," *Proc. ACL*, 2024.

[27] S. Korbak et al., "Enhancing Model Alignment for Legal Texts," *Proc. NeurIPS*, 2023.

[28] F. Ozgur Catak et al., "Adaptive Models for Legal Text Generation," *Proc. AAAI*, 2025.

[29] J. Irani et al., "WIBA: Weighted Argument Mining for Improved Legal Clause Classification," *Proc. EMNLP*, 2024.

[30] T. Wang et al., "Improving Text Augmentation for Multilingual Legal Applications," *Proc. IJCAI*, 2025.

[31] M. Jayanthi Kannan et al., "SwiftML: Streamlit-Powered Interface for Rapid Model Deployment in AI Systems," 2024.

[32] H. Zhou et al., "Large Model-Driven Data Augmentation for Text and Image Understanding," *Proc. NeurIPS*, 2024.

[33] N. Mishra et al., "LatticeML: A Dynamic Model Deployment Framework using Streamlit," *IEEE Access*, 2024.

[34] S. Chen et al., "Legal Text Analysis with Deep Neural Networks: Recent Progress and Future Directions," *Proc. ACL*, 2025.

[35] H. Nguyen et al., "Task-Adaptive Pretraining for Legal Text Understanding," *Proc. AAAI*, 2024.

[36] M. Gupta et al., "Explainable AI for Legal Argument Mining: A Review," *Proc. EMNLP*, 2024.

[37] B. Paneru et al., "Sentiment analysis of movie reviews: A flask application using CNN with RoBERTa embeddings," 2025 International Conference on Artificial Intelligence and Data Science (ICAIDS), 2025

[38] Liao, Xiaoxuan, et al. "A fine-tuning approach for T5 using knowledge graphs to address complex tasks." arXiv preprint arXiv:2502.16484 (2025).