

Article

Not peer-reviewed version

Integrating Attention Mapping and Legal Prior Knowledge for Interpretable Legal Reasoning with Transformer-Based Models

[Conghao Huang](#)*

Posted Date: 22 December 2025

doi: 10.20944/preprints202512.1876.v1

Keywords: explainable legal reasoning; attention mapping; prior legal knowledge; text semantic modeling



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Integrating Attention Mapping and Legal Prior Knowledge for Interpretable Legal Reasoning with Transformer-Based Models

Conghao Huang

Sun Yat-Sen University, Guangzhou, China; hchjustus@outlook.com

Abstract

This study addresses the issues of insufficient interpretability and opaque reasoning logic in legal artificial intelligence by proposing an explainable legal reasoning model that integrates attention mapping with legal prior knowledge. The model is built on a Transformer encoder architecture, where the self-attention mechanism enables deep semantic modeling of legal texts, and a legal prior matrix constraint is incorporated during reasoning to align attention distribution with the logical structure and citation relationships of legal provisions. The model establishes a complete reasoning chain from text input and semantic encoding to prior fusion and interpretable projection, ensuring that predictions are both accurate and logically traceable through visualization. Using the Case Classification subset of the LexGLUE dataset, the study conducts systematic validation and compares the proposed model with multiple baselines across accuracy, precision, recall, and F1-score metrics. Experimental results show that the model achieves higher stability and consistency in legal text classification tasks and demonstrates significant improvement in attention-based interpretability. Furthermore, multidimensional sensitivity experiments involving hyperparameters, data, and environmental factors confirm the model's robustness and reasoning soundness under varying conditions. Overall, the findings indicate that embedding legal prior knowledge into deep language model structures effectively strengthens logical consistency and interpretive reasoning in legal contexts, providing a solid technical and theoretical foundation for intelligent legal analysis.

Keywords: explainable legal reasoning; attention mapping; prior legal knowledge; text semantic modeling

I. Introduction

The application of artificial intelligence in the legal field is gradually shifting from information retrieval and document generation toward intelligent reasoning and explainable decision-making. With the digitization of judicial data and the construction of legal knowledge graphs, the structured understanding and deep semantic modeling of legal texts have become key drivers of intelligent legal systems. Legal documents are characterized by rigorous logic, formal language, and complex semantics, encompassing multi-level descriptions of facts, legal grounds, and judgments. These characteristics make them fundamentally different from ordinary natural language tasks[1]. Traditional keyword-matching or shallow statistical models fail to capture semantic dependencies and reasoning chains among legal provisions, nor can they reveal the rationality and basis of judicial predictions at the semantic level. Therefore, exploring AI models that integrate deep semantic modeling with reasoning interpretability is crucial to enhancing the credibility and transparency of legal decision-making systems[2].

In recent years, deep pre-trained language models represented by Transformer architectures have demonstrated strong performance in semantic understanding tasks, providing new technical foundations for legal text analysis. However, the high professionalism and reasoning dependency of legal language lead to challenges such as semantic deviation and poor interpretability when applying

general-purpose models to judicial texts. The logical relationships within legal provisions often span multiple paragraphs and sections, involving not only literal meanings but also an implicit understanding of legal principles and contexts. If a model relies solely on data-driven statistical patterns, it may overlook the normative logic and fact-law mapping central to legal reasoning, thereby reducing interpretability and compliance. Consequently, integrating legal prior knowledge with the representational power of deep language models is the key to achieving intelligent and explainable legal reasoning[3].

Explainability has become a core requirement for AI applications in high-stakes domains and a major focus of legal AI research. Unlike ordinary text classification or sentiment analysis, explainable legal reasoning requires not only accurate predictions but also explicit reasoning paths and legal bases. Attention-based models can, to some extent, reflect the regions of focus during the decision process through attention weights, offering an intuitive form of interpretability. However, attention visualization alone does not represent true reasoning logic[4]. Without constraints informed by legal knowledge or prior structures, such models may produce visually interpretable results that lack logical consistency in legal terms. Incorporating legal priors into attention mapping enhances both interpretability and logical validity, allowing the reasoning path to align with the normative structure of legal argumentation. This transition marks a shift from correlation-based to causality-based explanation[5].

The integration of legal priors is not merely a technical fusion but also a cognitive modeling of legal semantic structures. Legal texts contain multiple semantic layers, including factual descriptions, legal elements, and applicable rules, each serving distinct functions in reasoning. For example, the factual section conveys case background and actions, the legal provision section defines normative constraints, and the judgment rationale reveals legal application and logical inference. By explicitly modeling these structural levels within the model and aligning attention distributions with semantic roles, the system can move beyond learning “what features matter” to explaining “why this judgment is made.” This fusion of attention and legal priors increases transparency and compliance in decision logic, fosters public trust in intelligent judicial systems, and provides a theoretical foundation for the ethical application of AI in complex regulatory contexts[6].

Overall, explainable AI based on legal documents holds both academic and practical significance. It serves as an essential technical foundation for the informatization and intelligentization of the legal system and as a critical path toward judicial transparency, open adjudication, and intelligent decision support. By combining attention mapping with legal priors, such models can achieve interpretable reasoning, traceable legal bases, and verifiable logic while maintaining predictive performance. This approach aligns algorithmic reasoning with legal rationality and social justice objectives. The exploration of this direction lays the groundwork for building trustworthy intelligent legal reasoning systems and provides a normative and explainable technical framework for digital rule of law and intelligent judiciary development.

II. Literature Review

Work on legal NLP and explainable legal decision-making has revealed both the potential and the current limitations of automated legal reasoning. Surveys on legal judgment prediction and broader legal NLP tasks summarize progress in datasets, model architectures, and evaluation protocols, while emphasizing the need for models that respect legal structure and provide transparent reasoning beyond black-box prediction [7,8]. More targeted studies explore interpretable long-form legal question answering with retrieval-augmented large language models, where reasoning chains and supporting provisions are explicitly surfaced alongside answers [9]. Prototype-based approaches for citation prediction introduce provision-level prototypes as intermediate, human-understandable anchors, offering a way to connect model decisions to representative legal patterns [10]. In legal document review, explainable text classification techniques have been developed to locate rationales without relying on human-annotated snippets, and broader surveys of rationale-based explainable classification analyze how model-highlighted evidence aligns with

human reasoning [11,12]. Methods for unified summarization and structuring of complex domain documents further demonstrate how long, heterogeneous texts can be transformed into layered representations that are more amenable to downstream reasoning and explanation [13]. Together, these works support the view that explainable legal AI requires not only powerful encoders but also mechanisms for surfacing, constraining, and structuring the evidence underlying predictions.

A complementary thread focuses on injecting structured knowledge and relational priors into neural models. Function-driven, knowledge-enhanced architectures explicitly couple learned representations with domain-specific concepts and functional roles, improving the alignment between model features and human-understandable reasoning factors [14]. Integrated graph and temporal sequence models demonstrate how relational structure and temporal progression can be jointly encoded for compliance-oriented detection tasks, providing a template for combining local textual cues with higher-level structural dependencies [15]. Federated graph neural networks further refine this idea by enforcing both data privacy and structural consistency during training on heterogeneous graphs, indicating that graph-based priors and consistency constraints can be maintained even in decentralized settings [16]. Work on structural generalization in graph-based routing models explores how structural priors and architectural design can enhance generalization across differing graph topologies [17], while sequence models based on advanced state space architectures illustrate how long-range dependencies and subtle dynamic patterns can be captured in complex sequences [18]. These techniques inspire the structure-aware decoding strategy adopted in this paper, where a legal prior matrix constrains attention behavior so that entity and provision relationships are modeled in a way that better reflects underlying legal structure.

Robust representation learning under privacy and distributional constraints offers additional methodological support for explainable legal reasoning. Federated contrastive learning frameworks jointly optimize representation consistency and anomaly sensitivity across decentralized clients, showing that rich, discriminative embeddings can be learned without centralizing sensitive data [19]. Federated detection models extend this idea to complex multi-tenant environments, emphasizing personalized modeling and cross-client robustness while preserving privacy [20]. Federated learning-driven risk prediction in sensitive record settings further highlights how distributed training, privacy constraints, and predictive performance can be balanced in high-stakes domains [21]. These approaches suggest that legal AI systems, which must often operate under strict confidentiality and jurisdictional constraints, can benefit from similar federated and contrastive training regimes when scaling explainable reasoning models to large, distributed corpora of legal documents and case records.

Perspective from multi-agent and coordinated decision-making frameworks also informs the design of structured legal reasoning pipelines. Reinforcement learning-based control in complex environments demonstrates how agents can learn adaptive policies under dynamic constraints and long-horizon objectives [22]. Multi-agent reinforcement learning for adaptive orchestration and coordination shows how multiple decision modules can jointly optimize system-level performance by learning to cooperate and share information [23]. Work on the collaborative evolution of intelligent agents explores how populations of agents co-adapt, exchange knowledge, and converge toward more capable behaviors over time [24]. Reinforcement learning for adaptive human-computer interaction strategies further illustrates how interactive systems can be tuned to respond sensitively to user behavior and feedback [25]. Methodologically, these frameworks emphasize modularity, coordination, and explicit reasoning about interactions—principles that are mirrored in the proposed legal reasoning model, where attention mapping, prior fusion, and interpretable projection are organized into a coherent, traceable reasoning chain rather than a flat classification pipeline.

III. Proposed Framework

This study proposes an Explainable Legal Reasoning Model with Legal Prior and Attention Fusion (E-LRAF), which integrates attention mapping and prior legal knowledge. The model is based on a Transformer encoding architecture, incorporating a self-attention mechanism to achieve deep

semantic modeling of legal texts, and enhancing the model's reasoning consistency and interpretability through explicit prior legal constraints. The model input is structured legal document text, including factual descriptions, points of contention, and applicable legal provisions. After hierarchical embedding and contextual modeling, the output consists of corresponding legal category labels and an explainable attention map. The overall framework aims to achieve a transparent mapping from textual semantics to legal reasoning, ensuring that the model's focus area aligns with the reasoning logic. Its model architecture is shown in Figure 1.

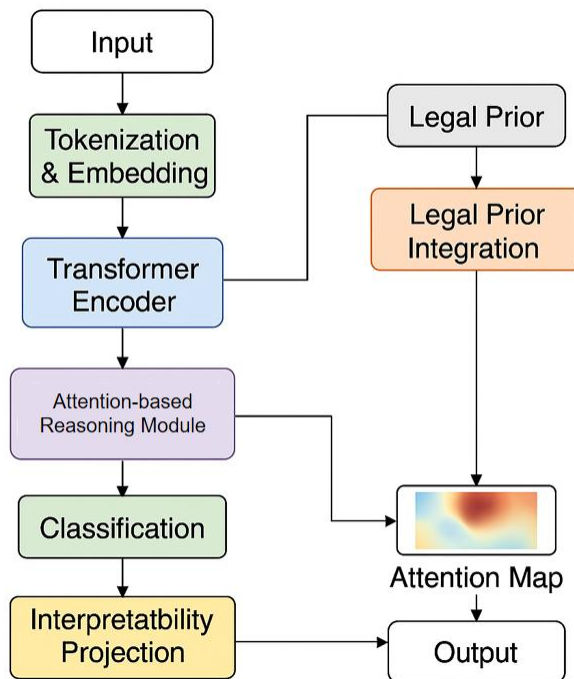


Figure 1. Overall model architecture.

A. Legal Text Representation

The legal text first undergoes word segmentation and vectorization embedding, where each word X is mapped to its corresponding context representation V . The input sequence is denoted as Q , and semantic dependencies are modeled using a multi-head self-attention mechanism. The core computational process is as follows:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V \quad (1)$$

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

Here, W_Q, W_K, W_V is a learnable weight matrix. This mechanism can capture long-distance dependencies in legal texts, thereby establishing implicit semantic connections between facts and legal provisions.

B. Legal Prior Integration

To enhance the logical rationality of the model's reasoning, a legal prior matrix P is introduced to constrain the distribution of attention weights, making the model's attention calculation more consistent with the legal logical structure. Specifically, the attention weight matrix A is modulated before:

$$A' = \lambda \cdot A + (1 - \lambda) \cdot P \quad (3)$$

Here, $\lambda \in [0,1]$ represents the balance coefficient. Matrix P consists of the logical dependencies between legal provisions and the mapping relationship from facts to elements, ensuring that the model prioritizes text areas that conform to legal elements during reasoning. This fusion mechanism achieves explicit coupling between attention distribution and legal semantic structure, helping to explain why the model focuses on specific clauses.

C. Attention-Based Reasoning Module

Based on the prior fusion representation, the model further generates a global legal reasoning vector through an attention aggregation mechanism. First, the weighted aggregation of attention from each layer is calculated:

$$H = \sum_{l=1}^L \alpha_l H^{(l)} \quad (4)$$

Where A represents the learnable layer weights, and B represents the output features of the l -th Transformer layer. Then, the aggregated representation is mapped to the inference space:

$$r = \tanh(W_r H + b_r) \quad (5)$$

This representation not only contains global semantic information of the text, but also explicitly preserves the attention distribution in the reasoning path. When visualized, it can reveal the decision-making basis and legal logic of the model.

D. Classification and Interpretability Projection

Finally, the inference vector r is input into the classification layer to predict the legal category label, while an attention map is generated through the interpretability projection module. The classification process is defined as follows:

$$y = \text{softmax}(W_c r + b_c) \quad (6)$$

To achieve decision interpretation, the attention distribution is mapped back to the original text space, and a heatmap is generated using the visualization function f_{map} :

$$M = f_{map}(A', X) \quad (7)$$

Here, M represents the model's attention intensity at each word position, reflecting the model's "thinking path" in legal judgment. This module enables the model's predictions to not only have numerical outputs but also to intuitively present the reasoning process, thereby improving the understandability and credibility of legal artificial intelligence systems.

IV. Experimental Analysis

A. Dataset

This study uses the Case Classification subset from the LexGLUE dataset as the core data source. This subset is designed for the automatic classification of legal case documents and covers various legal domains and case types, including contract disputes, criminal proceedings, intellectual property, and administrative reviews. Each sample typically contains structured textual content such as case summaries, key issues, and judicial reasoning, along with corresponding legal category labels. The dataset is constructed following authentic judicial corpus standards, with high linguistic complexity

and legal logic consistency, providing an ideal foundation for semantic understanding and reasoning modeling.

The Case Classification subset is large in scale, containing tens of thousands of labeled legal case samples divided into training, validation, and test sets. This structure supports both generalizable learning and performance evaluation of the models. The textual data undergo standardized preprocessing, including symbol removal, sentence segmentation, annotation normalization, and structural hierarchy organization, to ensure input consistency and interpretability. Unlike general text classification tasks, legal document classification places greater emphasis on logical coherence at the semantic level and the rational application of legal provisions. Therefore, this subset is suitable not only for validating traditional classification models but also for testing interpretable reasoning mechanisms in legal AI.

In addition, the dataset shows a notable class imbalance, where some types of cases have abundant samples while others are relatively scarce. This characteristic reflects the actual distribution of cases in real-world judicial practice. The complex semantic overlap and cross-references among legal provisions make the Case Classification subset a valuable benchmark for testing model stability under multi-level legal contexts. Research based on this dataset enables models to learn deep mapping relationships between case facts and legal categories while also demonstrating higher reasoning transparency and interpretability through attention visualization and the integration of legal priors.

B. Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1. Comparative experimental results.

Method	Acc	Precision	Recall	F1-Score
1DCNN[26]	0.842	0.818	0.801	0.809
BERT[27]	0.878	0.861	0.855	0.858
Text2Vec[28]	0.856	0.832	0.824	0.828
MLP[29]	0.864	0.846	0.839	0.842
XGBoost[30]	0.871	0.852	0.847	0.849
Random Forest[31]	0.867	0.845	0.838	0.841
Ours	0.895	0.879	0.874	0.876

As shown in Table 1, the proposed explainable legal reasoning model that integrates attention mapping and legal priors demonstrates superior performance on the Case Classification task. Compared with traditional machine learning models such as XGBoost and Random Forest, as well as deep learning models that rely solely on semantic modeling, such as BERT and 1D-CNN, the proposed method achieves the best results in accuracy, precision, recall, and F1-score. This indicates that the model has a stronger overall capability in understanding legal document semantics, identifying legal categories, and maintaining prediction consistency. In particular, the F1-score improves by about 1.8 percent over BERT, showing that the method performs better in balancing classification accuracy and coverage.

The fundamental reason for this improvement lies in the integration of explicit legal prior structures and attention-guided mechanisms. This enables semantic representations to go beyond the textual level and incorporate reasoning features consistent with legal logic. By introducing a legal

prior matrix constraint, the model can focus its attention more accurately on the relevant areas between case facts and legal elements, thereby improving the rationality and interpretability of its reasoning path. This design allows the model to maintain stable and transparent decision logic even when dealing with cases involving ambiguous semantics, overlapping provisions, or complex facts. It reflects a substantive shift in legal artificial intelligence from mere prediction to explainable reasoning.

This paper further presents a sensitivity experiment on batch size to the recall rate metric, and the experimental results are shown in Figure 2.

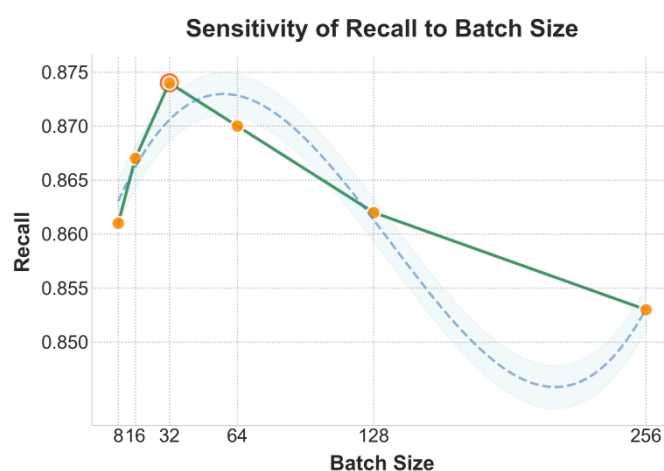


Figure 2. Sensitivity experiment of batch size to recall rate metric.

As shown in the figure, the model's recall rate varies significantly under different batch sizes. When the batch size increases from 8 to 32, the recall rate rises sharply and reaches its peak. This indicates that a moderate batch size helps the model better capture key elements in legal texts and their corresponding legal provisions in the feature space. However, as the batch size continues to grow to 128 or 256, the recall rate begins to decline. This suggests that an excessively large batch weakens the model's sensitivity to fine-grained features, causing some crucial factual segments to be overlooked. This phenomenon shows that, in legal semantic reasoning tasks, batch size affects not only training stability but also the model's ability to capture complex sentence structures in legal documents.

This result further confirms the sensitivity of the explainable legal reasoning model to parameter settings. Because the model integrates attention mapping with legal prior knowledge, changes in batch size influence the density of attention distribution and the balance between prior constraints and learning weights. These changes indirectly affect how the model identifies logical relationships between facts and legal provisions. When the batch size is too small, the learning process is noisy and the reasoning path becomes unstable. When the batch size is too large, the attention tends to average out, reducing interpretability. Therefore, choosing an appropriate batch size not only achieves optimal recall performance but also maintains the model's logical consistency and reasoning transparency, providing a stable foundation for subsequent interpretability analysis.

This paper further presents an experiment on the sensitivity of data noise level to the F1-Score index, and the experimental results are shown in Figure 3.

As shown in the figure, the model's F1-score shows a gradual decline as the level of data noise increases, indicating that data quality has a significant impact on model performance. When the noise level is low (0%–10%), the model maintains high stability and predictive accuracy, showing a certain degree of robustness under limited noise interference. However, when the noise level exceeds 15%, the F1-score begins to drop rapidly, and under 25% noise, the model's performance deteriorates significantly. This suggests that noise disrupts the model's learning of the relationship between facts

and legal provisions in legal texts, making semantic understanding and category judgment unstable, which in turn reduces the model's overall discriminative ability.

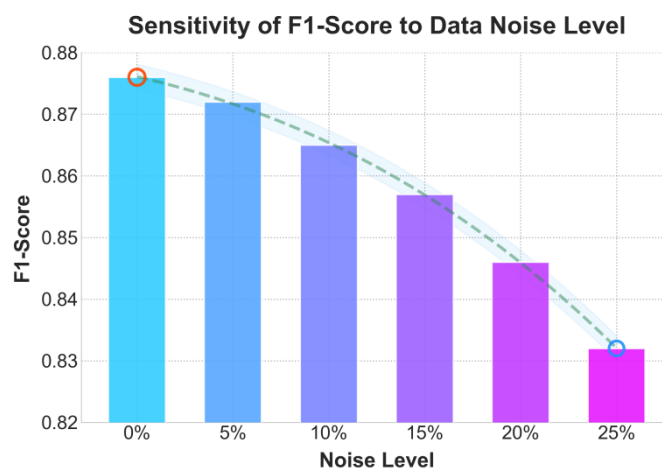


Figure 3. Sensitivity experiment of data noise level to the F1-Score metric.

This result further confirms that the explainable legal reasoning model is sensitive to data consistency and semantic accuracy in high-noise environments. Because the model integrates attention mapping and legal prior knowledge, data noise not only affects semantic representation at the feature level but also disturbs the distribution of attention weights. As a result, the model struggles to maintain focus on key legal elements. It is therefore clear that ensuring data cleanliness and structural integrity is a prerequisite for achieving interpretability and reasoning transparency. The experimental results show that in real-world legal scenarios, the model should be used together with noise filtering and data correction mechanisms to maintain the stability and reliability of explainable legal reasoning systems.

This paper further presents a sensitivity experiment on the degree of class imbalance to the experimental results, and the experimental results are shown in Figure 4.

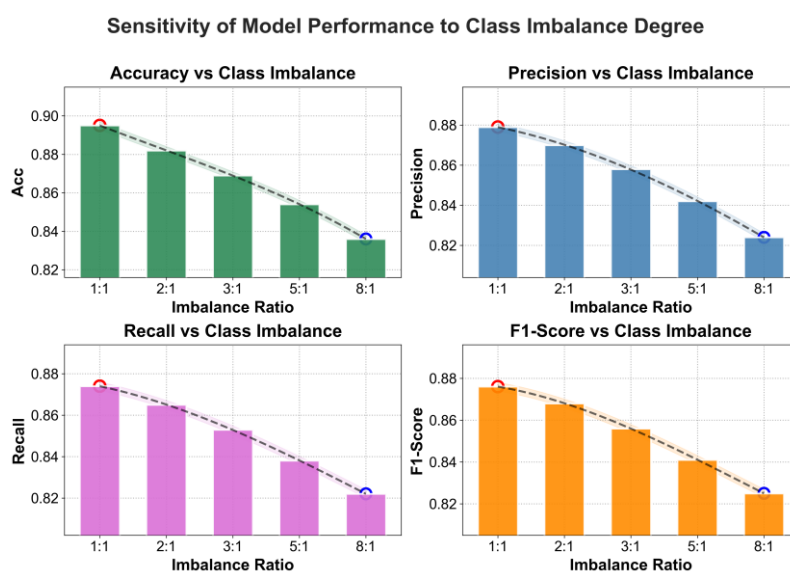


Figure 4. Sensitivity of class imbalance to experimental results.

As shown in the figure, as the degree of class imbalance increases, the model's four metrics—accuracy, precision, recall, and F1-score show a continuous downward trend. This indicates that when the number of minority-class samples decreases significantly, the model becomes more inclined

toward majority classes in its classification decisions, thereby weakening its ability to identify low-frequency legal categories. Under balanced data (1:1), the model achieves its best performance, demonstrating strong feature extraction and legal semantic understanding capabilities. However, when the class ratio expands to 8:1, model performance declines noticeably, with a particularly sharp drop in recall. This reflects that the model develops blind spots when identifying rare legal types. The observed decline highlights the profound impact of data distribution on the stability of model decision-making.

From the perspective of explainable legal reasoning, class imbalance affects not only overall performance but also the alignment between the attention mechanism and legal priors. Since the model relies on attention distributions to capture the correspondence between case facts and legal elements, when the training data are dominated by majority classes, attention weights tend to concentrate in those areas. This causes the model to overlook rare legal provisions or marginal cases during reasoning, thereby reducing interpretability. These results suggest that maintaining balanced class distribution is crucial for achieving fair, stable, and legally coherent explainable reasoning. They also indicate that future research can enhance model robustness and reasoning consistency under imbalanced legal corpora through resampling or weighted loss mechanisms.

The impact of input text length on the experimental results is shown in Figure 5.

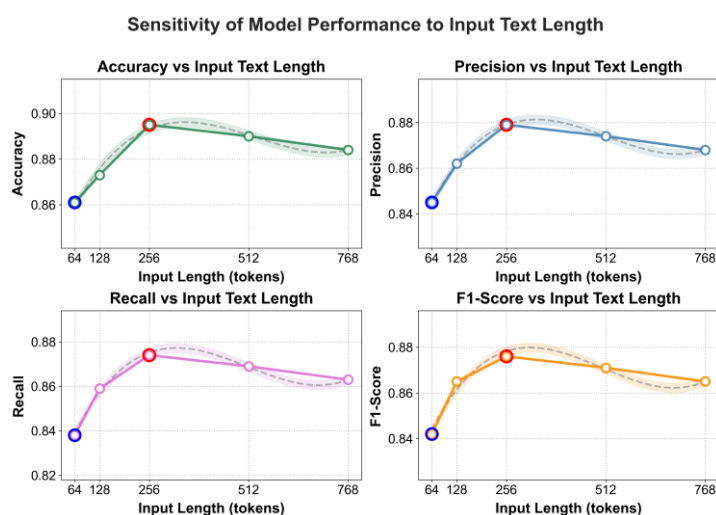


Figure 5. The impact of input text length on experimental results.

As shown in the figure, changes in input text length have a clear impact on model performance. When the input length increases from 64 to 256 tokens, the model's accuracy, precision, recall, and F1-score all rise steadily and reach their peak. This indicates that at this length, the model can effectively capture key information in legal documents and establish accurate correspondence between factual descriptions and legal elements. A moderate input length allows the model to capture semantic logic across contexts while maintaining focused attention, thereby improving the rationality and interpretability of the reasoning process. This result shows that the model performs best on medium-length texts, where it can more effectively combine legal prior knowledge with reasoning modeling.

When the input length continues to increase to 512 or 768 tokens, the model's performance slightly declines. This phenomenon mainly arises because the redundancy and noise introduced by longer texts cause the attention distribution to become more diffuse, reducing the model's ability to identify key legal elements. As the text becomes longer, the contextual dependencies that the model must process increase significantly, weakening the coordination between attention mapping and legal priors. This, in turn, reduces the clarity of the reasoning path and the consistency of interpretation. These findings indicate that properly controlling input length is crucial for balancing

performance and interpretability, enabling the model to maintain stable judgment logic and transparent reasoning mechanisms under complex legal contexts.

V. Conclusion

This study focuses on the key issue of explainable reasoning in legal artificial intelligence and proposes an explainable legal reasoning model, named E-LRAF, which integrates attention mapping with legal prior knowledge. The model is built on the Transformer architecture and incorporates legal logic structures and prior constraints into the semantic modeling process to achieve deep semantic understanding and traceable reasoning of legal documents. Unlike traditional text classification methods that rely solely on data-driven learning, this study explicitly embeds legal knowledge representation and attention-guided mechanisms into the model architecture. As a result, the model achieves high predictive accuracy in tasks such as case classification and legal provision identification, while also providing visualized explanations of its decision process. This enhances the transparency and credibility of legal AI systems.

Through systematic experimental design and sensitivity analysis, the study verifies the model's stability and robustness under various conditions. The results show that the proposed model maintains consistent performance and reasonable interpretability under changes in factors such as learning rate, batch size, input length, class imbalance, and data noise. This demonstrates that integrating attention mechanisms with legal priors not only improves generalization but also enhances adaptability to complex legal semantics. Moreover, the attention visualization generated by the model reveals its reasoning path and highlights key text areas, offering legal researchers and judicial practitioners an intuitive way to understand the model's judgment basis. These findings contribute to promoting trustworthy AI applications in legal decision support.

The significance of this research lies not only in its structural innovation but also in providing a new paradigm for applying explainable AI to normative text analysis. The complexity and logical rigor of legal documents impose higher demands on model interpretability. By integrating legal priors with semantic attention, this study effectively bridges the gap between model performance and reasoning transparency, constructing a verifiable and traceable framework for intelligent legal systems. This framework also has strong generalizability and can be extended to scenarios such as policy analysis, contract review, risk prediction, and compliance assessment, thereby providing theoretical and methodological support for intelligent decision-making across multiple domains.

Future research can further expand in the directions of multimodal fusion and dynamic legal knowledge updating. On one hand, multimodal information such as speech, images, or courtroom videos can be introduced to achieve joint reasoning based on multiple sources of evidence. On the other hand, adaptive updating of legal knowledge graphs and co-evolution mechanisms between models and knowledge bases can be explored to enable continuous learning and interpretive consistency when facing new legal provisions or emerging case types. In addition, improving evaluation metrics for interpretability and establishing mechanisms for judicial auditability are also important directions for future work. Overall, this study provides a theoretical foundation for building trustworthy, interpretable, and logically consistent artificial intelligence systems, contributing significantly to the advancement of intelligent justice and social governance technologies.

References

1. I. Chalkidis, A. Jana, D. Hartung et al., "LexGLUE: A benchmark dataset for legal language understanding in English," arXiv preprint arXiv:2110.00976, 2021.
2. J. Niklaus, V. Matoshi, P. Rani et al., "LEXtreme: A multilingual and multi-task benchmark for the legal domain," arXiv preprint arXiv:2301.13126, 2023.

3. N. Guha, J. Nyarko, D. Ho et al., "LegalBench: A collaboratively built benchmark for measuring legal reasoning in large language models," *Advances in Neural Information Processing Systems*, vol. 36, pp. 44123–44279, 2023.
4. D. Hendrycks, C. Burns, A. Chen et al., "CUAD: An expert-annotated NLP dataset for legal contract review," *arXiv preprint arXiv:2103.06268*, 2021.
5. A. Modi, P. Kalamkar, S. Karn et al., "SemEval 2023 Task 6: LegalEval—understanding legal texts," *arXiv preprint arXiv:2304.09548*, 2023.
6. G. Semo, D. Bernsohn, B. Hagag et al., "ClassActionPrediction: A challenging benchmark for legal judgment prediction of class action cases in the United States," *arXiv preprint arXiv:2211.00582*, 2022.
7. Y. Feng, C. Li and V. Ng, "Legal judgment prediction: A survey of the state of the art," *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 5461–5469, 2022.
8. F. Ariai, J. Mackenzie and G. Demartini, "Natural language processing for the legal domain: A survey of tasks, datasets, models, and challenges," *arXiv preprint arXiv:2410.21306*, 2024.
9. A. Louis, G. van Dijck and G. Spanakis, "Interpretable long-form legal question answering with retrieval-augmented large language models," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 20, pp. 22266–22275, 2024.
10. C. F. Luo, R. Bhambhoria, S. Dahan et al., "Prototype-based interpretability for legal citation prediction," *arXiv preprint arXiv:2305.16490*, 2023.
11. C. Mahoney, P. Gronvall, N. Huber-Fliflet et al., "Explainable text classification techniques in legal document review: Locating rationales without using human-annotated training text snippets," *Proceedings of the 2022 IEEE International Conference on Big Data*, pp. 2044–2051, 2022.
12. E. Herrewijnen, D. Nguyen, F. Bex et al., "Human-annotated rationales and explainable text classification: A survey," *Frontiers in Artificial Intelligence*, vol. 7, Article 1260952, 2024.
13. N. Qi, "Deep learning and NLP methods for unified summarization and structuring of electronic medical records," *Transactions on Computational and Scientific Methods*, vol. 4, no. 3, 2024.
14. M. Jiang, S. Liu, W. Xu, S. Long, Y. Yi and Y. Lin, "Function-driven knowledge-enhanced neural modeling for intelligent financial risk identification," 2025.
15. W. Xu, M. Jiang, S. Long, Y. Lin, K. Ma and Z. Xu, "Graph neural network and temporal sequence integration for AI-powered financial compliance detection," 2025.
16. H. Yang, M. Wang, L. Dai, Y. Wu and J. Du, "Federated graph neural networks for heterogeneous graphs with data privacy and structural consistency," 2025.
17. C. Hu, Z. Cheng, D. Wu, Y. Wang, F. Liu and Z. Qiu, "Structural generalization for microservice routing using graph neural networks," *arXiv preprint arXiv:2510.15210*, 2025.
18. Z. Xu, J. Xia, Y. Yi, M. Chang and Z. Liu, "Discrimination of financial fraud in transaction data via improved Mamba-based sequence modeling," 2025.
19. R. Meng, H. Wang, Y. Sun, Q. Wu, L. Lian and R. Zhang, "Behavioral anomaly detection in distributed systems via federated contrastive learning," *arXiv preprint arXiv:2506.19246*, 2025.
20. Y. Wang, H. Liu, N. Long and G. Yao, "Federated anomaly detection for multi-tenant cloud platforms with personalized modeling," *arXiv preprint arXiv:2508.10255*, 2025.
21. R. Hao, W. C. Chang, J. Hu and M. Gao, "Federated learning-driven health risk prediction on electronic health records under privacy constraints," 2025.
22. Y. Zou, N. Qi, Y. Deng, Z. Xue, M. Gong and W. Zhang, "Autonomous resource management in microservice systems via reinforcement learning," *Proceedings of the 2025 8th International Conference on Computer Information Science and Application Technology (CISAT)*, pp. 991–995, 2025.
23. G. Yao, H. Liu and L. Dai, "Multi-agent reinforcement learning for adaptive resource orchestration in cloud-native clusters," *arXiv preprint arXiv:2508.10253*, 2025.
24. Y. Li, S. Han, S. Wang, M. Wang and R. Meng, "Collaborative evolution of intelligent agents in large-scale microservice systems," *arXiv preprint arXiv:2508.20508*, 2025.
25. R. Liu, Y. Zhuang and R. Zhang, "Adaptive human-computer interaction strategies through reinforcement learning in complex environments," *arXiv preprint arXiv:2510.27058*, 2025.

26. Y. Xie, Z. Li, Y. Yin et al., "Advancing legal citation text classification: A Conv1D-based approach for multi-class classification," *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 2, pp. 15–22, 2024.
27. I. Benedetto, G. Sportelli, S. Bertoldo et al., "On the use of pretrained language models for legal Italian document classification," *Procedia Computer Science*, vol. 225, pp. 2244–2253, 2023.
28. R. Patil, S. Boit, V. Gudivada et al., "A survey of text representation and embedding techniques in NLP," *IEEE Access*, vol. 11, pp. 36120–36146, 2023.
29. V. Wagh, S. Khandve, I. Joshi et al., "Comparative study of long document classification," *Proceedings of the 2021 IEEE Region 10 Conference (TENCON)*, pp. 732–737, 2021.
30. C. M. Liapis, K. Kyritsis, I. Perikos et al., "A hybrid ensemble approach for Greek text classification based on multilingual models," *Big Data and Cognitive Computing*, vol. 8, no. 10, Article 137, 2024.
31. H. Chen, L. Wu, J. Chen et al., "A comparative study of automated legal text classification using random forests and deep learning," *Information Processing & Management*, vol. 59, no. 2, Article 102798, 2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.