Abstract

Focusing on the high incidence of labor dispute cases in contemporary judicial practice, this paper investigates the core challenge faced by Large Language Models (LLMs) in processing the corresponding adjudicative documents: the precise classification of the tripartite viewpoints of the "employee, employer, and court." Due to their unique tripartite structure, ambiguous and flexible legal terminology, and complex argumentative logic, labor dispute documents pose a formidable test for current natural language processing technologies. Existing models face three primary bottlenecks: (1) Domain Mismatch in Foundational Parsing, where general-purpose word segmentation tools fail to accurately process compound legal terms; (2) Failure to Model Discourse Logic and Legal Reasoning, with models struggling to capture the long-range "claim-defense-finding" argumentative chain; and (3) Severe Scarcity of High-Quality Annotated Data, where prohibitive costs and annotation complexity constrain model training.

To address these challenges, this paper proposes a systematic set of countermeasures. First, we advocate for a specialized text parsing system for the labor law domain to improve the accuracy of terminological recognition. Second, we innovatively introduce Graph Neural Networks (GNNs) to deconstruct adjudicative documents into an argumentation graph composed of "legal factors," thereby explicitly modeling discourse logic and the judicial syllogism. Finally, we propose a data strategy that combines semi-supervised and active learning to efficiently construct a large-scale, high-quality training corpus while controlling costs. This research aims to provide a viable technical path for LLMs to transition from general language understanding to specialized judicial reasoning, holding significant theoretical and practical importance for enhancing the precision and interpretability of judicial intelligence in the field of labor disputes.

Keywords: Labor Dispute; Tripartite Viewpoint Classification; Large Language Models (LLMs); Graph Neural Networks (GNNs); Discourse Logic; Legal Factors

1. Problem Background and Institutional Context

In contemporary judicial practice, the volume of labor dispute cases remains persistently high. In 2023 alone, mediation and arbitration bodies at all levels nationwide handled 3.85 million cases, involving 4.082 million workers and culminating in settlements and judgments valued at 82.99 billion RMB[1]. By 2024, these figures rose to 4.257 million cases, involving 4.549 million workers, with a total value of 93.47 billion RMB[2]. This translates to an average of over 11,000 new labor dispute cases filed each day, or approximately one new case every seven seconds. Such high frequency has made labor disputes a significant portion of the total caseload in grassroots courts, serving as a concentrated reflection of social conflict.

Structurally, judgments in labor disputes are largely consistent in form with general civil judgments, encompassing the plaintiff's claims, the defendant's response, the facts as ascertained, the reasoning for the judgment, and the final ruling. However, they differ substantially in substance. Ordinary civil cases emphasize the resolution of disputes between equal parties, primarily manifesting as a binary confrontation between plaintiff and defendant. In contrast, while labor dispute cases also present as adversarial proceedings between an employee and an employer, they are invariably accompanied by the institutional intervention of the government and its functional departments. For instance, matters such as the certification of work-related injuries, assessment of labor capacity, and social security contributions often require pre-emptive processing and determination by administrative bodies for human resources and social security or by work-related injury insurance agencies.z The conclusions of these bodies frequently have a direct bearing on the court's determination of facts and application of the law.

Consequently, a labor dispute judgment text effectively embodies a "tripartite" structure. The employee, often in a weaker position, presents claims tinged with strong emotional overtones. The employer relies on contractual clauses and internal regulations for its defense. The court, in responding to both parties, must also integrate the administrative determinations of government departments, social insurance systems, and labor protection policies. This interactive structure of "employee-employer-state" corresponds to the long-standing "tripartite coordination mechanism" in labor relations, wherein the government, trade unions, and employers' organizations jointly formulate and implement labor and social policies through communication and negotiation to ensure the stability of labor relations.[3] This institutional arrangement, which combines both judicial and administrative

---

[1] Ministry of Human Resources and Social Security of the PRC. 2023. *Statistical Communiqué on the Development of Human Resources and Social Security in 2023*. (In Chinese).

[2] Ministry of Human Resources and Social Security of the PRC. 2024. Statistical Communiqué on the Development of Human Resources and Social Security in 2024. (In Chinese).

[3] Tan, Qiuxia. 2013. "An Analysis of the Tripartite Coordination Mechanism for Labor Relations in China" [J]. Academic Exploration (10): 38-41.

characteristics, is the primary reason why labor dispute litigation is distinct from and relatively independent of general civil litigation.[4]

1.1. The Challenge of Ambiguity and the Need for Consistency

Labor law, as a quintessential form of protective legislation, has the fundamental mission of balancing the conflicts arising from the inherent inequality between employees and employers and safeguarding the legitimate rights and interests of workers[5]. The essence of the labor relationship is inequality; employees are often in a weaker economic and social position and thus require slanted protection under the law. For this reason, labor law is replete with ambiguous or flexible terms such as "serious violation of rules and regulations," "reasonable compensation," and "manifestly unfair," which are intended to be applied through judicial discretion in varying contexts.

Within adjudicative documents, these terms may appear in the employee's claims to highlight the infringement of their rights, in the employer's defense to justify a dismissal or other measures, or, most commonly, cited by the court in its reasoning as the basis for a normative evaluation. This overlapping use across multiple contexts often blurs the boundaries between the positions of the different parties. The ambiguity of these legal terms is often interpreted through a "factor model." Factors are key factual patterns abstracted from precedential cases that both define the application boundaries of vague legal concepts and provide a reference for reasoning in new cases[6]. As Russell also emphasized, an AI system must be able to reason effectively under uncertainty and ambiguity to cope with real-world complexity. Thus, the problem of ambiguity faced by legal AI is not an isolated phenomenon but a projection of a universal challenge for artificial intelligence within the legal domain[7]. Concurrently, ensuring "like cases are judged alike" is a crucial goal of institutional development. While case-based reasoning systems aim to achieve consistency and corrective functions by providing judges with similar cases, existing systems still suffer from deficiencies in precision, scope, and hierarchical differentiation, failing to fully support judicial justice[8].

---

[4] Zhou, Huyong, and Mao, Yong. 2013. "On the Theoretical Basis for the Relative Independence of Labor Litigation from Civil Litigation: A Substantive Law Perspective" [J]. Law Journal 34(05): 62-68.

[5] Wang, Quanxing (王全兴). 2017. Laodong Fa [Labor Law] (4th ed.). Beijing: Falü Chubanshe [Law Press]. (In Chinese).

[6] Ashley, K. D. 2017. Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge: Cambridge University Press.

[7] Russell, S., and Norvig, P. 2020. Artificial Intelligence: A Modern Approach (4th ed.). Pearson. (Note: This is the citation for the original English 4th edition, which is the standard practice for an English-language paper.)

[8] Zuo, Weimin (左卫民). 2018. "Ruhe tongguo rengong zhineng shixian lei'an leipan" [How to Achieve Like-Case-Like-Judgment Through Artificial Intelligence]. Zhongguo Falü Pinglun [China Law Review] (02): 26–32. (In Chinese).

1.2. The Foundational Role of Viewpoint Stratification

Against this backdrop, the primary task in advancing the application of large language models (LLMs) in the field of labor disputes is the accurate classification of the viewpoints of the different parties within adjudicative texts[9]. This step is not merely a technical pre-processing measure; it is a critical enabler for the effective functioning of legal AI in real-world judicial scenarios. Labor dispute cases are characterized by complex facts, intertwined points of contention, and lengthy texts, where the employee's claims, the employer's defense, and the court's reasoning are often interwoven and nested. Feeding this undifferentiated text into a model would inevitably blur the lines between factual statements and normative judgments. This would prevent the model from forming stable role representations during the learning phase, leading to insufficient accuracy in downstream reasoning and generation tasks.

Through stratification, the originally mixed information is decomposed into logically distinct units with clear boundaries. The employee's claims, the employer's defensive stance, and the court's legal application are separately labeled and categorized, resulting in a "purified" data corpus. During training, the model not only receives structured and stable input but also gradually develops an awareness of the distinctions between the different parties' perspectives. This lays a solid foundation for its subsequent performance in judicial reasoning and case analysis[10].

1.3.Enhancing Legal Reasoning and Interpretability

From the perspective of legal reasoning, viewpoint stratification holds even deeper value. The intrinsic logic of a judicial text lies in the "claim-defense-finding" reasoning chain, which is the fundamental structure upon which a judicial decision is built. Without role differentiation, a model can only operate at a narrative level, capturing scattered linguistic patterns without grasping the logical tension between the parties' positions, let alone reproducing the rational argumentation demonstrated by the court. After stratification, the model is explicitly exposed to the complete reasoning chain during the training phase and can replicate a similar logical structure during the generation phase. This allows its output to more closely reflect the syllogistic thinking characteristic of judicial professionals.

Viewpoint stratification is also directly related to the issue of interpretability. The unique nature of legal AI dictates that its output cannot be satisfactory on the basis of a conclusion alone. What truly determines its admissibility in judicial practice is whether the reasoning behind the conclusion is justifiable and can be traced back to

---

9 *Ajunwa, Ifeoma. (2021). An Auditing Imperative for Automated Hiring Systems. Harvard Journal of Law & Technology, 34(2), 381–446.*

10 Gray, M., Zhang, L., & Ashley, K. D. (2025). *Generating Case-Based Legal Arguments with LLMs*. In *Proceedings of the 2025 Symposium on Computer Science and Law (CSLAW '25)*, ACM, pp. 160–168

specific facts and legal grounds. By stratifying the three viewpoints, the model's response can clearly indicate that "this fact originates from the employee's claim," "this reason comes from the employer's defense," or "this normative judgment is from the court's finding," thereby forming a complete evidentiary chain. This traceability not only enhances the transparency of the output but also addresses the fundamental requirement of open and justifiable reasoning in judicial practice, significantly improving the applicability of LLMs within the legal system.

1.4. Managing Complexity in Voluminous Texts

The complexity of labor dispute documents—which can often run to tens of thousands of words and simultaneously involve multiple points of contention such as wage payments, work-related injury certification, contract termination, overtime pay, and social security contributions, encompassing both substantive and procedural issues—further highlights the importance of stratification. Unprocessed long-form texts can easily cause a model to lose its grasp of the overall context, leading to fractures in memory and logical coherence, ultimately affecting its comprehension and the stability of its output. Role-based classification effectively mitigates this challenge by segmenting vast and convoluted texts into ordered, logical units. This provides the model with learnable, structured samples, enabling it to maintain strong global comprehension and output capabilities even when faced with exceptionally long and complex corpora.

1.5. Conclusion

In summary, the stratification of tripartite viewpoints is not merely a text-processing step but a foundational requirement for the effective functioning of legal large models. It ensures the purity and clear demarcation of input data, provides training samples for judicial logic at the reasoning level, and establishes an interpretable evidentiary chain at the output level. This enables an LLM not only to accurately understand the facts and reasoning within labor dispute documents but also to progressively master the reasoning methods and argumentative styles of judicial decision-makers, thereby achieving the leap from a general-purpose language model to a specialized legal model.

2. Challenges and Bottlenecks in Tripartite Viewpoint Classification

In recent years, with the rapid advancement of natural language processing (NLP) technologies, the research paradigm for legal text analysis has comprehensively shifted from traditional symbolic and retrieval-based models to deep semantic modeling. Throughout this process, the technological trajectory has undergone a clear evolution from shallow to deep methods. Early explorations were primarily based on symbolic AI, employing manually crafted rulebases, keywords, and Boolean logic for text matching[11]. While effective for tasks like matching explicit legal articles, these methods were heavily reliant on expert knowledge for knowledge acquisition and struggled to handle the ambiguity and diversity of legal language, thus exhibiting clear limitations[12].

To overcome this bottleneck, statistical machine learning methods based on feature engineering emerged as the mainstream in the second phase. Researchers trained classifiers such as Support Vector Machines (SVM) by extracting features like Bag-of-Words, N-grams, and topic models to predict judgment outcomes or identify legal elements[13]. However, the performance of such methods was inherently capped by the quality of the handcrafted features, and their capacity for capturing deep semantics was insufficient for the nuanced expressions and complex logic prevalent in legal texts.

The introduction of deep learning brought about a fundamental turning point, with its core advantage being the automated learning of features. Models represented by Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) were rapidly applied to legal tasks such as judgment and charge prediction, achieving results that surpassed traditional methods. Nevertheless, these early deep models had their own structural deficiencies: RNNs faced the problem of long-term dependencies [14]due to vanishing gradients when processing long texts, while CNNs were more focused on capturing local patterns, with a weaker ability to perceive global logical structures.

The true catalyst for advancing legal text processing to a new stage was the Transformer architecture and its core self-attention mechanism[15]. Pre-trained language models (PLMs) like BERT, pre-trained on massive general-purpose corpora, acquired powerful bidirectional contextual representation capabilities, significantly raising the baseline for semantic understanding. To adapt to domain specificity, academia and industry successively released domain-specific models pre-trained on specialized legal corpora, such as Legal-BERT[16], which demonstrated marked superiority over general models in understanding legal terminology and capturing logical relationships. More recentlyz, the advent of large-scale generative language models (LLMs) like GPT-3 has, by virtue of their immense scale and powerful few-shot learning capabilities, opened upDespite the rapid progress in NLP, the capability boundaries of

---

[11] *Gardner, A. v. d. L. (1987). An Artificial Intelligence Approach to Legal Reasoning. MIT Press.*
[12] *Ashley, K. D. (2017). Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge University Press.*
[13] *Aletras, N., et al. (2016). Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective. PeerJ Computer Science, 2, e93.*
[14] *Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780. Vaswani, A., et al. (2017). Attention Is All You Need. Advances in NeurIPS, 30*
[15] *Devlin, J., et al. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.*
[16] *Chalkidis, I., et al. (2020). LEGAL-BERT: The Muppets straight out of Law School. In Findings of EMNLP 2020.*

existing models remain evident when confronted with specific, complex tasks. In the task of tripartite viewpoint classification within labor dispute documents, the performance of current models is still far from satisfactory.

2.1. Domain Mismatch and Information Distortion in Foundational Parsing

All advanced NLP tasks are built upon precise foundational text parsing, a sensitivity that is particularly acute for the tripartite viewpoint classification task. However, mainstream Chinese parsing tools—such as the Language Technology Platform (LTP) and Jieba, which are widely used in general-purpose Chinese NLP—exhibit significant limitations when faced with highly specialized legal texts[17]. Their direct application to the highly professionalized and stylized language of labor dispute documents results in a severe "domain mismatch," causing information distortion from the very source. The core mechanism of mainstream Chinese word segmentation is typically based on a maximum probability path search using a prefix dictionary, supplemented by statistical models like Hidden Markov Models (HMM) or Conditional Random Fields (CRF) to handle "out-of-vocabulary" (OOV) words[18].

Under this technical framework, a prominent challenge is the incorrect segmentation of compound legal terms. Labor dispute texts are replete with terms composed of multiple common words that carry a specific meaning in a legal context, such as 劳动能力鉴定结论 (conclusion of labor capacity assessment), 无固定期限劳动合同 (open-ended employment contract), and 劳务派遣单位 (labor dispatch agency). Lacking these specialized terms in their dictionaries, general-purpose segmentation models often mechanically break them down into constituent parts like 劳动能力 (labor capacity), 鉴定 (assessment), and 结论 (conclusion). This "semantic fragmentation" directly undermines the integrity of the legal concept, preventing subsequent models from understanding and modeling it as a unified legal element, thereby leading to the loss of critical information.

Secondly, the domain-specificity of word meaning and contextual ambiguity pose a significant challenge. Legal language is characterized by high precision, and many common words assume entirely different, specific meanings in a legal context. For example, the character 审 (shěn) is often a prefix in general language (e.g., 审稿,

---

[17] *Yan, Qian (严倩). 2018. "Mianxiang falü wenshu de zhongwen fenci fangfa yanjiu" [Research on Chinese Word Segmentation Method for Legal Documents]. Dissertation, Suzhou Daxue [Suzhou University]. (In Chinese).*
[18] *Xing, Ling (邢玲), and Cheng, Bing (程兵). 2023. "Jiyu jieba fenci de lingyu zishiying fenci fangfa yanjiu" [Research on Domain Adaptive Word Segmentation Method Based on Jieba]. Jisuanji Fangzhen [Computer Simulation] 40(04): 310–316, 503. (In Chinese).*

to review a manuscript), but in legal documents, it is frequently a suffix (e.g., 一审, first instance; 二审, second instance). As general-purpose segmentation models are primarily trained on corpora from news and encyclopedias[19], they lack specialized training on legal texts, particularly those concerning labor disputes, and thus perform poorly on professional expressions like 工伤认定 (work-related injury certification) and 社会保险 (social insurance). This lack of domain knowledge readily leads to errors in part-of-speech tagging and boundary delineation. For a word like 解除 (termination/rescission), the agent of the action (employer or employee) is key to determining viewpoint attribution, but a context-insensitive segmentation tool cannot capture this distinction at the foundational parsing level.

Finally, the failure to recognize new and OOV words is particularly acute. The legal field, especially labor law, constantly generates new concepts and expressions in response to socioeconomic development. These neologisms are not covered by pre-constructed static dictionaries and become OOV words. While models attempt to guess using algorithms like HMM, their performance is often poor for professional terms longer than two characters[20]. This systematic bias at the foundational parsing stage means that the data fed into subsequent classification models is flawed and distorted from the outset. No matter how powerful the model, it cannot perform accurate viewpoint attribution and stance detection on an erroneous foundation.

2.2. Failure to Model Discourse Logic and Legal Reasoning

A fundamental bottleneck in the processing of labor dispute documents by existing models is their inability to effectively model the overall structure and internal logic of the text. This failure is deeply rooted in their deficiencies in handling discourse-level argumentative structures, the legal syllogism, and ultra-long texts, leading to poor performance in the tripartite viewpoint classification task. Labor dispute documents possess a highly structured argumentative paradigm, centered on the logical chain of "claim-defense-finding," where viewpoints are often interwoven and nested, forming a complex dialogical network. The design of most current NLP models is confined to processing local semantics at the sentence or paragraph level, lacking the ability to capture these long-range argumentative relationships that span multiple paragraphs. They fail to understand the role a specific sentence plays within the entire argumentative structure—a challenge known in academia as "argumentation mining" and recognized as a formidable task.

This is especially evident in the "This Court finds" (本院认为) section, where, to provide thorough reasoning, the court often first fully restates the plaintiff's claims

---

[19] *Liu, Ting (刘挺), Che, Wanxiang (车万翔), and Li, Zhenghua (李正华). 2011. "Yuyan jishu pingtai" [Language Technology Platform]. Zhongwen Xinxi Xuebao [Journal of Chinese Information Processing] 25(06): 53–62. (In Chinese).*

[20] *Bai, Fengbo (白凤波); Chang, Lin (常林); Wang, Shifan (王世凡); et al. 2020. "Cai pan wenshu guanjianci tiqu de gaijin fangfa yanjiu" [Research on an Improved Method for Keyword Extraction from Judicial Documents]. Jisuanji Gongcheng yu Yingyong [Computer Engineering and Applications] 56(23): 153–160. (In Chinese).*

and evidence, then cites the defendant's defense, and finally makes its judgment and findings based on this foundation. A human reader can easily distinguish between the court's direct opinions and the viewpoints of others cited for argumentation. However, for a model lacking discourse-level comprehension, these are merely adjacent text fragments. The model struggles to identify speech acts such as citation, refutation, or confirmation, and thus frequently misattributes the viewpoints of other parties appearing in the same paragraph to the court, causing systematic classification errors[21].

This incapacity is particularly pronounced when judicial decisions follow the syllogistic reasoning of "minor premise (fact) – major premise (rule) – conclusion." The core training objective of general-purpose pre-trained models is linguistic probability optimization, not logical inference. This reflects a fundamental mismatch between current statistical-based NLP methods and the symbolic reasoning required in the legal domain[22]. By learning statistical patterns in massive texts to predict the next word or a masked word, these models excel at capturing surface-level linguistic regularities but cannot truly understand or reconstruct the rigorous reasoning chain—from fact to norm to conclusion—upon which judicial decisions depend[23]. A model might identify the fact that "the plaintiff did not sign a labor contract" and the legal norm that "the Labor Contract Law requires a written labor contract," but it cannot establish the logical entailment between the two—that the former is a precondition for the application of the latter. Consequently, when performing viewpoint classification, the model can often only conduct fuzzy matching based on keywords or surface semantics. This results in classification outcomes that are not only lacking in stability and robustness but, more importantly, are devoid of interpretability. In a field like law, which demands high transparency and logical rigor, a "black box" model unable to clearly display its reasoning path will have its applied value and credibility severely diminished.

These two problems are further amplified when dealing with long texts and become intertwined with the technical limitations of the models themselves, forming an almost insurmountable barrier. The length of labor dispute documents often reaches thousands or even tens of thousands of characters, far exceeding the input length limit of mainstream models like BERT (approximately 512 tokens). The very emergence of long-text models developed by the academic community (e.g., Longformer) attests to the severe limitations of the standard Transformer architecture in handling long documents[24]. While splitting a long text into multiple chunks for

---

[21] *An, Zhenwei (安震威); Lai, Yuxuan (来雨轩); and Feng, Yansong (冯岩松). 2022. "Mianxiang falü wenshu de ziran yuyan lijie" [Natural Language Understanding for Legal Documents]. Zhongwen Xinxi Xuebao [Journal of Chinese Information Processing] 36(8): 1–11. (In Chinese).*

[22] Ashley, K. D. (2017). Artificial intelligence and legal analytics: new tools for law practice in the digital age. Cambridge University Press.

[23] Li, Jinchen (李瑾晨); Li, Yanling (李艳玲); Ge, Fengpei (葛凤培); et al. 2023. "Mianxiang falü lingyu de zhineng xitong yanjiu zongshu" [A Survey of Intelligent Systems for the Legal Domain]. Journal of Computer Engineering & Applications 59(7). (In Chinese).

[24] Beltagy, I., Peters, M. E., & Cohan, A. (2020). Longformer: The long-document transformer. arXiv preprint arXiv:2004.05150.

independent processing solves the input length problem, it comes at the cost of sacrificing global information, forcibly severing the intrinsic logical connections between paragraphs. A core claim made by the plaintiff at the beginning of a document might only be addressed by the court pages later at the end. Such long-distance viewpoint dependencies are easily lost in fragmented processing. A model might see a claim in one window and a finding in another, but due to the lack of a global perspective, it cannot link the two. This remains a stubborn bottleneck even for models specifically pre-trained on legal corpora (e.g., LEGAL-BERT), leading to a fragmented grasp of the overall logic and, ultimately, the failure of viewpoint classification[25].

2.3. Severe Scarcity of High-Quality Annotated Data

As a typical supervised learning task, the performance of tripartite viewpoint classification is highly dependent on large-scale, high-quality annotated corpora. However, in the field of labor disputes, acquiring such data faces a nearly irreconcilable dual dilemma, which not only constrains the training effectiveness of existing models but also fundamentally restricts the exploration of technological pathways.

First, the cost of annotation is prohibitively high, preventing the formation of large-scale corpora. Labor dispute documents are typically lengthy, with complex case facts and interlocking logical chains. To perform fine-grained stance attribution annotation on any single document requires a significant investment of time for deep reading and analysis. This is fundamentally different from annotation tasks in general domains, where judgments can be made quickly through simple reading. The former requires the annotator not only to understand the literal meaning but also to sort out the complete argumentative relationships between the parties and between the parties and the court. More critically, this annotation work cannot be accomplished through simple crowdsourcing; it must rely on costly professionals with a deep legal background (such as law graduate students, lawyers, or judicial assistants) to ensure annotation accuracy. This makes the economic cost of building a large-scale legal annotated corpus prohibitive, causing research and development to be confined to small-scale datasets that cannot satisfy the "feeding" demands of deep learning models for massive data. This reflects a major challenge in the data construction for judicial AI today: although a vast number of legal documents are publicly available, this raw data is merely "raw material" that requires extensive cleaning and annotation, far from being a "finished product" ready for use. A huge gap exists between massive raw data and high-quality datasets suitable for model training[26]. Consequently, research in legal AI currently faces a universal predicament of data scarcity, which

[25] Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. arXiv preprint arXiv:2010.02559.

[26] Zuo, Weimin (左卫民). 2020. "Cong tongyonghua zouxiang zhuanyehua: Fansi Zhongguo sifa rengong zhineng de yun yong" [From Generalization to Specialization: A Reflection on the Application of Judicial Artificial Intelligence in China]. Faxue Luntan [Legal Forum] 35(2): 17–23. (In Chinese).

directly leads to insufficient model training and an inability to learn the multifaceted and complex language patterns of labor dispute scenarios, ultimately compromising the model's generalization ability and its effectiveness in real-world applications.

Second, the ambiguity and complexity of legal text lead to low inter-annotator agreement. Even if cost issues could be overcome, the intrinsic complexity of legal text creates a formidable obstacle to obtaining "high-quality" data. Due to the intricacy of case facts and the subtlety of legal language, achieving complete consensus among different annotators is exceptionally difficult. This is particularly true in ambiguous contexts where the court restates a party's opinion—for instance, "The defendant argues that its actions were based on company regulations..." Different annotators may subjectively disagree on whether this sentence should be attributed to the "defendant's viewpoint" or as "the defendant's claim cited by the court during its summary of case facts." In addition to stance attribution, there are also disputes over the delineation of viewpoint boundaries. A complete argument may span multiple sentences or even paragraphs, interspersed with descriptions of evidence and comments on facts. Precisely defining boundaries in such texts makes it extremely difficult for different annotators to reach a uniform standard. This problem of "low inter-annotator agreement" results in training data that is filled with a significant amount of noise and inconsistency. For a machine learning model, this means it receives numerous contradictory signals during training—for example, the same or similar sentences being labeled with different stances in the training set. This severely disrupts the learning process, making it difficult for the model to converge to a stable state. The resulting trained model not only performs poorly but also behaves erratically and irreproducibly when faced with new, unseen documents. This predicament is fundamentally rooted in the complexity of law itself and the ambiguity of human language, which technological means cannot entirely overcome[27]. Furthermore, legal practice itself involves open and contestable activities such as legal interpretation and factual discretion. The attempt to convert it entirely into precise, uncontroversial annotated data may itself overlook the intrinsic characteristics of legal practice[28]. Therefore, the data problem has become one of the core bottlenecks constraining breakthrough progress in the task of tripartite viewpoint classification in labor law.

3. Countermeasures and Institutional Implications

3.1. A Specialized Path for Text Parsing in the Labor Law Domain

---

[27] Zuo, Weimin (左卫民), and Pan, Xin (潘鑫). 2023. "Tongguo jishu guixun sifa: Jinbu yu tiaozhan" [Disciplining the Judiciary Through Technology: Progress and Challenges]. Faxue Pinglun [Law Review] 41(04): 57–65. (In Chinese).

[28] Zuo, Weimin (左卫民). 2021. "AI faguan de shidai hui daolai ma: Jiyu zhongwai sifa rengong zhineng de duibi yu zhanwang" [Will the Era of AI Judges Arrive? A Comparison and Outlook Based on Judicial Artificial Intelligence in China and Abroad]. Zhengfa Luntan [Tribune of Political Science and Law] 39(5): 3–13. (In Chinese).

In the task of tripartite viewpoint classification for labor dispute documents, a solid foundation for subsequent modeling can only be established by constructing a high-quality, domain-specific parsing system at the input stage. Existing general-purpose word segmentation and syntactic parsing tools are predominantly derived from open corpora such as news and encyclopedias, leading to a natural divergence in parsing standards and lexical coverage when applied to the legal domain. Therefore, the direction for adaptation should be to deeply specialize these tools, enabling them not only to recognize exclusive labor law terminology but also to maintain logical consistency at the discourse level, thereby minimizing semantic noise and information loss.

The first step in this adaptation is to construct a specialized dictionary for the labor law domain. Diverging from manual reliance on expert experience, terms can be automatically mined from large-scale corpora using unsupervised methods. Metrics such as Pointwise Mutual Information (PMI) or the improved Normalized Pointwise Mutual Information (NPMI) can be used to measure word cohesion, while left and right information entropy can assess the stability of candidate phrases in context[29]. This process allows for the screening of high-quality terms like 劳动能力鉴定结论 (conclusion of labor capacity assessment) and 无固定期限劳动合同 (open-ended employment contract). Once these terms are consolidated as integral units, they are not only protected from being mechanically fragmented but can also serve as distinct legal elements in downstream tasks, enhancing the sense of semantic boundaries during viewpoint classification[30].

On the other hand, dictionary expansion alone is insufficient to handle complex legal expressions; the parsing framework must integrate multi-feature information. For instance, part-of-speech tagging can help differentiate the semantic roles of "termination" in different contexts, while dependency parsing can reveal the subordinate relationship between the verb and object in "terminate the labor contract," thereby determining the agent and the applicable legal effect. Through multi-feature fusion, the model no longer relies on a single vocabulary for segmentation but is equipped with the ability to robustly identify key legal concepts in complex contexts. To further enhance the model's robustness, a joint learning mechanism can be introduced, training it on data from labor dispute documents alongside data from adjacent fields like social insurance and civil contracts. By learning general expressions in a shared layer and capturing the unique, fine-grained features of labor law in a specialized layer, this cross-domain transfer learning not only alleviates the

---

[29] PMI (Pointwise Mutual Information) measures the difference between the probability of two words co-occurring and the probability of them occurring independently. A higher value indicates a stronger association between the words. However, PMI is sensitive to low-frequency words, where an accidental co-occurrence can result in a high score. NPMI (Normalized PMI) builds on PMI by normalizing the score to a range of [-1, 1]. This normalization addresses the issue of incomparability between different word pairs and suppresses noise from low-frequency events, making it more suitable for automatically discovering stable, domain-specific terms in large-scale corpora.

[30] Bommarito M J, Katz D M, Bommarito J. KL3M Tokenizers: A Family of Domain-Specific and Character-Level Tokenizers for Legal, Financial, and Preprocessing Applications[J]. arXiv preprint arXiv:2503.17247, 2025. https://arxiv.org/abs/2503.17247

problem of insufficient labor law corpora but also enhances the model's adaptability to new types of cases.[31]

At the same time, parsing consistency and global coherence are crucial. Relying solely on local probability models often leads to shifting term boundaries across different paragraphs. To address this, global optimization strategies such as integer linear programming can be superimposed on the results of segmentation and syntactic analysis, incorporating discourse-level logical constraints into the solving process. For example, within the same case, "social insurance" should be segmented uniformly. If "termination" has been previously identified as an act of the employer, it should be interpreted consistently in subsequent mentions. Through global constraints, not only is parsing consistency maintained at a technical level, but symbolic legal knowledge can also be introduced as a supplement. For example, "labor capacity assessment" is invariably linked to work-related injury certification, and "economic compensation" usually co-occurs with contract termination. These explicit rules can play a corrective role during optimization, making the parsing results more aligned with legal logic.

Consequently, foundational parsing is no longer an isolated technical step but an integrated framework that fuses statistical learning with symbolic reasoning, capable of both preserving the integrity of legal terms and maintaining logical stability at the discourse level. After such domain-specific adaptation, the semantic precision and consistency of the input data will be significantly improved, allowing for subsequent discourse logic modeling and tripartite viewpoint classification to proceed on a clearer structural foundation. More importantly, this approach is not limited to the labor law context but holds universal value for other specialized legal domains, thus not only addressing the specific dilemmas in labor dispute documents but also providing a generalizable paradigm for the application of legal AI in a broader range of fields[32].

3.2. Reconstructing Discourse Logic and Reasoning with Graph Neural Networks and Legal Factors

In the automated analysis of labor dispute documents, relying solely on traditional sequence modeling methods can easily lead to semantic fragmentation and the loss of logical chains. To effectively reconstruct the discourse logic of legal texts, Graph Neural Networks (GNNs) can be introduced, with legal factors serving as the fundamental units for constructing an argumentation graph. This enables the model to perform reasoning and classification within a structured graph environment. The core of this approach is to remap the originally linear long-form text into a structured graph composed of nodes and edges. Through the propagation and aggregation of

---

[31] Hua W, Eger S, Gurevych I. Mixed-domain Language Modeling for Processing Long Legal Documents[C]//Findings of the Association for Computational Linguistics: EACL 2023. 2023: 83-95. https://aclanthology.org/2023.nllp-1.7

[32] Zhang G, Nulty P, Lillis D. Argument Mining with Graph Representation Learning[C]//Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law (ICAIL 2023). ACM, 2023: 325-329. https://eprints.bbk.ac.uk/id/eprint/53089/

information within the graph, deep relationships spanning across sentences and paragraphs can be captured, thereby improving the accuracy of viewpoint attribution and the interpretability of reasoning[33].

Legal factors are abstractions of key semantic units within the text. They include not only factual elements of the case (e.g., "status of labor contract signing," "wage payment records") but also encompass the parties' viewpoints (e.g., "plaintiff's claim," "defendant's defense"), and extend to legal bases and judgment outcomes (e.g., "Article 10 of the Labor Contract Law," "court finding"). By defining these factors as nodes in a graph and establishing edges based on their logical relationships, a legal element graph can be formed. For instance, when a plaintiff claims "no labor contract was signed," the defendant responds "remuneration was paid," and the court ultimately makes a finding by citing Article 10 of the Labor Contract Law, these three types of factors form a "contention-citation-finding" path, clearly illustrating the unfolding of the reasoning process. Compared to linear sequence input, a graph structure can more naturally express this multi-party interactive argumentative pattern. This approach aligns with the LUIMA framework proposed earlier at the International Conference on AI and Law, which provided an experimental basis for the structured modeling of complex legal reasoning through factor-based annotation[34].

Building on this, GNNs offer an effective learning mechanism. Through Graph Convolutional Networks (GCNs) or Graph Attention Networks (GATs), each node continuously aggregates information from its neighbors when computing its own representation. Thus, a single "court finding" node contains not only the semantics of the judgment conclusion but also integrates information from related plaintiff claims, defendant defenses, and cited articles, forming a contextually consistent comprehensive representation. As the number of layers increases, the model can aggregate information across multi-hop relationships, achieving global logical integration. This propagation mechanism allows the model to naturally learn relationships of support and refutation along the chain of "plaintiff's viewpoint – evidence – court finding," thereby possessing stronger explanatory power during classification.

In terms of implementation, candidate legal factors can first be identified using NLP techniques, such as extracting "agent-action-object" triplets through named entity recognition and dependency parsing, and then mapping them to a set of nodes[35]. The construction of edges can combine both rule-based and data-driven methods: on one hand, using a priori knowledge rules, such as "citing a legal article" typically points to a "court finding"; on the other hand, automatically learning edge weights through co-occurrence and semantic relationships in the training set. After the graph is constructed, GNNs are used for iterative propagation and representation updates,

---

[33] Zhang G., Nulty P., Lillis D. Argument Mining with Graph Representation Learning. ICAIL 2023.f

[34] Grabmair M., Ashley K. D., Chen R., et al. Introducing LUIMA. ICAIL 2015

[35] Dong, Yuhong (董玉红); Lu, Peng (卢鹏); Chen, Jing (陈静); et al. 2024. "Jiyu sifa caipan wenben de falü yaosu tiqu fangfa" [A Method for Legal Element Extraction Based on Judicial Judgment Texts]. Journal of China Academy of Electronics and Information Technology 19(06): 552–558, 569. (In Chinese).

ultimately outputting classification results at the node or graph level to discriminate the viewpoint attribution of different sentences or fragments.

The advantage of this method is its ability to explicitly preserve and utilize the logical structure of the legal text. First, the construction of nodes and edges presents argumentative relationships in a structured manner, allowing the model's reasoning process to be explained through visualizable paths. Second, the propagation mechanism of GNNs compensates for the deficiencies of traditional models in handling long-range dependencies, enabling a plaintiff's claim and a court's conclusion, separated by several paragraphs, to be connected through intermediate nodes, thus maintaining the integrity of the reasoning chain. Furthermore, the graph structure offers the possibility of cross-task transfer: the same legal factor graph can be shared across tasks such as viewpoint classification, legal article recommendation, and case matching, thereby reducing the overall modeling cost[36]. The introduction of GNNs is not merely a simple replacement of model architecture but signifies a paradigm shift in legal text analysis. It transforms unstructured text into an intermediate layer that combines symbolic and vectorized representations, retaining the logical transparency required for legal reasoning while leveraging the powerful semantic modeling capabilities of deep learning. This approach can not only improve the accuracy of tripartite viewpoint classification in labor dispute scenarios but also provide a replicable path for broader judicial AI applications. With the continuous accumulation of large-scale labor law corpora and legal knowledge graphs, the reconstruction of discourse logic based on legal factors and GNNs is poised to become a significant direction in legal NLP[37].

3.3. Optimizing Data Strategy with Semi-Supervised and Active Learning

In the data construction phase, scale can be expanded through semi-supervised learning, while the precision and efficiency of manual annotation can be enhanced with active learning. The idea behind semi-supervised learning is to use a small amount of manually annotated "seed data" to drive the automated processing of a large volume of unlabeled documents. Specifically, this method first maps all documents (both labeled and unlabeled) into a high-dimensional semantic space using word vector techniques. Then, based on clustering methods like K-means, the texts are aggregated according to semantic similarity[38]. On this basis, using the label

[36] Zhang G., Nulty P., Lillis D. Enhancing Legal Argument Mining with Domain Pre-training and Neural Networks. JDMDH, 2021

[37] Li, Xin (李鑫). 2025. "Falü da moxing goujian de moshi xuanze he shijian lujing" [Model Selection and Practical Paths for the Construction of Large Legal Models]. Jishou Daxue Xuebao (Shehui Kexue Ban) [Journal of Jishou University (Social Sciences Edition)] 46(01): 21–36. (In Chinese).

[38] Ma, Yao (马尧). 2023. "Mianxiang falü wenti shuju de K-means julei fenxi jishu de yanjiu" [Research on K-means Clustering Analysis Technology for Legal Problem Data]. Wangluo Anquan Jishu yu Yingyong [Network Security Technology & Application] (09): 35–37. (In Chinese).

distribution of the small set of existing "seed data," a "pseudo-label" with the highest confidence is assigned to an entire cluster, thereby efficiently transferring the knowledge from manual annotations to thousands of unlabeled samples[39]. This approach can rapidly generate large-scale training data at a low cost, allowing the model to learn rich semantic patterns at an early stage. To control the potential noise introduced by pseudo-labels, a confidence threshold can be set during the generation phase, or consistency regularization can be used to constrain the prediction consistency of multiple models on the same input, retaining only high-confidence samples for the training set to strike a balance between data scale and stability.

The active learning mechanism can further enhance the utilization of limited human resources, achieving the goal of "using good steel for the blade's edge." Its core lies in shifting from random sampling to model-driven intelligent sampling, allowing the model to automatically select the samples it is "most uncertain" or "most confused" about for expert annotation. For instance, through uncertainty sampling, samples where the model's predicted probability distribution is closest to uniform—i.e., where the model is "hesitating" between multiple options—can be selected. Through query-by-committee (or disagreement sampling), text fragments where different models or models from different training epochs produce highly divergent predictions can be identified. These strategies ensure that precious manual annotation time is prioritized for the boundary cases that are most difficult for the model to judge. This allows each incremental update of the training data to maximally fill the model's "knowledge gaps," continuously strengthening its classification capabilities. This is particularly valuable in labor dispute documents, where common expressions like 协商解除 (termination by mutual agreement) and 单方解除 (unilateral termination) have drastically different legal meanings. Active learning can effectively identify these high-value samples, avoiding the waste of annotation resources on a large number of simple, repetitive examples[40]

The combination of these two methods forms a dynamic, iterative data optimization mechanism. Semi-supervised learning acts like "casting a wide net," continuously expanding the data scale to provide the model with broad corpus coverage. Active learning, in contrast, is like "precision fishing," refining data quality at key points to provide corpus depth. As the model's performance improves, its ability to predict uncertainty becomes stronger, and the accuracy of pseudo-labels gradually increases, which in turn enhances the quality of the data generated by semi-supervised learning. Meanwhile, active learning continuously supplements the training set with new, difficult samples, creating a positive feedback loop for the entire corpus in terms of both scale and precision. Such a strategy enables the dataset

[39] Sun, Jiankai (孙建凯). 2013. "Shuju wajue zhong banjiandu K junzhi julei suanfa de yanjiu" [Research on Semi-supervised K-means Clustering Algorithm in Data Mining]. Dissertation, Zhejiang Ligong Daxue [Zhejiang Sci-Tech University]. (In Chinese).

[40] Zhang, Yifei (张亦菲); Li, Yanling (李艳玲); and Ge, Fengpei (葛凤培). 2025. "Jiyu tu shendu xuexi de sifa panjue yuce zongshu" [A Survey on Judicial Judgment Prediction Based on Graph Deep Learning]. Jisuanji Kexue yu Tansuo [Journal of Computer Science and Exploration] 19(08): 2024–2042. (In Chinese).

to evolve from "small and refined" to "large and superior" within a manageable cost framework.

From a long-term perspective, this mechanism also has the potential for continuous updating and cross-domain transfer. As new case types and terminologies emerge in legal practice, the model can rapidly absorb new corpora through semi-supervised methods and then use active learning to supplement annotations for key samples, maintaining adaptability to new linguistic phenomena. At the same time, the accumulated experience in methodology and data construction can be transferred to other adjacent domains, such as social insurance or contract disputes, providing unified corpus support for multi-domain judicial intelligence[41].

4. Conclusion

The structuring and interpretability of judicial texts mark a pivotal transition for intelligent justice, signaling a shift from the phase of data accumulation to the phase of knowledge governance. Focusing on labor dispute adjudicative documents as its object of study, this paper has explored feasible paths for the reconstruction of judicial texts from semantic, logical, and institutional perspectives. By constructing a domain-specific corpus system for labor law, an argumentative structure centered on legal factors, and a dynamically evolving data optimization mechanism, this paper attempts to build a traversable bridge between computational models and jurisprudential logic, enabling the linguistic expressions and legal reasoning within judicial texts to be mapped and reconstructed at an algorithmic level. The research indicates that the structuring of judicial texts based on legal factors not only enhances the precision and stability of tripartite viewpoint identification in labor dispute cases but also promotes the interpretability and normative consistency of judicial reasoning at an institutional level.

From a broader perspective, the key to judicial intelligence lies not in "replacement" but in "empowerment." Artificial intelligence should serve as a technological extension of judicial rationality, enabling the organic unification of legal logic, facts, and values within data-driven governance. Future research in judicial AI must seek a balance among algorithmic transparency, rule legitimacy, and institutional compatibility, advancing the digital representation and continuous evolution of legal knowledge systems. This exploration into the structuring of labor dispute documents based on legal factors provides a template for constructing computable, verifiable, and traceable judicial knowledge graphs. It also offers new theoretical insights and methodological support for the institutional construction of intelligent justice and the digital transformation of the rule of law in China.

---

[41] Jiang, Mingqi (江明奇); Yan, Qian (严倩); and Li, Shoushan (李寿山). 2019. "Jiyu lianhe xuexi de kua lingyu falü wenshu zhongwen fenci fangfa" [A Cross-Domain Chinese Word Segmentation Method for Legal Documents Based on Joint Learning]. Zhongwen Xinxi Xuebao [Journal of Chinese Information Processing] 33(09): 17–23. (In Chinese).