

# Research on the Dynamic Update and Intelligent Governance Mechanism of Legal Corpus

**Qinglin Ma, Chao Meng**

School of Foreign Languages, Northwest University of Political Science and Law, Xi'an 710122, Shaanxi, China

**Copyright:** © 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

**Abstract:** With the widespread application of artificial intelligence and big data technology, legal corpus plays an increasingly important role in legal intelligent systems, rule of law research, and judicial informatization. As a fundamental resource for processing legal texts, the quality and timeliness of legal corpus directly affect the training effectiveness of artificial intelligence models and the level of intelligence in legal services. However, the rapid update speed of legal texts, the complex content structure, and the variability of legal definitions have led to significant challenges in the updating and governance of existing legal corpus. How to achieve dynamic updates and intelligent governance of legal corpus has become an urgent issue that needs to be addressed. This paper conducts a systematic study on the dynamic updating and intelligent governance mechanisms of legal corpus, proposing a multi-dimensional, multi-level, and scalable updating strategy for legal corpus, which combines privacy computing and blockchain technology to construct an intelligent governance framework that supports compliance across multiple legal domains and monitors the use of legal text data. The research results indicate that this mechanism can effectively enhance the timeliness and controllability of legal corpus, providing a safer and smarter legal text data system for legal AI applications.

**Keywords:** Legal Corpus; Dynamic Update Mechanism; Intelligent Governance Mechanism; Privacy Computing

**Online publication:** June 26, 2025

## 1. Introduction

### 1.1. Research background and significance

As an important resource supporting legal text analysis, legal reasoning systems, and legal intelligent question-answering, legal corpus have received widespread attention in recent years. With the rapid development of artificial intelligence and natural language processing technologies, the role of legal corpus in judicial services, legal research, and policy analysis has become increasingly prominent. However, the dynamic characteristics of legal texts determine that legal corpus must have the capability for continuous updates to ensure the timeliness and applicability of their data services. Moreover, the sensitivity of legal data dictates that the governance mechanisms during the construction process must be rigorous to ensure the legality and compliance of the corpus content<sup>[1]</sup>.

Currently, most legal corpora are still constructed in a static manner, making it difficult to cope with the challenges posed by the frequent updates of laws and regulations. For example, since the implementation of China's Civil Code in 2021, it has had a profound impact on numerous legal cases, and the relevant legal texts and data need to be updated in real time. Meanwhile, the European General Data Protection Regulation (GDPR) has imposed strict compliance requirements

on the management of personal information within legal corpus. Therefore, establishing a legal corpus mechanism that can be dynamically updated and achieve intelligent governance has become an important issue in the current legal technology field<sup>[2]</sup>.

## **1.2. Research issues and challenges**

The core focus of this research is how to construct a framework for a legal corpus that can be dynamically updated and intelligently governed. Based on existing research and technology, the main challenges can be summarized in the following aspects:

First, the dynamic updating of legal data relies on efficient text collection and annotation systems, but the diversity of legal texts (including case records, regulations, judgments, legislative materials, etc.) and the complexity of language (involving a large number of legal terminology, implied meanings and complex legal logic) pose high demands for data processing. How to achieve automated updates while ensuring the accuracy and completeness of the texts is one of the current research difficulties.

Second, the issues of legal data security and compliance are becoming increasingly prominent. Due to the fact that legal data often involves personal privacy, trade secrets, and personal identification information, how to balance data security and usability during the construction of legal corpus has become an urgent issue to be addressed. The existing data architecture has obvious shortcomings in terms of security and compliance, especially in the governance of legal data across multiple jurisdictions, where current technological means cannot achieve automated compliance detection and adjustment.

## **1.3. Research objectives and innovations**

The main research objectives of this paper include the following three aspects:

- (1) Establishing a dynamic update mechanism for legal corpus: design an automated data collection, updating and annotation system suitable for the legal field, enabling the legal corpus to respond quickly to changes in laws and regulations, achieving real-time data updates and quality control.
- (2) Achieving intelligent governance of legal corpus: construct an intelligent governance framework that supports multi-jurisdictional regulation, data traceability and privacy protection in combination with privacy computing, blockchain and natural language processing technologies, ensuring the security and compliance of legal corpus.
- (3) Promoting the intelligent development of legal AI systems: enhance the efficiency and applicability of legal corpus by establishing dynamic updating and intelligent governance mechanisms, to provide more accurate and compliant data support for the training and optimization of legal intelligent systems.

The innovations of this research are mainly reflected in the following three aspects:

- (1) Multi-source data fusion and intelligent updating: this paper proposes a cross-agency or multi-type legal data fusion method based on legal terminology graphs and intelligent text classification, enabling the legal corpus to have the capability for automatic updates from multiple sources.
- (2) Intelligent design of legal governance: by introducing privacy computing and blockchain technology, this paper constructs a governance system that integrates data security and legal compliance, capable of automatically identifying and processing illegal data access.
- (3) Synergistic development of legal corpus and legal AI: this paper proposes a path for the synergistic development of legal corpus and legal intelligent systems, focusing on the dynamic updating of corpora and real-time training of AI models, providing data support for the intelligent development of legal AI.

## **2. Design of Dynamic Update Mechanism for Legal Corpus**

### **2.1. Necessity of dynamic updating of legal corpus**

The dynamic updating of legal corpus is a key link to ensure that its content is synchronized with changes in laws and regulations. Legal texts have the characteristic of rapid updates, especially the publication of legislative provisions, judicial interpretations, and temporary policies, all of which have a significant impact on the construction and use of legal corpus. Traditional legal corpus often relies on static collection, which cannot timely reflect the latest legal developments. Therefore, the dynamic update mechanism of legal corpus can not only enhance the quality of the corpus but also strengthen its support capabilities for legal AI systems and legal research<sup>[3]</sup>.

### **2.2. Core elements of dynamic update mechanism**

The dynamic update mechanism for legal corpus proposed in this paper mainly includes the following core elements:

#### **2.2.1. Data collection and preprocessing**

The dynamic updating of the legal corpus must first address the issue of data collection. Since legal texts come from multiple institutions, including courts, government agencies, and legal research institutions, it is necessary to build a data collection system that can interface with the data of each institution in real time. This system should have the capability to automatically identify, classify, clean, and store data, ensuring that legitimate data can be efficiently entered into the corpus.

#### **2.2.2. Classification and topic identification of legal texts**

The types of legal corpus are diverse, including civil law, criminal law, administrative law, economic law, and social law, and different types of legal texts have varying requirements for information updating and annotation. Therefore, this paper designs a method for text classification and topic identification based on Natural Language Processing (NLP), which can classify legal texts according to legal terminology and semantic features to facilitate subsequent data annotation and updating.

#### **2.2.3. Construction and maintenance of legal terminology graphs**

Accurate legal terminology is the foundation for the dynamic updating of the legal corpus. This paper proposes a terminology management method based on legal terminology graphs, which can automatically identify new legal terminology, update existing legal terminology, and achieve structured annotation of legal corpus. By establishing a legal terminology graph, the corpus can automatically adjust its terminology system when laws and regulations change, ensuring the consistency and accuracy of the data.

#### **2.2.4. Automatic annotation and rule construction**

The core of the legal corpus lies in how to annotate legal texts with high fidelity. This paper proposes an annotation mechanism based on a combination of rules and machine learning, designing a rule library that supports dynamic adjustments, capable of automatically optimizing annotation standards according to changes in legal policies. This mechanism can automatically adjust the annotation methods for relevant terminology when legal terminology is updated or changed, ensuring the data quality of the legal corpus.

#### **2.2.5. Automated updates and version control**

The updates of legal corpus often have temporal and version differences, thus it is necessary to establish a version control system to ensure that the data update process of the legal corpus is traceable and auditable. This paper introduces blockchain technology, which constructs data snapshots and change histories of legal corpus to achieve intelligent tracking and version management of the update process.

### **2.3. Implementation path of dynamic update mechanism**

To achieve dynamic updates of the legal corpus, this paper proposes a multi-stage implementation path, including the following key steps:

#### **2.3.1. Automation enhancement of data collection**

Construct an automated legal text collection system that can connect in real time to multiple legal databases, such as the China Judgments Online, legal industry databases, and legal consulting platforms. Using natural language processing technology, the system can automatically identify new legal texts and perform data cleaning, deduplication and format standardization.

#### **2.3.2. Optimization of classification system of legal text**

During the collection of legal texts, preliminary classification of the legal texts is required to ensure the efficiency of subsequent annotation work. This paper adopts a BERT-based classification method of legal text, achieving high-precision classification of legal texts and enhancing the organization and management efficiency of the legal corpus.

#### **2.3.3. Dynamic maintenance of legal terminology graphs**

Construct a legal terminology graph system to manage the legal terminology within the legal corpus. This system not only supports the automatic identification and annotation of legal terminology but also allows for automatic updates and version management of legal terminology. By introducing a change tracking mechanism, the terminology graph can automatically adjust the annotation standards of legal texts following changes in laws and regulations<sup>[4]</sup>.

#### **2.3.4. Intelligent adjustment of legal corpus annotation rules**

The annotation work of legal corpus needs to continuously adapt to changes in legal policies. This paper proposes an automated annotation rule adjustment mechanism based on legal text analysis and machine learning, achieving intelligent optimization of the annotation standards for legal corpus.

## **3. Intelligent Governance Mechanism for Legal Corpus**

### **3.1. Necessity of governance of legal corpus**

The intelligent governance of legal corpus is a prerequisite for ensuring their safe, legal, and controllable use. Since legal data often involves personal privacy, commercial secrets and national security, a lack of effective governance mechanisms may expose legal corpus to risks such as data breaches, unauthorized access, and data misuse. Especially in the current data-driven legal technology application environment, the governance of legal corpus must not only manage the scope of data usage but also possess the capabilities for monitoring, auditing, and rectifying data usage.

### **3.2. Design principles of governance mechanism**

The intelligent governance mechanism for legal corpus proposed in this paper is primarily based on the following technical principles:

#### **3.2.1. Application of privacy computing technology**

In order to ensure the data security of the legal corpus, this paper introduces privacy computing technologies such as Differential Privacy and Federated Learning to ensure that legal data is not leaked during collection, processing, and usage. Privacy computing technologies can ensure the security of sensitive information while allowing data sharing through encryption and anonymization.

### **3.2.2. Introduction of blockchain technology**

Blockchain technology, with its characteristics of distributed storage, data traceability, and immutability, provides a reliable technical foundation for the governance of the legal corpus. This paper uses blockchain technology to construct a tracking system for legal corpus data flow, making the processes of collecting, storing, using, and updating legal texts traceable, ensuring the transparency of legal data usage.

### **3.3. Technical implementation of governance mechanism**

To construct an intelligent governance framework, this paper proposes a governance mechanism based on technological integration, which includes the following technical modules:

#### **3.3.1. Construction of privacy computing system**

During the construction of the legal corpus, privacy computing technologies are used to ensure data security. This paper employs Differential Privacy technology to inject noise into legal texts to reduce the risk of sensitive information leakage. In addition, by combining Federated Learning technology, different institutions are allowed to conduct joint model training without sharing raw data, ensuring the security and compliance of data governance.

#### **3.3.2. Implementation of Blockchain data tracking system**

Blockchain technology is used to build a data tracking system for legal corpus, ensuring the traceability of the data lifecycle. By introducing a smart contract mechanism, automated management of data usage, access, and sharing can be achieved. This system can record the history of data usage, ensuring the transparency and auditability of legal data flow.

#### **3.3.3. Integration of machine learning and NLP**

This paper utilizes natural language processing technology to perform semantic analysis on legal texts, and combines machine learning models for automatic classification and labeling. By training a legal text classification model, it can automatically identify illegal data access behaviors and automatically block sensitive content, ensuring that the use of the legal corpus complies with legal regulations.

### **3.4. Application examples of governance mechanisms**

To verify the effectiveness of the intelligent governance mechanism proposed in this paper, a series of experiments and application examples are conducted. First, conduct intelligent governance testing on the legal judgments of the court system. The experiment has selected 100,000 legal judgments obtained from the China Judgments Online, recording the collection, storage, and usage process of the judgments through a blockchain system, and using Differential Privacy technology to process sensitive information to ensure that personal data will not be leaked.

The experiment has automatically verified the compliance of legal texts. The experiment has selected several updated legal provisions and compared them with data in the legal corpus. The compliance rule engine can automatically determine whether the data in the corpus complies with the latest laws and regulations, and intelligently filter or replace content that does not meet legal standards<sup>[5]</sup>.

The experimental results indicate that the intelligent governance mechanism proposed in this paper can effectively enhance the compliance and controllability of the legal corpus, making it suitable for data management needs in different legal environments.

## **4. Synergistic Effect of Dynamic Updates and Intelligent Governance**

### **4.1. Role of dynamic update mechanism in intelligent governance**

The dynamic update mechanism for the legal corpus not only addresses the timeliness issue of legal text data but also

plays an important role in promoting intelligent governance. Since updates to legal texts often accompany changes in legal terminology, legal logic, and legal norms, correct data updates can not only ensure the accuracy of the corpus content but also provide a solid data foundation for intelligent governance.

#### **4.2. Supporting role of intelligent governance in dynamic updates**

The intelligent governance mechanism provides strong technical support for the dynamic updating of the legal corpus by establishing a secure, traceable, and controllable data management system. First, the intelligent governance mechanism can monitor the usage of the legal corpus in real-time, ensuring that data is not accessed illegally or used inappropriately. This monitoring capability also helps the system identify which data needs to be prioritized for updates and which data has expired or is no longer applicable, thereby enhancing the efficiency of dynamic updates.

#### **4.3. Synergistic effect of two mechanisms**

The synergy between the dynamic update mechanism and the intelligent governance mechanism enables the legal corpus to achieve more efficient, intelligent, and controllable data management. The dynamic update mechanism ensures that the content of the legal corpus remains consistent with changes in laws and regulations, while the intelligent governance mechanism safeguards the legality and security of data usage through monitoring and management. The combination of two mechanisms allows for multi-layered security and compliance control during the data flow process, ensuring that the legal corpus reflects the latest legal developments while providing high-quality and highly available legal text data.

### **5. Conclusion and Outlook**

#### **5.1. Research summary**

This paper conducts a systematic study on the dynamic update and intelligent governance mechanisms of legal corpus, proposing a data management and governance solution suitable for the legal field. By constructing a dynamic update mechanism for legal corpus, this paper ensures the timeliness of legal texts, allowing the legal corpus to adapt to the rapid changes in laws and regulations and the need for system updates. The intelligent governance mechanism also ensures the security, compliance, and controllability of the legal corpus by incorporating technologies such as privacy computing, blockchain, and NLP.

The research finds that the dynamic updating of legal corpus not only enhances the quality and applicability of legal data but also plays a key role in the governance process of legal data, achieving automatic supervision and lawful use of legal text data. For example, by combining legal terminology graphs, intelligent annotation, and blockchain, this paper proposes a personalized governance strategy that can automatically identify sensitive content and support the requirements of different legal jurisdictions. These strategies not only enhance the intelligence level of the legal corpus but also increase its practical value in scenarios such as legal AI training, legal research, and judicial practice<sup>[6]</sup>.

#### **5.2. Practical application value**

The dynamic update and intelligent governance mechanisms of the legal corpus proposed in this paper have significant practical application value. Firstly, its dynamic update mechanism can strengthen the responsiveness of the legal corpus to changes in laws and regulations, ensuring the timeliness and consistency of legal text data. This is of great significance for the training and optimization of legal AI systems, as the training of AI models relies on high-quality and accurate legal data, and the dynamism of the legal corpus can meet this demand.

Secondly, the intelligent governance mechanism can provide more comprehensive technical support for the management of legal corpus. Through the collaborative application of privacy computing, blockchain, and natural language processing technologies, the system can achieve monitoring, auditing, and compliance assessment of legal data usage,



thereby enhancing the credibility and security index of legal corpus. For example, in scenarios such as legal consulting services, intelligent legal Q&A, and training legal reasoning models, intelligent governance can effectively prevent the illegal circulation and misuse of legal data, ensuring the legality and compliance of legal AI system applications.

## Funding

This paper is a phased outcome of the National Social Science Fund Project “Translation and Construction of a Chinese-English Parallel Corpus of the Regulations and Rules of the Communist Party of China and the Regulations of Major Political Parties in the World” (Project No.: 19XYY014).

## Disclosure statement

The author declares no conflict of interest.

## References

- [1] Chen J, Yang L, 2025, Ideological Risks of Generative Artificial Intelligence and Its Legal Responses. *Journal of Xi'an Jiaotong University (Social Sciences Edition)*, 45(04):158-168. DOI:10.15896/j.xjtuskxb.202504014.
- [2] Zhu Y, Liu Y, 2022, A Comparative Study of the English Translation of the Civil Code of the People's Republic of China Based on Corpora: Taking the English Translation of Core Civil Concepts and Related Terms as an Example. *Foreign Languages (Journal of Shanghai International Studies University)*, 45(05): 87-99.
- [3] Meng C, Ma Q, 2019, An Empirical Study on the Teaching Model of Legal English Vocabulary Based on Online Corpora. *Foreign Language Electronic Teaching*, (02):82-89. DOI:10.20139/j.issn.1001-5795.2019.02.012.
- [4] Cheng L, Hoffman G, 2017, A Semiotic Interpretation of the Principles of Legal Terminology Translation. *Journal of Zhejiang Gongshang University*, (06):15-24. DOI:10.14134/j.cnki.cn33-1337/c.2017.06.002.
- [5] Song L, 2023, Linguistic Data Foundation, Methods, and Applications of Digital Jurisprudence: Taking the Birth and Development of Legal Corpus Linguistics as an Example. *Eastern Law*, (06):118-129. DOI:10.19404/j.cnki.dffx.20231116.004.
- [6] Cheng Y, Wang H, 2012, Research on Copyright Protection of English Corpora in China. *Editorial Friend*, (11):87-90. DOI:10.13786/j.cnki.cn14-1066/g2.2012.11.009.

### Publisher's note

*Whioce Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.*