Hybrid Legal Clause Classification: Integrating Supervised BERT with Reinforcement Learning

Shruti Bhushan
Apex Institute of Technology (CSE)
Chandigarh University, Punjab, India
bhushanshruti46@gmail.com

Hemaditya AVS
Apex Institute of Technology (CSE)
Chandigarh University, Punjab, India
hemadityaavs@gmail.com

Arpita Verma
Apex Institute of Technology (CSE)
Chandigarh University, Punjab, India
varpita1118@gmail.com

2025 3rd International Conference on Disruptive Technologies (ICDT) | 979-8-3315-1958-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICDT63985.2025.10986477

***Abstract*—Detecting ambiguity or classifying as clear or vague is a crucial yet important task in legal document analysis, which often requires human feedback. This research involves a unique approach by integrating supervised Machine Learning with Reinforcement Learning for automatic legal clause classification. This work proposes a goal to develop a system which is able to accurately identify ambiguity in legal documents by pre-trained BERT embeddings and dynamic learning methods. The initial part of the system is trained using labeled data that predicts the clause as ambiguous or non-ambiguous. This model uses BERT for clause representation. The latter part of the system inculcates Reinforcement Learning (RL) algorithm such as Q-Learning to interact with the environment created by legal dataset. The agent used in implementation of Reinforcement learning learns to classify clauses by receiving rewards as feedback. After multiple training episodes of RL agent, it dynamically improves its decision making by maximizing cumulative reward. This integrated system of Supervised Machine Learning and Reinforcement Learning helps the system to handle situations with well-labeled datasets and environments with available feedback but limited direct supervision. The proposed model showcases the potential of the integrated system for legal document analysis and scalable solution for clause classification task.**

***Index Terms*—Supervised Machine Learning, Reinforcement Learning, Legal Clause, Ambiguity, Textual data.**

I. INTRODUCTION

The task of legal clause classification holds a prominent place in the domain of legal document analysis, specifically in contract review and legal risk assessment. The legal contract is a document holding the rules and conditions on which any bond or contract is signed. It acts as proof in case of any discrepancy or violation of rules. This document ensures that both the parties among which the contract has been signed is to follow the mentioned guidelines, violation of which can lead to serious legal actions to be taken against the party which violates the rules. Legal contract is composed of many complex and dense language clause which ordinarily include many types of clauses each referring to one specific field such as: confidentiality, liability, indemnity etc. These contracts are principally large documents which if misinterpreted may lead to misunderstandings or even disputes.

Traditionally this task of reading, summarizing and understanding of legal documents were done by legal professionals, but somehow this method of legal clause classification is time consuming and error prone. The traditional method of assessing the legal document is not a good fit for today's exponentially increasing legal cases all across the world. The existing models used for legal document summarization work well with providing the summary of the entire documentation but not assessing the ambiguity and risk.

To address this severe problem of ambiguity, Machine Learning has sprung up as an encouraging approach. Machine Learning models deal with wide variety of data of every size depending on the model being used. Machine Learning approach has shown a promising result in classification task in different fields of healthcare, finance etc. Machine Learning techniques involving Supervised Learning, Natural Language Processing including pre-trained language models like BERT (Bidirectional Encoder Representations Transformers) have been a great success in recognizing the nuances of legal documentation.

Supervised Learning is an effective approach for classification tasks but is limited where labeled data is sparse or unavailable. In this research, we are dealing with legal documents which are dynamic in nature containing clauses which may vary in structure and meaning depending on the aspect through which the contract is being signed. In such scenarios, Supervised Learning solely used can be less effective, which is why the integration of Reinforcement Learning with the Supervised Learning model is required. Since Reinforcement Learning is based on feedback-based learning, it can dynamically adjust the agent based on the environment and provided reward. This integration of Supervised Learning and Reinforcement Learning can more efficiently classify the clauses based on ambiguity or risk.

We will describe the hybrid system's architecture, the techniques for training the supervised and RL models, and the outcomes of our tests in the sections that follow. According to our research, the hybrid model outperforms classic supervised learning techniques in a number of situations, especially when labeled data is scarce or non-existent. The RL agent's flexibility also helps it to more effectively generalize in many legal situations, which makes the system a strong option for classifying legal clauses.

II. LITERATURE REVIEW

The legal document analysis, specifically the classification of legal clauses is a topic of considerable research. The legal cases have been increasing over time exponentially, which has led to integrating technologies such as Machine Learning and Natural Language Processing with the legal domain[1-4]. Having a look at the traditional ways of parsing or summarizing legal documents or assessing risk within the same has shown immense consumption of time and was highly error prone as it was more often accomplished by human labor.

Legal documents consist of complex clauses which are not easily interpretable and have different meanings with different jurisdiction, the clauses may vary over time. One of the reasons to include an integrated system are the variation in clauses over time[7]. Traditional ways of legal clause classification and risk assessment involved feedback from legal experts which sometimes, weren't convincible and the matter of dispute between the involved parties.

The introduction of Machine Learning into this domain has facilitated the development of efficient and scalable solutions. The researches on Supervised Learning and Reinforcement Learning have paved the way for improved classification tasks and reduction of ambiguity.

A. Supervised Learning in Legal Document Analysis

Supervised Learning is one of the prominent Machine Learning Techniques which has particularly been used in text classification in various areas of research also for legal document analysis. The initial text classification relied heavily on bag-of-words or TF-IDF (Term Frequency-Inverse Document Frequency) which were integrated with classical classification algorithms such as Support Vector Machine (SVM) and Logistic Regression. These models were used in figuring out simple text patterns in any documentation but were not highly effective for capturing deeper semantics required for text classification tasks.

With this limitation in the previous models, the advancement of Deep Learning models has led to proper interpretation and understanding of complex legal texts[2]. Models like Recurrent Neural Network (RNNs) and Long Short-Term Memory (LSTM) were brought into play for the proper and specific interpretability of legal clauses[5]. These models were able to understand the sequential nature of the lengthy legal texts, handling dependencies between words over longer text sequences. Somehow, these models were found limited to the parallel training and loosing information from previous texts as the input lengths increased.

To overcome the limitation of deep learning models of losing information due to increase in length of text, paved the way for development of Transformer Based Model such as BERT (Bidirectional Encoder Representations from Transformers) have brought transformable change in cutting-edge technologies like Natural Language Processing (NLP). BERT inculcated concepts of word embeddings, where the semantic parsing of the word was the matter of concern, meaning of each word was determined by context in the entire sentence, dealing with complex words and their relationships with other words or sentences in the content.

Though BERT and similar Transformer models have been more effective in understanding the complex nature of texts but still heavily rely on large labeled data for training. These models work well with sufficient amount of training data, however in legal domains the legal clauses vary over time and lead to scarcity in labeled dataset, especially in few areas such as intellectual property law or contract negotiation. The requirement of labeled data limits the scalability of purely supervised models, particularly when applied to real-world legal tasks where ambiguity and varied interpretations are common.

B. Text Classification Using Reinforcement Learning

Reinforcement Learning is another prominent Machine Learning technique that recently seen a boom in the field of research[10]. The researches related to reinforcement learning have shown that integration of technique in any machine learning models help the model to perform more efficiently and result in providing high accuracy. Reinforcement Learning is never limited to labeled data, instead an agent in RL makes decision through trial and error and mostly considers the phenomenon of Markov Decision Process[10]. The agent performs particular action in some state and gains rewards or penalty based on the action performed. The agent's prominent goal is to learn an optimal policy which could yield maximum cumulative reward in the entire process. This ability to learn with the environment, rather than through direct supervision, makes RL suitable for applications where labeled data is scarce or expensive to obtain.

In legal domain, RL has seen growing interest as it is flexible in handling ambiguous or incomplete information. There are many applications of Reinforcement learning in legal domain, one such is contract negotiation[11], here agent itself interacts with other parties and makes decision based on the reward and penalty gained. When RL is combined with the advanced models like BERT, it results in dynamic learning and increase in effectiveness of any task. In our case, RL is being used with Supervised Learning model in order to improve the clause classification and reduce the ambiguity.

TABLE I
COMPARATIVE STUDY OF CONTRIBUTIONS AND LIMITATIONS OF DIFFERENT FRAMEWORKS

Study	**Methodology**	**Limitations**
Traditional Legal clause classifications	Manual reviews	Time consuming, error-prone
Supervised Learning	TF-IDF,SVM, Logistic Regression	Requires large dataset
Transformer based NLP models (BERT)	Capturing complex cause semantics	Extensive training data, Less adaptability
RL for text classification	Agent classifies legal clause via reward	Excessive training time
Hybrid approach (Supervised+RL)	Integration for adaptive clause classification	Computationally expensive

III. Methodology

This section offers a complete methodological framework that uses both supervised deep learning and reinforcement learning to solve the problem of temporal semantic variability in legal clause interpretation. The architecture combines contextual language understanding from BERT (Bidirectional Encoder Representations from Transformers) with adaptive decision-making through Q-learning, thus enabling both static pattern recognition and dynamic policy optimization.

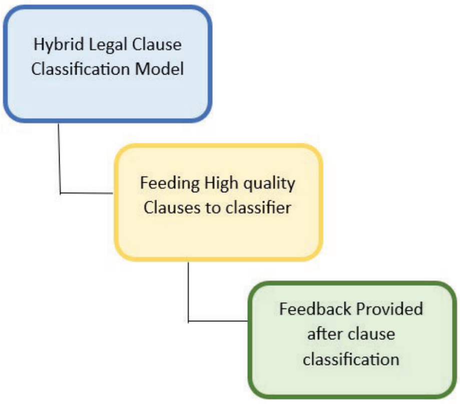
Hybrid Legal Clause Classification Model
Feeding High quality Clauses to classifier
Feedback Provided after clause classification

Fig. 1. Model Design

A. Supervised Learning Framework with BERT

The core element utilizes BERT (Bidirectional Encoder Representations from Transformers) using learning on labeled datasets to build fundamental classification abilities.

1) Text Representation Formalism: In this case, let

$$C = \{c_1, c_2, \ldots, c_n\}$$

be the set of legal clauses, where each clause c_i is represented as a sequence of tokens $t_1, t_2, \ldots, t_m$. The BERT tokenizer uses a subword tokenization method

$$\tau : C \rightarrow \mathbb{Z}^k$$

which converts clauses into a pre-determined length of token sequences:

$$\tau(c_i) = [[\text{CLS}], t'_1, t'_2, \ldots, t'_{k-1}, [\text{SEP}]]$$

where t_j stands for Word Piece tokens and k denotes the maximum sequence length, which is usually set at 512 characters in length. The [CLS] token combines the representations of the sequences at a level, for tasks involving classification.

2) Contextual Embedding Generation: The transformer architecture of BERT computes contextualized embeddings by stacking self-attention layers. The model generates hidden representations of an input token sequence $\tau(c_i)$:

$$H_l = \text{TransformerBlock}(H_{l-1}), \quad l \in \{1, \ldots, L\}$$

where $L = 12$ for BERT-base, and each transformer block implements:

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

with Q, K, V being learned query, key, and value matrices respectively. The final hidden state,

$$h_{[\text{CLS}]} \in \mathbb{R}^d$$

where $d = 768$ serves as the clause representation.

3) Classification Layer Optimization: A dense classification head projects the [CLS] embedding into decision space:

$$p(y|c_i) = \sigma(W^T h_{[\text{CLS}]} + b)$$

where,

$$W \in \mathbb{R}^{d \times 2}$$

$$b \in \mathbb{R}^2$$

and σ denotes softmax activation. The model minimizes categorical cross-entropy loss:

$$\mathcal{L}_{\text{sup}} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=0}^{1} y_{i,c} \log p(y = c|c_i)$$

through adaptive moment estimation (Adam) optimization with weight decay.

B. Reinforcement Learning Component

The additional element utilizes Q-Learning to adjust strategies in time for situations with limited rewards and changes, in temporal patterns.

1) Markov Decision Process Formulation: We define the classification task as a Markov Decision Process.

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$$

:

- **State Space** $\mathcal{S}$: BERT embeddings

$$h_{[\text{CLS}]} \in \mathbb{R}^d$$

- **Action Space** $\mathcal{A}$: {0 (ambiguous), 1 (clear)}
- **Reward Function** $\mathcal{R}$:

$$r_t = \begin{cases} +1 & \text{if } a_t = y_{\text{true}} \\ -1 & \text{otherwise} \end{cases}$$

- **Transition Dynamics** $\mathcal{P}$: Deterministic state progression through clause sequence

2) Q-Learning Update Mechanism: The agent learns optimal action-value function.

$$Q^* : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$$

through temporal difference learning:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right]$$

where α is the learning rate and $\gamma \in [0, 1)$ the discount factor. We stabilize training with experience replay and mini-batch updates:

$$\mathcal{L}_{\text{RL}} = \mathbb{E}_{(s,a,r,s') \sim \mathcal{D}} \left[\left(r + \gamma \max_{a'} Q_{\text{target}}(s', a') - Q(s, a) \right)^2 \right]$$

3) Hybrid Architecture Integration: The complete system combines both components through:

1) **Warm Start Initialization**: Supervised BERT model provides initial policy.

$$\pi_0 : \mathcal{S} \to \mathcal{A}$$

2) **Embedding Space Alignment**: Frozen BERT encoder maps clauses to state representations.
3) **Dual Optimization**: Joint learning of classification boundaries (supervised) and reward maximization (RL)

The hybrid objective function becomes:

$$\mathcal{L}_{\text{total}} = \lambda \mathcal{L}_{\text{sup}} + (1 - \lambda) \mathcal{L}_{\text{RL}}$$

where λ controls component weighting.

C. Training Protocol

1) Supervised Phase:

- **Dataset**:

$$\mathcal{D}_{\text{train}} = \{(c_i, y_i)\}_{i=1}^N$$

with balanced class distribution

- **Hyperparameters**:
  - Learning rate: 2×10^{-5}
  - Batch size: 32
  - Epochs: 5 (early stopping)

2) Reinforcement Phase:

- **Exploration Strategy**: ϵ-greedy with linear decay

$$\epsilon_t = \max(0.01, 1 - t/T)$$

- **Q-Network Architecture**: Two-layer MLP with ReLU activation
- **Training Regimen**:
  - Episode length: M clauses
  - Replay buffer size: 10,000
  - Target network update frequency: 100 steps

D. Evaluation Metrics

1) Supervised Performance:

- **Classification Metrics**:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

2) Reinforcement Learning Performance:

- **Convergence Metric**: Average discounted return per episode.

$$G_t = \sum_{k=0}^{M-t} \gamma^k r_{t+k+1}$$

- **Exploration Efficiency**: State-space coverage metric.

$$\mathcal{C} = \frac{|\{s \in \mathcal{S}\}|}{|\mathcal{S}|}$$

This methodology enables to adapt continuously to the changing legal semantics while keeping robust baseline performance by the complementary learning mechanisms of the hybrid architecture. The proposed approach integrates deep contextual representations with model-free reinforcement learning for a novel way of dealing with temporal concept drift in legal NLP tasks.

Algorithm Supervised Pre-training for Legal Clause Classification

```
Require: Legal clauses dataset $\mathcal{D} = \{(c_i, y_i)\}_{i=1}^N$ with balanced labels.
Require: Pre-trained BERT model with tokenizer $\tau$ and transformer encoder.
Ensure : Trained BERT classifier for legal clause interpretation.

for each clause $(c_i, y_i) \in \mathcal{D}$ do
    Tokenization:
      $\tau(c_i) = [\text{[CLS]}, t'_1, t'_2, \ldots, t'_{k-1}, \text{[SEP]}]$
    Embedding Extraction:
      Compute hidden states: $H = \text{BERT}(\tau(c_i))$
      Extract representation: $h_{\text{[CLS]}} \in \mathbb{R}^d$
    Classification:
      Compute logits: $z = W^T h_{\text{[CLS]}} + b$
      Compute prediction: $p(y|c_i) = \text{softmax}(z)$
    Loss Computation:
      Compute cross-entropy loss:
        $\mathcal{L}_{\text{sup}} \mathrel{+}= -\log(p(y_i|c_i))$
end
Optimize BERT and classification head parameters using Adam with weight decay.
return Trained BERT classifier.
```

Algorithm Reinforcement Learning Adaptation for Legal Clause Interpretation

```
Require: Trained BERT classifier with frozen encoder for
         state extraction.
Require: Legal clauses sequence $\{c_i\}$ with true labels $\{y_i\}$.
Require: Q-network with parameters $\theta$, replay buffer $\mathcal{D}_{\text{RL}}$.
Ensure : Adapted Q-network optimizing clause classification
         decisions.
Initialize exploration rate $\epsilon \leftarrow 1.0$ and replay buffer $\mathcal{D}_{\text{RL}}$
for each training episode do
    for each clause $c_i$ in sequence do
        State Extraction:
          Compute state: $s_i = h_{[\text{CLS}]}$ (via frozen BERT encoder).
        Action Selection:
          Select action $a_i \in \{0, 1\}$ using an $\epsilon$-greedy policy.
        Reward Assignment:
          Set reward:
          $r_i = \begin{cases} +1, & \text{if } a_i = y_i \\ -1, & \text{otherwise} \end{cases}$
        Transition Storage:
          Store $(s_i, a_i, r_i, s_{i+1})$ in $\mathcal{D}_{\text{RL}}$.
        Q-Learning Update:
          Sample a mini-batch from $\mathcal{D}_{\text{RL}}$ and update:
          $Q(s_i, a_i) \leftarrow Q(s_i, a_i) + \alpha \Big[ r_i + \gamma \max_{a'} Q(s_{i+1}, a') - Q(s_i, a_i) \Big]$
        Update exploration rate: $\epsilon \leftarrow \max(\epsilon_{\text{min}}, \epsilon - \Delta\epsilon)$
    end
end
return Adapted Q-network.
```

IV. RESULT

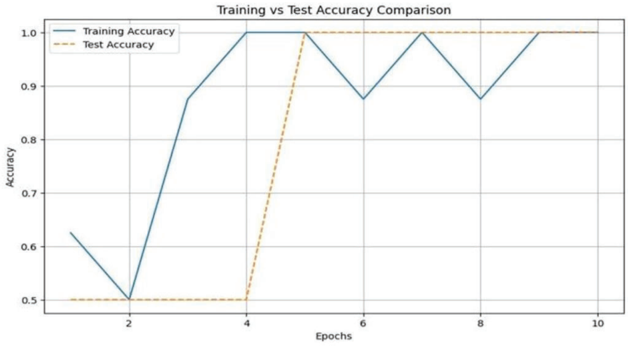

Fig. 2. Training vs Test accuracy

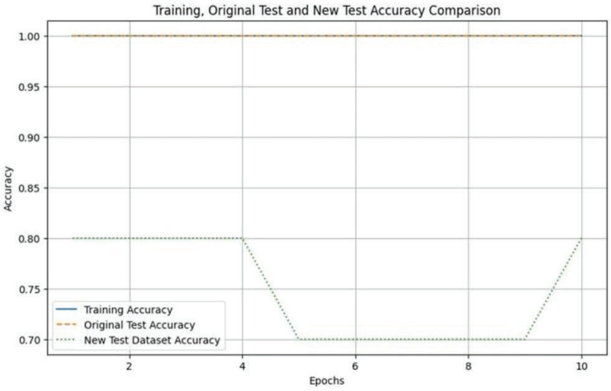

Fig. 3. Training, Original and New Test accuracy

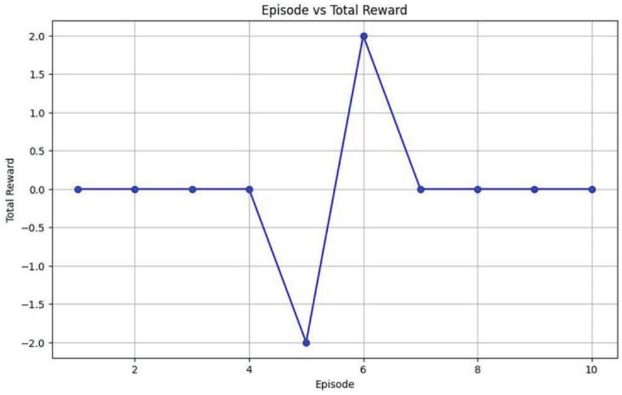

Fig. 4. Episode vs Total Reward

The training accuracy reaches 100 percent by the 4th epoch, stabilizing for the remainder of training. Test accuracy follows the same trend, reaching almost 99 percent accuracy by the 5th epoch and remaining stable, indicating the model effectively generalizes on the test data. (fig. 2) Both training and test sets converged rapidly, indicating efficient learning.

Training, Original and New Test Accuracy Comparison

The training and original test accuracy reach 100 percent in training and remains same, demonstrating effective learning on the original test set. (fig. 3) The new test dataset accuracy shows a slight decrease in mid-training but recovers toward the later epochs, fluctuating between 70 percent and 80 percent. The drop in new test accuracy suggests room for improvement in handling unseen data.

V. CONCLUSION

The study effectively showcases a hybrid model that employs reinforcement learning (RL) and supervised learning using BERT to categorize legal clauses as either clear or ambiguous. Its ability to comprehend legal text is demonstrated by the supervised BERT model, which attains near-perfect accuracy (100 percent) on both the training and test sets. But the model performs in a range of 70 percent to 80 percent on fresh test data that has not yet been seen, suggesting that the initial dataset may have been over-fitted and that more effective generalization strategies are required. Better consistency and long-term learning will require more tuning, as the RL agent, which is meant to dynamically improve through feedback, exhibits inconsistent performance with rewards changing across episodes.

VI. FUTURE WORK

In the future, performance on untested data may be enhanced by regularization, cross-domain training, and data augmentation to further improve the generalization of the BERT model. More steady and efficient learning may also result from improving the RL agent's reward system, exploration-exploitation strategy, and training procedure. Investigating explainability mechanisms and human-in-loop reinforcement learning will also be essential for practical legal applications. All things considered, this hybrid approach offers a promising platform for automated legal clause classification, with room to grow into a more robust and adaptable system in the future.

REFERENCES

[1] Bansal, N., Sharma, A., Singh, R.K. (2019). A Review on the Application of Deep Learning in Legal Domain. In: MacIntyre, J., Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds) Artificial Intelligence Applications and Innovations. AIAI 2019. IFIP Advances in Information and Communication Technology, vol 559. Springer, Cham. https://doi.org/10.1007/978-3-030-19823-7_31

[2] Chen, H., Wu, L., Lu, W., Chen, J., Ding, J. (2022). A comparative study of automated legal text classification using random forests and deep learning. Information Processing and Management, 59(2)

[3] Graziella De Martino, Gianvito Pio, and Michelangelo Ceci. 2023. Multi-view overlapping clustering for the identification of the subject matter of legal judgments. Inf. Sci. 638, C (Aug 2023). https://doi.org/10.1016/j.ins.2023.118956

[4] Ahmed Elnaggar, Christoph Gebendorfer, Ingo Glaser, and Florian Matthes. 2018. Multi-Task Deep Learning for Legal Document Translation, Summarization and Multi-Label Classification. In Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference (AICCC '18). Association for Computing Machinery, New York, NY, USA, 9–15. https://doi.org/10.1145/3299819.3299844

[5] R.Qin, M. Huang and Y. Luo, "A Comparison Study of Pre-trained Language Models for Chinese Legal Document Classification," 2022 5th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2022, pp. 444-449, doi: 10.1109/ICAIBD55127.2022.9820466.

[6] Riya Sil, Abhishek Roy. (2021). Machine Learning Approach for Automated Legal Text Classification. International Journal of Computer Information Systems and Industrial Management Applications, 13, 10. Retrieved from https://cspub-ijcisim.org/index.php/ijcisim/article/view/487

[7] Neha Bansal, Arun Sharma, R. K. Singh. A Review on the Application of Deep Learning in Legal Domain. 15th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), May 2019, Hersonissos, Greece. pp.374-381, ff10.1007/978-3-030-19823-$7_31ff.ffhal - 02331336f$

[8] Zhong, B., Shen, L., Pan, X., Zhong, X., He, W. (2024). Dispute Classification and Analysis: Deep Learning–Based Text Mining for Construction Contract Management. Journal of Construction Engineering and Management, 150(1), 04023151.

[9] R. Saha and S. Jyhne, "Interpretable Text Classification in Legal Contract Documents using Tsetlin Machines," 2022 International Symposium on the Tsetlin Machine (ISTM), Grimstad, Norway, 2022, pp. 7-12, doi: 10.1109/ISTM54910.2022.00011.

[10] M. Naeem, S. T. H. Rizvi and A. Coronato, "A Gentle Introduction to Reinforcement Learning and its Application in Different Fields" in IEEE Access, vol. 8, pp. 209320-209344, 2020, doi: 10.1109/ACCESS.2020.3038605.

[11] N. Huber-Fliflet et al., "Explainable Text Classification for Legal Document Review in Construction Delay Disputes," 2023 IEEE International Conference on Big Data (BigData), Sorrento, Italy, 2023, pp. 1928-1933, doi: 10.1109/BigData59044.2023.10386240.

[12] *R. Sil, A. Roy, B. Bhushan and A. K. Mazumdar, "Artificial Intelligence and Machine Learning based Legal Application: The State-of-the-Art and Future Research Trends," 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2019, pp. 57-62, doi: 10.1109/ICCCIS48478.2019.8974479.*