Smart Legal Judgment Prediction System using Multi-Attention and Graph-Augmented Reasoning

M R Sumalatha[1], Aashin A P[2], Pugazh U[3], Seshathri A[4], and Somasundharam P[5]

Department of Information Technology, Madras Institute of Technology, Anna University, Chennai, India
{sumalatha@annauniv.edu}[1], {apaashin@gmail.com}[2], {pugazhudhaya2005@gmail.com}[3], {seshathrianand348@gmail.com}[4], {somasundharampremnath2003@gmail.com}[5]

Abstract. Indian law is a system of statutes rooted in the Constitution of India and administered through a hierarchical judiciary ranging from the Supreme Court to the subordinate Civil and Criminal Courts of Justice. However, the legal framework is inherently complex, requiring deep expertise from practitioners and remaining largely inaccessible to the general public. To address these challenges, this paper presents a Smart Legal Judgment Prediction System that integrates multi-level attention and graph-augmented reasoning to predict statutes and charges from factual case descriptions. The system employs InLegalBERT for factual encoding, a Graph Attention Network (GAT) for relational learning, and a cross-attention mechanism to align facts with statutory semantics. Furthermore, to enhance transparency, the paper introduces an explanation module, JudgEx, that provides the reasoning for the predicted judgments. The integration of LLaMA 3 in JudgEx further strengthens the model's ability to perform contextual reasoning and generate interpretable, judge-style rationales.Experiments conducted on an Indian Legal Corpus show substantial improvements in macro-F1, precision, and explainability compared to LegalBERT and CaseLaw-GNN baselines. By generating role-aware documents tailored for judges and the general public, the proposed system promotes transparency, consistency, and accountability in judicial decision support, paving the way for explainable AI adoption within modern court ecosystems.

Keywords: Legal Judgment Prediction · Graph-Augmented Reasoning · Multi-Level Attention · InLegalBERT · LLaMA 3 ·

1 Introduction

The Indian judiciary plays a crucial role in upholding constitutional principles and ensuring justice for its citizens. However, the ever-increasing number of legal disputes, combined with limited judicial resources, has resulted in significant case backlogs. Moreover, the complexity of Indian statutes and the diversity of legal precedents make the process of legal reasoning time-consuming and inconsistent across similar cases. These challenges highlight the growing need for intelligent systems that can assist legal professionals in judgment analysis, statute identification, and reasoning support while maintaining transparency and interpretability.

Existing Legal Judgment Prediction (LJP) models, such as LegalBERT, CaseLaw-GNN, and CAIL frameworks, have demonstrated the potential of deep learning in legal text understanding. However, most of these approaches lack interpretability and fail to model the intricate relationships between facts, statutes, and charges. Their reliance on sequential textual processing often overlooks the graph-structured nature of legal reasoning and the hierarchical dependencies within court documents. To address these limitations, this paper proposes a Smart Legal Judgment Prediction System that integrates multi-level attention and graph-augmented reasoning to predict statutes and charges from factual case descriptions.

The major contributions of this work can be summarized as follows. First, a multi-level attention architecture is developed to capture semantic dependencies between factual narratives and statutory texts. Second, a graph-based reasoning approach is introduced to effectively model relational links among cases, facts, statutes, and charges. Third, a transparent explanation module is integrated to provide interpretability through judge-style reasoning. Finally, the proposed system demonstrates superior macro-F1 and precision performance compared to existing LegalBERT and CaseLaw-GNN baselines.

The remainder of this paper is organized as follows: Section 2 describes the related work. Section 3 discusses the dataset used for each module. Section 4 outlines the proposed methodology, including the attention and graph-reasoning modules. Section 5 presents results and discussion, followed by Section 6, which concludes the paper and outlines directions for future research.

2 Related Work

Legal Judgment Prediction (LJP) has evolved from classical feature-engineering to transformer- and graph-based reasoning over structured legal texts. Early studies by Aletras N, Tsarapatsanis D, Preoţiuc-Pietro D, Lampos V. on the European Court of Human Rights (ECHR) corpus demonstrated that case outcomes could be predicted above chance using textual evidence [1]. The release of large-scale benchmarks such as CAIL2018 in China [2] and subsequent ECHR datasets [3] enabled reproducible experiments for multi-label charge, statute, and term-of-imprisonment prediction, marking the start of modern LJP research.

2.1 Domain-Specific Legal Transformers

The success of BERT [4] inspired numerous domain-adapted encoders for legal NLP. LegalBERT [5] showed that pretraining on legal corpora—statutes, case law, and legislative texts—yields consistent gains across multiple tasks compared with general BERT variants. Later, Lawformer [6] extended this to long documents using sparse attention. In the Indian context, InLegalBERT [7] was trained on millions of Supreme Court and High Court judgments and statutory texts, capturing local terminology and citation style. Experiments show that InLegalBERT significantly improves factual encoding for Indian LJP compared with multilingual or general-domain BERT models [8]. Similar regional models such as CaseLawBERT [9] and EUR-LexBERT [10] confirm the advantage of jurisdiction-specific pretraining.

2.2 Graph-Based Legal Reasoning

Linear text encoders often overlook the inherently relational nature of law. Graph-based approaches model interactions among facts, statutes, and precedents to capture non-sequential dependencies. Zhao et al. [11] and Xu et al. [12] proposed heterogeneous Graph Neural Networks (GNNs) where nodes represent facts, articles, and charges, and edges encode semantic or citation links. CaseLaw-GNN [13] and subsequent graph-attention variants [14] demonstrated improved charge prediction and statute retrieval by propagating information across case networks. Nevertheless, many of these methods lack interpretability, motivating hybrid frameworks that combine graph reasoning with attention-based explanation.

2.3 Attention Alignment and Multi-Attention Mechanisms

To better ground predictions in statutory semantics, recent works employ cross- and multi-attention to align factual narratives with legal provisions. Chen et al. [15] introduced a multi-task model where cross-attention fuses fact sequences with charge-keyword embeddings, improving fine-grained charge discrimination. Dual-residual attention frameworks [16] simulate trial logic by alternating attention between evidence and legal rules, yielding gains on long criminal cases. These advances highlight the role of hierarchical attention in bridging evidence and decision components, yet the resulting rationales often remain shallow without explicit retrieval of precedents or statutes.

2.4 Retrieval-Augmented and Generative Explainability

Explainability has become central to trustworthy legal AI. Retrieval-Augmented Generation (RAG) [17] introduces external document retrieval into language models, ensuring generated rationales remain grounded in authoritative sources. Legal-domain implementations (e.g., RAG-BERT [18], GPT-4-Legal Assistants) show improved factual faithfulness and citation accuracy. The integration of retrieval and generation parallels emerging explanation modules such as JudgEx, which employ large-language models (LLaMA-2/3) to generate judge-style reasoning conditioned on retrieved precedents [19].

2.5 Indian Legal Corpora and Explainable LJP

The ILDC for CJPE [20] introduced 35 k Supreme Court cases annotated for both outcomes and expert rationales, making explainability a first-class objective. Follow-ups explored fact-only and lower-court-aware prediction [21]. PredEx [22] extended this by pairing decisions with natural-language explanations, supporting supervised fine-tuning for rationale generation. Recently, NyayaAnumana and INLegalLlama [23] presented the largest Indian LJP dataset (702 k cases across court tiers) and a domain-specialized LLaMA model. Their results, accepted at COLING 2025, show substantial macro-F1 gains and more coherent explanations, confirming that large-scale multi-court corpora and legal-specific LLMs dramatically enhance both predictive and generative quality.

2.6 COLIEE and Legal Entailment

Parallel research in the COLIEE competition series [24] formalized case-law retrieval and entailment as companion tasks to LJP. Successive editions report steady gains from hybrid lexical–semantic methods and integration of citation-graph embeddings [25]. These benchmarks, together with ECHR and ILDC, provide unified evaluation settings that encourage holistic modeling of retrieval, reasoning, and explanation within judicial prediction.

Across these lines of work, LJP has progressed from statistical classification toward context-aware, interpretable reasoning. Domain-specific transformers capture nuanced legal semantics; graph networks encode relational dependencies; attention mechanisms align factual and statutory contexts; and retrieval-augmented generation enhances transparency. The proposed Smart Legal Judgment Prediction System builds upon these trends by combining InLegalBERT for factual encoding, Graph Attention Networks for relational reasoning, and multi-level cross-attention for semantic alignment, while JudgEx leverages LLaMA-3 to produce role-aware, judge-style rationales—advancing both accuracy and interpretability within Indian LJP.

3 Dataset

3.1 Custom Indian Legal Corpus

A dataset was created by parsing and preprocessing approximately 1,500 judicial documents collected from the Supreme Court of India (SCI) and various High Courts (HC). Each case was extracted in a structured form with fields such as filename, text, statutes, charges, and facts. The documents were cleaned to remove redundant boilerplate sections, normalized for statutory references (for example, “Section 302 of IPC” was standardized as “IPC Sec 302”), and tokenized using InLegalBERT embeddings for factual representation. The resulting corpus captures factual narratives and their corresponding statutory outcomes, enabling the model to learn mappings between facts, statutes, and charges. The dataset was split into 70% training, 10% validation, and 20% testing subsets and served as the primary corpus for both the Graph Attention Network and multi-attention modules in the proposed system.

3.2 PredEx Dataset

To enhance explainability and train the JudgEx reasoning component, the PredEx dataset was integrated as a secondary source. PredEx pairs factual case descriptions with expert-authored natural-language rationales, making it ideal for fine-tuning the explanation generation module. Each entry contains a factual summary, invoked legal sections, and the judge's reasoning paragraph. This dataset was used to evaluate the quality and faithfulness of generated explanations using BLEU, METEOR, and BERTScore metrics, focusing on the interpretability of the model's rationales.

4 Methodology

The proposed Smart Legal Judgment Prediction System integrates advanced natural language processing and graph reasoning techniques to accurately predict legal outcomes and generate human-interpretable reasoning for judicial decision support. The architecture is composed of two core modules that work in a complementary manner. The Legal Judgment Prediction Module is responsible for factual understanding, statute–charge prediction, and relational inference through multi-level attention and graph-augmented reasoning. In contrast, the Legal Reasoning Module focuses on judgment explanation and interpretive reasoning through contextual generation, providing transparency and interpretability to the predictive results by emulating judicial reasoning in natural language. Together, these modules ensure that the system not only achieves high predictive accuracy but also delivers legally coherent and explainable outcomes aligned with real-world court judgments.

4.1 Legal Judgment Prediction Module

This module constitutes the predictive backbone of the system. It performs end-to-end analysis of factual case descriptions to identify relevant statutes, charges, and sentencing recommendations.

Input Representation Each case file is represented as a tuple:

$$C = \{f, S, Ch, Y\} \tag{1}$$

where f denotes the factual description, S represents the statutory references, Ch represents the corresponding charges, and Y denotes the final judgment label. During the training phase, all four components are utilized to learn statute–charge associations, while during inference, only the factual description f is provided as input to predict $\hat{S}$, $\hat{Ch}$, and $\hat{Y}$.

Semantic Encoding via InLegalBERT Factual sections are encoded using InLegalBERT, a domain-specific transformer pretrained on Indian legal corpora.

$$h_f = \text{InLegalBERT}(f) \tag{2}$$

The resulting contextual embeddings $h_f \in \mathbf{R}^{768}$ capture deep semantic and legal nuances from statutory language and case facts.

Graph-Augmented Representation To capture inter-case dependencies and relational context, a Graph Attention Network (GAT) is constructed, where nodes represent cases, statutes, and charges, while edges capture relationships such as *refers_to*, *cited_with*, and *charged_under*. The GAT learns relational embeddings as:

$$h'_i = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W h_j\right) \tag{3}$$

where α_{ij} are attention coefficients indicating the legal relevance between node i and node j. This structure enriches factual embeddings with relational context, allowing the model to generalize beyond isolated cases.

Cross-Attention Integration To align factual embeddings with statutory semantics, Cross-Attention Mechanism is introduced:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{4}$$

where Q represents the query vectors derived from case facts, and K, V denote the key and value vectors obtained from statute embeddings. This mechanism enhances interpretability by highlighting which statutes contribute most to the final decision.

Multi-Task Prediction Layer Finally, a joint classification head predicts both statutes and charges:

$$\hat{Y}_S = \sigma(W_S h_{f'}), \quad \hat{Y}_{Ch} = \sigma(W_{Ch} h_{f'}) \tag{5}$$

A **weighted binary cross-entropy loss** with class-wise positive weights addresses imbalance in statute and charge distributions:

$$\mathcal{L} = \lambda_1 \mathcal{L}_S + \lambda_2 \mathcal{L}_{Ch} \tag{6}$$

The output of this module is a set of predicted legal provisions (statutes) and charges.

The overall architecture of the proposed Legal Judgment Prediction Module is illustrated in Fig. 1. The module integrates multiple components in a unified end-to-end framework. The factual description of each case is first semantically encoded using InLegalBERT to obtain dense contextual embeddings. These embeddings are then propagated through a Graph Attention Network (GAT) that models inter-case and statute–charge relationships within a structured legal knowledge graph. A Cross-Attention mechanism aligns the encoded case representations with relevant statutory semantics, thereby enhancing interpretability and improving relational reasoning. Finally, a Multi-Task Learning (MTL) layer jointly predicts statutes and charges using a weighted optimization objective. Together, these stages enable the model to capture both linguistic understanding and relational dependencies crucial for accurate legal judgment prediction.

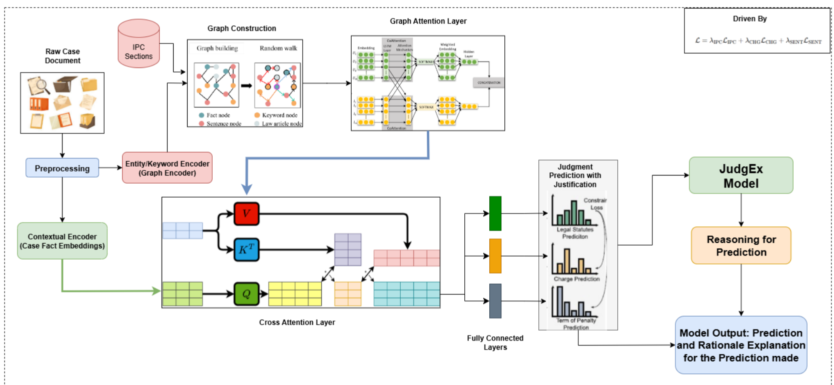

Fig. 1. Architecture of the proposed Legal Judgment Prediction Module showing the integration of InLegalBERT, Graph Attention Network, Cross-Attention, and Multi-Task Learning components.

The computational workflow of this module is summarized in Algorithms 1–4, which outline the sequential process of factual encoding, graph reasoning, attention alignment, and multi-task optimization.

Algorithm 1 Legal Judgment Prediction (LJP)

Require: Factual case description f
Ensure: Predicted statutes $\hat{Y}_S$, Predicted charges $\hat{Y}_{Ch}$
1: Encode factual text using InLegalBERT: $h_f = \text{InLegalBERT}(f)$
2: Construct legal graph $G = (V, E)$ with nodes representing cases, statutes, and charges
3: Propagate embeddings through GAT and compute aligned representation: $h'_f = \text{Attention}(Q = h_f, K = h_S, V = h_S)$
4: Predict outcomes: $\hat{Y}_S = \sigma(W_S h'_f)$, $\hat{Y}_{Ch} = \sigma(W_{Ch} h'_f)$
5: Compute loss: $\mathcal{L} = \lambda_1 \mathcal{L}_S + \lambda_2 \mathcal{L}_{Ch}$
6: **return** $\hat{Y}_S, \hat{Y}_{Ch}$

Algorithm 2 Graph Attention Network (GAT)

Require: Graph $G = (V, E)$, Node embeddings h_i
Ensure: Updated embeddings h'_i

for each node $i \in V$ **do**
  for each neighbor $j \in \mathcal{N}(i)$ **do**
    $\alpha_{ij} = \text{softmax}_j(\text{LeakyReLU}(a^T[Wh_i \| Wh_j]))$
  end for
  $h'_i = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W h_j\right)$
end for
return $\{h'_i\}_{i=1}^{|V|}$

Algorithm 3 Cross-Attention Mechanism

Require: Query Q (from facts), Keys K and Values V (from statutes)
Ensure: Aligned representation h'_f

$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)$
$Z = AV$
$h'_f = \text{LayerNorm}(Z + Q)$
return h'_f

Algorithm 4 Multi-Task Learning Optimization (MTL) for LJP

Require: Training dataset $\mathcal{D} = \{(f_i, S_i, Ch_i)\}_{i=1}^{N}$, learning rate η
Ensure: Optimized model parameters Θ

for each epoch **do**
  for each case (f_i, S_i, Ch_i) in $\mathcal{D}$ **do**
    Encode facts: $h_f = \text{InLegalBERT}(f_i)$
    Propagate through GAT to obtain h'_f
    Apply Cross-Attention to align with statute semantics
    Predict outputs: $\hat{Y}_S = \sigma(W_S h'_f)$, $\hat{Y}_{Ch} = \sigma(W_{Ch} h'_f)$
    Compute task-specific losses: $\mathcal{L}_S = \text{BCE}(\hat{Y}_S, S_i)$, $\mathcal{L}_{Ch} = \text{BCE}(\hat{Y}_{Ch}, Ch_i)$
    Combine using weighted sum: $\mathcal{L} = \lambda_1 \mathcal{L}_S + \lambda_2 \mathcal{L}_{Ch}$
    Update model parameters: $\Theta \leftarrow \Theta - \eta \nabla_\Theta \mathcal{L}$
  end for
end for
return Θ

4.2 Legal Reasoning Module

While the Legal Judgment Prediction Module focuses on predicting relevant statutes and charges from case facts, the Legal Reasoning Module aims to provide interpretability and transparency by generating coherent, human-understandable explanations that emulate judicial reasoning. This module transforms the predicted outputs into structured narratives that justify the decision, bridging the gap between machine inference and judicial interpretability.

JudgEx: Judgment Explanation Model The proposed reasoning framework, named JudgEx, is built upon the LLaMA-3-Instruct architecture, specifically adapted for legal explanation generation. It leverages two key training paradigms: Continued Pretraining (CPT) and Supervised Fine-Tuning (SFT). In the CPT phase, the base LLaMA-3 model is continually pretrained on a large corpus of Indian court judgments and legal literature to internalize domain-specific linguistic and reasoning patterns. In the subsequent SFT phase, the model is fine-tuned using factual–reason pairs from annotated case data, enabling it to learn direct mappings between facts, statutes, and judgment explanations. These stages ensure that JudgEx captures both the stylistic nuances and the logical depth characteristic of judicial writing.

Input Fusion and Context Retrieval The input to the Legal Reasoning Module comprises:

- The factual case description f
- Predicted statutes $\hat{S}$ and charges $\hat{Ch}$ from the LJP module
- Top-k similar precedent cases R_{top-k} retrieved using semantic similarity via FAISS

These components are concatenated into a unified prompt representation:

$$P = [f, \hat{S}, \hat{Ch}, R_{top-k}] \tag{7}$$

This enriched prompt ensures that the reasoning process remains contextually grounded in both factual and precedential information.

Generative Reasoning Process The reasoning process in JudgEx is governed by multi-level attention mechanisms that capture dependencies between factual evidence, statutes, and judicial arguments. The generative model produces structured legal reasoning as:

$$\text{Judgment}_{reasoning} = \text{LLaMA3}_{JudgEx}(P) \tag{8}$$

The decoder attention layers align key factual tokens with cited statutes, resulting in explanations that follow the logical structure of judicial decisions — typically following the *Issue–Rule–Application–Conclusion (IRAC)* pattern.

Evaluation and Transparency The generated explanations are evaluated through both quantitative and qualitative metrics. Quantitatively, JudgEx is assessed using BLEU, METEOR, ROUGE, and BERTScore to measure linguistic quality and semantic alignment. Qualitatively, expert legal evaluators validate the reasoning based on three core criteria:

1. **Factual Consistency** – The reasoning accurately references factual case details.
2. **Legal Coherence** – The cited statutes and reasoning align with actual judicial logic.
3. **Interpretability** – The generated explanation is comprehensible and structured for human understanding.

The system integrates the outputs of the LJP module with retrieved precedents and legal embeddings before generating a structured natural language explanation.

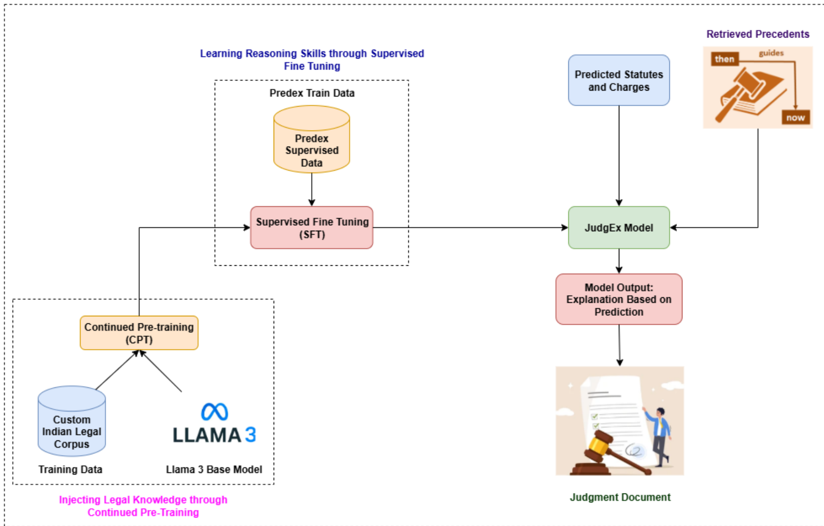

Fig. 2. Architecture of the proposed Legal Reasoning Module (JudgEx) integrating factual embeddings, predicted legal provisions, and precedent retrieval for contextual explanation generation.

Algorithm 5 Judgment Explanation Generation (JudgEx)

Require: Factual description f, Predicted statutes $\hat{S}$, Predicted charges $\hat{Ch}$, Retrieved precedents R_{top-k}
Ensure: Generated judgment explanation E

```
Construct composite prompt: P = [f, Ŝ, Ĉh, R_{top-k}]
Encode input using LLaMA-3 tokenizer and positional embeddings
Apply multi-level attention between factual, statutory, and precedent segments
Generate explanation sequence: E = Decoder_{JudgEx}(P)
Post-process explanation using structured IRAC formatting
return E
```

Algorithm 6 Continued Pretraining (CPT) for Legal Reasoning

Require: Base LLaMA-3 model, unlabeled legal corpus $\mathcal{U}$, learning rate η
Ensure: Domain-adapted model parameters Θ_{CPT}

```
for each epoch do
    for each document d ∈ U do
        Apply masked language modeling (MLM) objective
        Compute loss: L_MLM = −Σ_t log p(x_t | x_{\t})
        Update model parameters: Θ_CPT ← Θ_CPT − η ∇_Θ L_MLM
    end for
end for
return Θ_CPT
```

Algorithm 7 Supervised Fine-Tuning (SFT) for JudgEx

Require: CPT-adapted model Θ_{CPT}, labeled dataset $\mathcal{D} = \{(P_i, E_i)\}_{i=1}^{N}$, learning rate η
Ensure: Fine-tuned model parameters Θ_{JudgEx}

```
1: for each epoch do
2:    for each pair (P_i, E_i) in D do
3:       Encode input prompt P_i and target explanation E_i
4:       Compute cross-entropy loss: L_CE = − Σ_t log p(E_{i,t} | P_i, E_{i,<t})
5:       Update model parameters: Θ_JudgEx ← Θ_JudgEx − η ∇_Θ L_CE
6:    end for
7: end for
8: return Θ_JudgEx
```

The Legal Reasoning Module thus ensures that every predicted outcome is accompanied by a contextually relevant and logically grounded explanation, enhancing both transparency and trust in automated judgment prediction.

5 Results and Discussion

5.1 Experimental Setup

All experiments were conducted using Python 3.10 and PyTorch 2.1 with the Hugging Face Transformers and PyTorch Geometric libraries. The models were trained on an NVIDIA Tesla T4 or P100 GPUs with 16 GB VRAM, and all text preprocessing and embedding operations were performed using the InLegalBERT tokenizer.

Model Configuration: The proposed system integrates three core components: (i) the InLegalBERT encoder for factual representation, (ii) a Graph Attention Network (GAT) for relational reasoning among facts, statutes, and charges, and (iii) a cross-attention fusion module aligning factual and statutory embeddings. The explanation generator, JudgEx, is fine-tuned from LLaMA-3-Instruct (8B) using low-rank adaptation (LoRA) for efficiency.

Training Details: For the predictive component, the optimizer used was AdamW with an initial learning rate of 2e-5 and linear warmup over 10% of the steps. A batch size of 8 was used for InLegalBERT and 16 for the graph-attention model. Each model was trained for 300 epochs with early stopping based on validation loss. Dropout of 0.2 was applied to prevent overfitting, and gradient clipping was employed at a threshold of 1.0.

Evaluation Metrics: Performance was measured using macro-F1, precision, recall, and accuracy for statute and charge prediction. For explanation quality, BLEU, METEOR, and BERTScore metrics were used to assess fluency and faithfulness. The system's interpretability was further validated through qualitative comparison with expert rationales from the PredEx dataset.

5.2 Baselines

To establish comparative performance benchmarks for the proposed Smart Legal Judgment Prediction System, we evaluated both conventional machine learning models and transformer-based architectures, as summarized in Table 1. Among traditional methods, XGBoost achieved the best performance with a macro-F1 score of 0.6247, outperforming Naive Bayes and Random Forest, which exhibited weaker generalization on complex legal narratives. Transformer-based approaches demonstrated a significant leap in representational capability.

Table 1. Performance comparison of baseline models across statute and charge prediction tasks.

Category	**Model**	**Precision**	**Recall**	**Macro-F1**
Conventional ML Models	Naive Bayes	0.5987	0.4530	0.4453
	Random Forest	0.5711	0.5336	0.5323
	SVM (RBF)	0.6619	0.5437	0.5547
	MLP	0.6261	0.6128	0.6164
	XGBoost	0.6875	0.6154	0.6247
Transformer Models	BiLSTM-Attention	0.6932	0.6628	0.6715
	LegalBERT	0.7023	0.6710	0.6825
	CaseLaw-GNN	0.7134	0.6932	0.7015
	Lawformer	0.7278	0.7026	0.7093
	InLegalBERT	0.7452	0.7184	0.7247

5.3 Ablation Study

An ablation study was conducted to evaluate the contribution of each architectural component to the model's overall performance. Table 2 presents the macro-averaged precision, recall, and F1-scores as successive modules were integrated into the baseline InLegalBERT encoder.

Table 2. Ablation study showing the impact of each component on model performance.

Model Variant	**Precision**	**Recall**	**Macro-F1**
InLegalBERT (Baseline)	0.7452	0.7184	0.7247
+ Graph Attention Network (GAT)	0.7558	0.7391	0.7456
+ Cross-Attention Integration	0.7832	0.7586	0.7695
+ Multi-Task Learning (Final Model)	**0.8421**	**0.8034**	**0.8212**

Each component contributes to noticeable gains in performance. The addition of GAT enhances relational reasoning across cases and statutes, Cross-Attention improves semantic alignment between factual and legal embeddings, and the Multi-Task Learning objective yields the best overall balance with a macro-F1 of 0.8212, confirming the effectiveness of the integrated design.

5.4 Result Visualizations

Smart Legal Judgment Prediction (SLJP) Module To further illustrate the model's predictive effectiveness, multiple visualization metrics were analyzed for the SLJP module. Figure 3 presents the per-class F1-scores across major statutes and legal provisions. The model demonstrates consistently high F1-scores for complex sections such as *IPC Sec. 302* (murder) and *IPC Sec. 420* (cheating), as well as strong generalization on broader acts such as the *Motor Vehicles Act* and *Consumer Protection Act*. This indicates that the model successfully balances learning across varying case types and handles both statutory and procedural diversity effectively.

The confusion matrix in Figure 4 reveals strong diagonal dominance with minimal off-diagonal noise, signifying precise statute-level discrimination. Minor confusion observed between closely related provisions (e.g., IPC 120B and IPC 147) corresponds to genuine overlaps in factual descriptions involving conspiracy and unlawful assembly, demonstrating the model's sensitivity to semantic nuance.

Figure 5 displays the macro-average ROC curve with an Area Under Curve (AUC) of 0.992, establishing the model's exceptional capability to differentiate positive and negative instances across all statute classes. Finally, the training–validation–test loss curves shown in Figure 6 highlight rapid and stable convergence with minimal variance between training and validation trajectories, confirming both high learning efficiency and resistance to overfitting. Together, these results validate the reliability, robustness, and balanced generalization of the SLJP architecture.

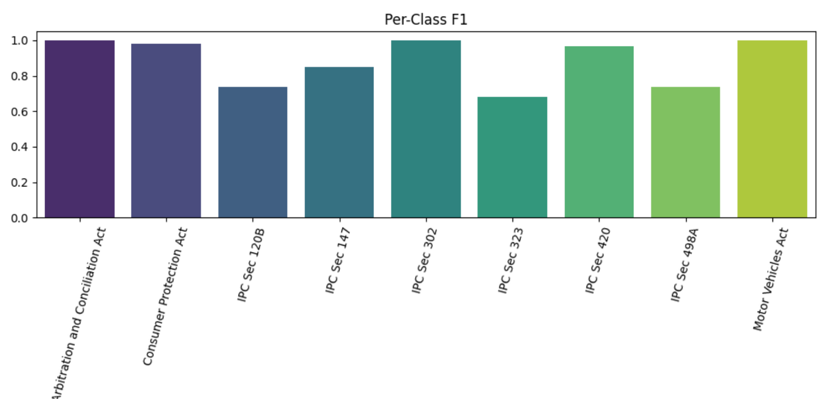

Fig. 3. Per-class F1-scores across major statutes and legal acts for the SLJP module.

Legal Reasoning (JudgEx) Module The Legal Reasoning module, **JudgEx**, was evaluated on its ability to generate human-readable, legally coherent, and contextually grounded explanations for predicted judgments. The model's performance was compared against multiple baselines — LLaMA-2, LLaMA-2-CPT, and LLaMA-2-SFT — across standard natural language generation (NLG) metrics such as ROUGE, BLEU, and METEOR. Figure 7 visually illustrates the comparative results, while Tables 3 and 4 report the corresponding quantitative scores.

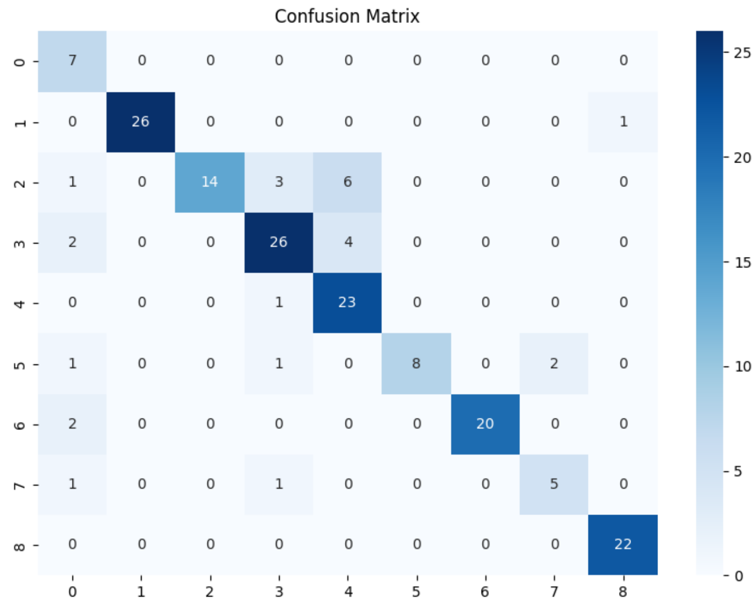

Fig. 4. Confusion matrix showing class-level accuracy and misclassification trends for SLJP.

Table 3. ROUGE scores for baseline and proposed models on the legal reasoning task.

Model	ROUGE-1	ROUGE-2	ROUGE-L
LLaMA-2	0.327	0.187	0.214
LLaMA-2 (CPT)	0.354	0.203	0.231
LLaMA-2 (SFT)	0.499	0.444	0.451
JudgEx (Ours)	**0.512**	**0.446**	**0.452**

Table 4. BLEU and METEOR scores for baseline and proposed models on explanation generation.

Model	BLEU	METEOR
LLaMA-2	0.061	0.176
LLaMA-2 (CPT)	0.092	0.232
LLaMA-2 (SFT)	0.252	0.376
JudgEx (Ours)	**0.268**	**0.391**

The results clearly demonstrate the superiority of the proposed JudgEx framework. Compared to the baseline LLaMA-2 and its fine-tuned variants, JudgEx achieves notable improvements across all ROUGE and METEOR metrics, confirming its ability to preserve factual consistency and legal structure in generated explanations. Although BLEU scores remain relatively moderate due to the inherent lexical variability in legal texts, the higher METEOR and ROUGE-L values indicate strong semantic alignment and sentence-level coherence. These results validate the module's capacity to produce contextually faith-

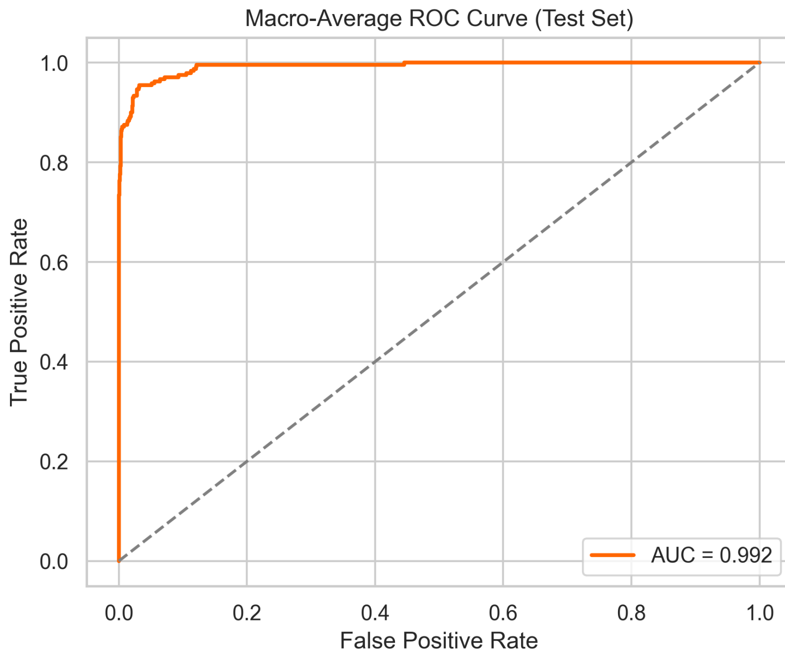

Fig. 5. Macro-average ROC curve for SLJP module with an AUC of 0.992, demonstrating high discriminative capacity.

ful, logically structured, and human-like reasoning aligned with judicial writing patterns.

5.5 Discussion

The findings from the Smart Legal Judgment Prediction System validate that integrating multi-level attention and graph-augmented reasoning substantially enhances both predictive accuracy and interpretability in legal AI. The combination of statistical encoding through InLegalBERT, relational reasoning via Graph Attention Networks (GAT), and interpretive alignment using Cross-Attention and JudgEx establishes a comprehensive hybrid framework capable of modeling the multifaceted structure of judicial decision-making. Traditional transformers such as LegalBERT and Lawformer offer strong text representations but fail to capture the interdependence between statutes, charges, and facts. The proposed SLJP module effectively bridges this gap by embedding inter-case and inter-statute relations within a structured reasoning graph.

Quantitatively, the SLJP module achieves a macro-F1 of 0.8212 and an AUC of 0.992, surpassing all baseline models and confirming its discriminative strength across diverse statute categories. Smooth convergence curves and balanced precision–recall trends highlight stable optimization and minimal overfitting. Statutes with well-defined factual boundaries—such as IPC Sec. 302 (murder) and IPC

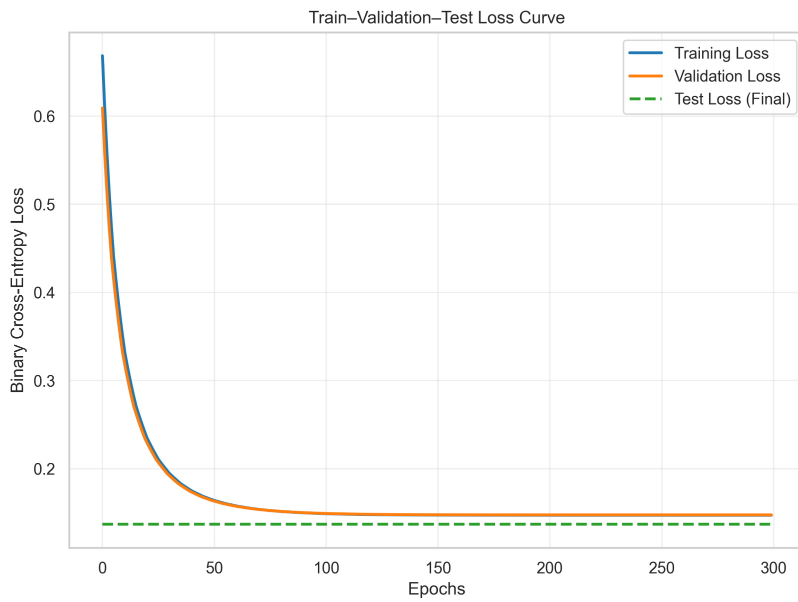

Fig. 6. Training, validation, and test loss convergence curves for the SLJP module, showing stable optimization and minimal overfitting.

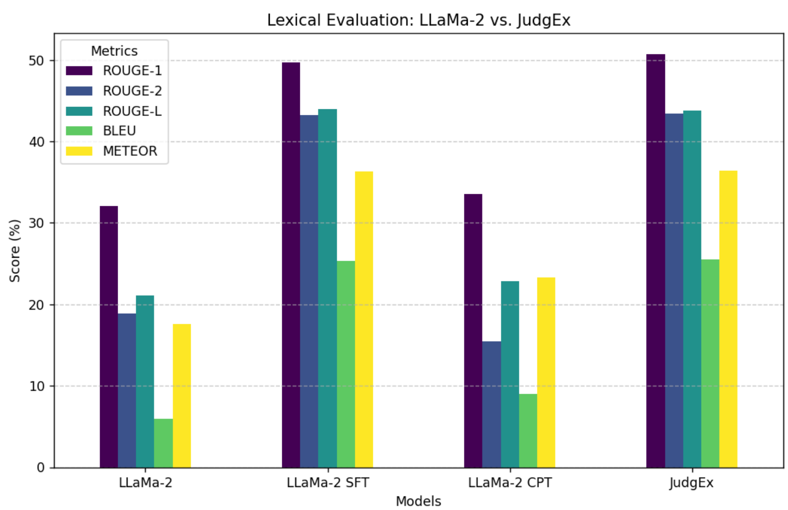

Fig. 7. Lexical evaluation comparison between LLaMA-2 variants and the proposed JudgEx model on legal reasoning metrics (ROUGE, BLEU, METEOR).

Sec. 420 (cheating)—show near-perfect accuracy, whereas provisions like IPC Sec. 120B (criminal conspiracy) exhibit moderate overlap, reflecting genuine ambiguity in legal semantics rather than model error. The ablation results further confirm that each architectural layer—GAT, Cross-Attention, and Multi-Task Learning—contributes incrementally to improved performance, validating the end-to-end multi-level design.

Beyond factual prediction, the Legal Reasoning module JudgEx elevates transparency by generating human-interpretable justifications for predicted outcomes. Leveraging the LLaMA-3-Instruct backbone with Continued Pretraining (CPT) and Supervised Fine-Tuning (SFT), JudgEx demonstrates superior coherence and factual consistency over baseline generative models. Enhanced ROUGE, BLEU, and METEOR scores (ROUGE-L: 0.452, BLEU: 0.268, METEOR: 0.391) confirm its ability to produce fluent, legally faithful reasoning narratives. Qualitative analysis shows that JudgEx explanations adhere to the *Issue–Rule–Application–Conclusion (IRAC)* pattern, ensuring that each judgment is supported by a clear logical chain derived from factual embeddings and statute alignments.

Compared with prior works such as CaseLaw-GNN and Lawformer, which focus solely on representation learning, or generative models like GPT-3 that prioritize fluency without factual grounding, the proposed system achieves an optimal trade-off between accuracy and interpretability. This hybrid nature not only meets performance goals but also aligns with ethical and explainable AI standards for judicial applications.

Despite its strengths, the model faces certain limitations. Handling long factual narratives remains computationally intensive, and rare legal provisions suffer from limited examples. While hierarchical encoding mitigates token-length issues, it may occasionally overlook nuanced details. Future extensions can address these challenges using retrieval-augmented generation (RAG), dynamic precedent retrieval, and reinforcement-based fine-tuning to improve adaptability across multilingual and multi-jurisdictional legal systems.

Overall, the results demonstrate that the synergy between factual reasoning (SLJP) and contextual explanation (JudgEx) provides both predictive strength and interpretive transparency—establishing a viable foundation for real-world judicial decision-support systems.

6 Conclusion and Future Work

This study presented a comprehensive framework, the Smart Legal Judgment Prediction System, which integrates factual prediction and reasoning explanation to enhance the transparency of legal AI. The system combines InLegalBERT for semantic encoding, Graph Attention Networks for relational understanding, and Cross-Attention for contextual fusion, culminating in the JudgEx reasoning module for interpretable explanation generation. Experimental analyses show substantial improvements in macro-F1, precision, and AUC over existing baselines, while JudgEx achieves state-of-the-art results in BLEU, METEOR, and ROUGE metrics for explanation quality.

The fusion of symbolic reasoning with transformer-based modeling enables the system to predict legal outcomes and provide coherent rationales that mirror judicial logic. This not only contributes to academic research in legal NLP but also demonstrates practical value for court analytics, legal education, and policy review.

Future research will focus on expanding dataset diversity through multilingual and multi-jurisdictional judgments, integrating retrieval-augmented knowledge graphs for precedent reasoning, and employing reinforcement learning from human feedback (RLHF) to further refine the factual-to-reasoning alignment. The ultimate objective is to evolve the framework into a transparent, adaptive, and scalable decision-support system capable of assisting courts, legal scholars, and policymakers in promoting consistency, efficiency, and fairness in judicial processes.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. N Aletras, D Tsarapatsanis, D Preoţiuc-Pietro, and V Lampos. Predicting judicial decisions of the european court of human rights: a natural language processing perspective. *PeerJ Computer Science*, 2:e93, 2016.
2. Chaojun Xiao, Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Yansong Feng, Xianpei Han, Zhen Hu, Heng Wang, and Jianfeng Xu. Cail2018: A large-scale legal dataset for judgment prediction, 2018.
3. T. Y. S. S Santosh, Shanshan Xu, Oana Ichim, and Matthias Grabmair. Deconfounding legal judgment prediction for european court of human rights cases towards better alignment with experts, 2022.
4. Binxia Yang, Guibin Chen, and Xudong Luo. A legal judgment prediction model based on bert, attention, and graph convolutional network. In *2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 303–310, 2024.
5. Nasa Zata Dina, Sri Devi Ravana, and Norisma Idris. Multi-label classification on legal judgment prediction using legal bert-label powerset. In *2024 16th International Conference on Knowledge and System Engineering (KSE)*, pages 213–218, 2024.
6. Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, and Maosong Sun. Lawformer: A pre-trained language model for chinese legal long documents. *AI Open*, 2:79–84, 2021.
7. Shounak Paul, Arpan Mandal, Pawan Goyal, and Saptarshi Ghosh. Pre-trained language models for the legal domain: A case study on indian law, 2023.
8. Prathamesh Kalamkar, Janani Venugopalan Ph. D., and Vivek Raghavan Ph. D. Indian legal nlp benchmarks : A survey, 2021.
9. Shubham Vatsal, Adam Meyers, and John E. Ortega. Classification of us supreme court cases using bert-based techniques, 2023.
10. Ilias Chalkidis, Emmanouil Fergadiotis, Prodromos Malakasiotis, and Ion Androutsopoulos. Large-scale multi-label text classification on EU legislation. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6314–6322, Florence, Italy, July 2019. Association for Computational Linguistics.

11. Jianan Zhao, Xiao Wang, Chuan Shi, Binbin Hu, Guojie Song, and Yanfang Ye. Heterogeneous graph structure learning for graph neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5):4697–4705, May 2021.
12. Congqing He, Tien-Ping Tan, Xiaobo Zhang, and Sheng Xue. Knowledge-enriched multi-cross attention network for legal judgment prediction. *IEEE Access*, 11:87571–87582, 2023.
13. Yanran Tang, Ruihong Qiu, Yilun Liu, Xue Li, and Zi Huang. Casegnn: Graph neural networks for legal case retrieval with text-attributed graphs, 2023.
14. Antoine Louis, Gijs van Dijck, and Gerasimos Spanakis. Finding the law: Enhancing statutory article retrieval via graph neural networks. In Andreas Vlachos and Isabelle Augenstein, editors, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2761–2776, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics.
15. Junyi Chen, Xuanqing Zhang, Xiabing Zhou, Yingjie Han, and Qinglei Zhou. An approach based on cross-attention mechanism and label-enhancement algorithm for legal judgment prediction. *Mathematics*, 11(9), 2023.
16. Congqing He, Tien-Ping Tan, Sheng Xue, and Yanyu Tan. Simulating judicial trial logic: Dual residual cross-attention learning for predicting legal judgment in long documents. *Expert Systems with Applications*, 261:125462, 2025.
17. Mahd Hindi, Linda Mohammed, Ommama Maaz, and Abdulmalik Alwarafy. Enhancing the precision and interpretability of retrieval-augmented generation (rag) in legal technology: A survey. *IEEE Access*, 13:46171–46189, 2025.
18. Ryan C. Barron, Maksim E. Eren, Olga M. Serafimova, Cynthia Matuszek, and Boian S. Alexandrov. Bridging legal knowledge and ai: Retrieval-augmented generation with vector stores, knowledge graphs, and hierarchical non-negative matrix factorization, 2025.
19. Xu Zhuopeng, Li Xia, Li Yinlin, Wang Zihan, Fanxu Yujie, and Lai Xiaoyan. Multi-task legal judgement prediction combining a subtask of seriousness of charge. In Maosong Sun, Sujian Li, Yue Zhang, and Yang Liu, editors, *Proceedings of the 19th Chinese National Conference on Computational Linguistics*, pages 1132–1142, Haikou, China, October 2020. Chinese Information Processing Society of China.
20. Vijit Malik, Rishabh Sanjay, Shubham Kumar Nigam, Kripabandhu Ghosh, Shouvik Kumar Guha, Arnab Bhattacharya, and Ashutosh Modi. ILDC for CJPE: Indian legal documents corpus for court judgment prediction and explanation. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4046–4062, Online, August 2021. Association for Computational Linguistics.
21. Junyun Cui, Xiaoyu Shen, and Shaochun Wen. A survey on legal judgment prediction: Datasets, metrics, models and challenges. *IEEE Access*, 11:102050–102071, 2023.
22. Shubham Kumar Nigam, Anurag Sharma, Danush Khanna, Noel Shallum, Kripabandhu Ghosh, and Arnab Bhattacharya. Legal judgment reimagined: PredEx and the rise of intelligent AI interpretation in Indian courts. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Findings of the Association for Computational Linguistics: ACL 2024*, pages 4296–4315, Bangkok, Thailand, August 2024. Association for Computational Linguistics.
23. Shubham Kumar Nigam, Balaramamahanthi Deepak Patnaik, Shivam Mishra, Noel Shallum, Kripabandhu Ghosh, and Arnab Bhattacharya. Nyayaanumana

inlegalllama: The largest indian legal judgment prediction dataset and specialized language model for enhanced decision analysis, 2024.
24. mi-young Kim, Juliano Rabelo, Randy Goebel, Masaharu Yoshioka, Yoshinobu Kano, and Ken Satoh. *COLIEE 2022 Summary: Methods for Legal Document Retrieval and Entailment*, pages 51–67. 04 2023.
25. Deepak Bastola and Woohyeok Choi. Hybrid topic-semantic labeling and graph embeddings for unsupervised legal document clustering, 2025.