Legal Precedent Assistant for Indian Family Courts.

1. Abstract:

In today's world, due to the ever-rising population and inability of the economy to cope up with it has led to a rise in crimes, law violations, struggle for resources, etc. One of the main concerns is the massive backlog in the Indian courts, pertaining to both the population and the inefficiency of the Indian Judiciary. Legal Professionals such as Judges, Lawyers, Consultants often must go through past court documents in order to come across "Precedents" (Past Rulings) which can help in passing informed judgements in Ongoing trials. This is a tedious process and can take months if not years, leading to delay in Judgements and as the proverb goes "Justice delayed is Justice denied."

NLP (Natural Language Processing) is a branch of ML (Machine Learning) that deals with making machines comprehend Human language and its context. By feeding aforementioned models with Court Datasets, we can train them to not only find Legal Precedents but can also be used to Summarize and Predict the actual verdict.

In this project, we aim to create a Legal Precedent Assistant using NLP that will aid in the process of finding Precedents and predicting verdicts based on presented evidence to reduce the burden of the Indian Judiciary and improve its efficiency. This project has 3 main goals: Mapping the IPC codes found in a Court Document, Using the Mapped IPC codes alongside the context in the document to predict an outcome(Appellant/Defendant Wins) coupled with the Judgment And finally Summarizing the entire document and Judgement passed in Simplified terms by removing Legal Jargon, thus allowing General Public to understand the Court's Proceedings. The performance of the model was found to be accurate 85\% of the time but can be improved further with larger models and more refined Datasets.

2. Introduction:

2.1 Motivation:

Due to the massive Population and issues with corruption, the Indian Judiciary has become woefully inefficient. Since, passing Judgements require one to look for Legal Precedents, the process is slowed down even further. Furthermore, due to the complicate Jargon used in Legal documents, it has often been inaccessible to the general public in India. There were over 44 million pending cases in the Indian courts in 2023**\cite{b1}**. Hence, there is need for a tool, specifically trained on Indian Legal Corpus that can help improve the efficiency of the Judiciary and also help the public understand the Nuances of the Legal landscape.

The dataset has been procured from AWS Registry for Indian Supreme court and High Court judgements in the form of court case docs(PDF format) and their associated metadata files(in .json format). These PDFs are then converted into texts, cleaned and their content is mapped to their metadata file to create an Indian Corpus dataset, that consist of the facts in the document. The facts include the IPC codes, Number of Judges, Acts, Disposal nature, Verdict label, etc.

Most of the existing NLP models for Legal Document processing such as LegalBERT, RoBERTA, DistillBERT, etc, are trained either on European, American or Chinese Legal corpus, thus making them difficult for use in the Indian Legal landscape. To combat this, we have proposed the development and design of a \textbf{Legal Precedent Assistant for Indian Judiciary}, focusing on Determining truth from presented evidence, IPC mapping, Summarization and Verdict Prediction.

DAPT (Domain-Adaptive Pre-training) will be used in conjunction with SCM(Similarity Case Matching) and Contrastive learning to train a LegalBERT/Llama model on the Indian Legal Corpus dataset. This model will then be finetuned further for our 5 stated goals. Since, the Indian Legal process is descended from the British Legal system, LegalBERT/Llama could be adapted/transferred from ECHR to IPC; provided we have a substantially large and refined dataset.

The system will take case documents (.pdf, .docx, .txt, etc) as input; convert and clean them into a text files and feed them to the 5 stage system pipeline. The 1st stage will be for Evidence extraction and using the alibis and statements provided by both parties to try and find any contradictions in the case. The 2nd stage will be for mapping the IPC codes found in the input document. In parallel to IPC mapping, we will also have the 3rd stage which is used to find legal precedents/case law for similar types of cases in the past. The 4th stage will be verdict prediction; to use the mapped IPC codes and additional context extracted from the 1st and 3rd stages, to predict an outcome **\cite{b5};** Chance of Appellant/Defendant winning and what the judgement/punishment will be in a short sentence. The 5th and final stage will be to summarize the document in 150 words or less. The 2 Initial stages are aimed towards Legal Professionals, whereas the last stage is primarily for the General public's understanding. NER for IPC mapping should not be used as Standalone, as Judges utilise more context to Adjudicate for a given circumstance **\cite{b3}**.

The system can also take in Case documents in Marathi as well as Hindi, however, due to lack of specialized and established models for Indic languages their Prediction and Summarization capabilities would be greatly limited and more raw data and metadata would be required to improve this part of the system.

2.2 Problem Statement

Despite growing interest in Legal NLP, current systems exhibit critical limitations that hinder educational adoption:

1. High Computational Costs:

Commercial Legal NLP models, often forego BERT based models and smaller specialized LLMs such as Llama-3-Legal and rely heavily on popular LLMs such as ChatGPT and Gemini, which have exponentially higher compute requirements and concerns related to privacy and data hallucination. BERT and Legal Llama have cost per million tokens ranging from \$0.20 to \$0.80, whereas popular LLMs have a cost ranging from \$2.50 to \$10.

2. Lack of Datasets:

Most BERT and Llama based Legal models are primarily trained on US and ECHR datasets, due to lack of large scale Indian legal corpus. Thus, there is a need to create datasets for Indian High Courts and District courts.

3. Lack of Evidence Scrutiny:

Existing Legal models are used primarily for summarization or verdict prediction, they do not have the ability to distinguish between any contradictions based on the provided evidence/alibis and must treat the provided input as facts, which they cannot scrutinize.

4. Lack of Precedent in Verdict Prediction:

Existing models tend to utilise the statements and acts mentioned in the case documents and map them to IPC to predict a generalized sentence. However, this does not take into account established precedents (case law) for similar cases in the past to provide a verdict more relevant in the current context.

5. Limited Language Support:

Popular Legal NLP models are based on either English or Mandarin. Indic Language support is exceedingly rare due to the lack of dataset based on Indian legal corpus.

There exists a clear need for a thorough, multi-purpose, scalable, and ipc-accurate Legal NLP model that bridges the computational overhead gap, scrutinizes presented evidence, finds relevant precedents, supports multiple languages, and provides a framework aligned with judiciary standards—all while maintaining the privacy of both parties and provide a fair verdict.

2.3 Research Objectives:

The primary objective of this project is to design and validate a Legal NLP model that provides verdict prediction, ipc mapping, simplified legal summary, etc using fine tuned legal models and data collected from Bombay HC and various District courts in the state of Maharashtra. The specific research objectives are:

1. Legal Dataset Creation:

Develop a dataset of Indian Family courts based on Taluka, District and High court cases. The AWS registry provides Supreme court case documents in abundance, but not for High courts and District courts. The dataset should contain columns such as bench, IPC codes, disposal nature, case duration, verdict, arguments, evidence sections, etc.

2. Evidence Scrutinization:

Develop a fine-tuned Llama model, that cross-references data and performs a consistency audit on the statements and evidence presented by both sides to detect any contradictions. This will allow the system to determine the truth and provide clear context to help with verdict prediction.

3. Precedent Search:

Implement Contrastive learning and Similar Case Matching(SCM) to allow the model to find older cases which are similar to the current case, and take into account the context/precedent of older cases while providing a verdict.

4. IPC Mapping:

Implement NER to extrapolate the IPC applicable to the current case, based on the different sections of evidence, arguments, accusations, etc.

5. Verdict Prediction:

Predict the verdict while also taking into account Precedents(Case law), IPC mappings and the contradictions found in the evidence.

6. Legal Document Summarization:

Summarize the case and verdict in 150 words or less, while also eliminating any legal jargon, to make the system's output more interpretable/accessible.

2.4 Scope & Limitations:

2.4.1 In-Scope:

The model development encompasses:

- Dataset building via data collected from District and High courts and filtering based on Family court cases.
- Perform case evidence scrutinization on both sides using Llama for consistency audit to find any contradictions.
- Precedent search via Contrastive learning and Similar Case Matching (SCM).
- IPC mapping using NER via Bi-LSTM encoder/Llama.
- Verdict Prediction using Llama/LegalBERT.
- Summarization using Llama.
- Indic language support limited only to Marathi and Hindi.
- Acts as replacement/tool for a hired legal assistant.

2.4.1 Out of Scope:

The following are explicitly excluded from the current project phase:

- No support for other Local languages such as Gujarati, Kannada, Tamil, etc.
- Evidence scrutinization limited only to evidence and case docs provided as input.
- Consistency audit for evidence prone to LLM hallucinations.
- Only focused for Family court, no support for Criminal, Industrial, Copyright proceedings, etc.
- Quality of dataset dependent upon raw data provided by Family/District/High courts.
- Not meant to be used as a replacement for an actual Attorney/Lawyers/Judge.
- Verdict and summary might need to be reviewed, as the sentence passed might not have all the context despite using Precedents.
- The Verdict does not take into account the current financial, personal and other circumstances of individuals involved when passing the sentence.

2.5 Research Organization:

The remainder of this document is organized as follows:

Chapter 2 (Literature Review): Presents a comprehensive synthesis of 23 research papers on Legal NLP models such as BERT, RoBERTa, Distill-BERT, BigBird, Legal-BERT, Legal-Llama, etc, using approaches such as DAPT (Domain Adaptive Pre-Training), Contrastive Learning, SCM (Similar Case Matching), UDA (Unsupervised Data Augmentation), DAM (Dual Attention Mechanism), etc. These existing systems primarily focus on summarizing, predicting verdicts, finding similar cases and named entity recognition (NER). This section includes tabular summary of key findings, gaps, and relevance, followed by systematic gap analysis identifying 15 specific research gaps across hardware, methodology, pedagogy, technical, and regulatory domains.

Chapter 3 (System Analysis and Design): Details functional and non-functional requirements derived from gap analysis and educational needs. Presents design alternatives comparison (DAPT+SCM) with justified selections. Includes system architecture, model parameter specifications, operational workflow, and design rationale explicitly mapping technical decisions to research gaps.

Chapter 4 (Proposed Methodology): Outlines six-phase development plan (Dataset building and refinement, Evidence scrutinization, IPC Mapping, Precedent search, Verdict prediction, Summarization) with deliverables, timelines, and evaluation metrics. Defines technical performance benchmarks, user experience measures, testing procedures, and risk mitigation strategies.

extbfReferences: Comprehensive bibliography of cited literature using biblatex with consistent IEEE-style formatting.

This structured presentation aims to provide both academic rigor through literature grounding and practical feasibility through detailed technical planning, positioning the project for successful implementation and potential contribution to the Indian legal landscape.

3. Literature Review(Survey)/Background/Related Work/:

Citation method on Overleaf: \cite{b1}

Replace [\b1] by \cite{b1}

The use of AI-ML in the Legal field has attracted considerable attention. Since the rise of NLP and LLMs it has become feasible to process large amounts of data and produce results in line with Verdict predictions, Summarization, Precedent finding. There is a large amount of legal data from the last century that has been scanned and digitized which has led to the rise of a Sub-field in the NLP domain known as LJP (Legal Judgement Prediction) **\cite{b1}** to make the judiciary more efficient and faster. PLMs such as LegalBERT tend to outperform with 88.3 MaF compared to less than 80.2 MaF for generic LLMs and HANs. For documents exceeding 512 tokens, the Longformer-based **Lawformer** achieved superior results in criminal cases (**95.4 MaF** for charges) by capturing long-distance dependencies. Integrated frameworks like **LADAN + MPBFN** reach **96.60%** accuracy for article prediction and **96.42%** for charge prediction. Despite the 43 datasets and 16 evaluation metrics used, there is still need for a Legal NLP model that is trained primarily on Indian Legal corpus as all existing models are trained on EU, US or Chinese corpora. Black box nature of the output leads to low interpretability which is contrary to the justification based legal landscape. **Prison term** prediction remains suboptimal (e.g., **42.55 MaR** in few-shot settings) due to data distribution challenges. Out of 36 official global languages, support is missing for 27 of them. Only uses judgement summaries and not raw evidence.

A Hybrid 2 stage model that includes RobERTa+DGCNN(stage 1) and T5 PEGASUS (stage 2) was used to create summary of Legal news**\cite{b2}**. The 1st stage is used to extract sentences and given a vector representation that contains critical information by using the Stage 1 models in conjunction with Average Pooling layer for rapid text vectorisation. DGCNN takes the dimensionality reduced vector representation of extracted sentences. The 2nd layer model(T5) takes the encoded vector, passes them through a Dense layer to produce a single embedding vector (short summary). ROUGE (Recall-Oriented Understudy for Gisting Evaluation) was used as the evaluation measure. The model was found to be effective in summarizing relatively short Legal articles, it displayed a limited ability extract deep/full context (arguments, entities) from longer/complex inputs.

A circumstance aware LJP framework known as NeurJudge was proposed to assist judicial decision making by separating crime facts into adjudicating, statutory and discretionary circumstances to better model what decisions are suitable\cite{b3}. The system introduces a novel approach called Circumstances of Crime aware Fact Separation (CCFS) to extract the facts from the input. An improved model NeurJudge+ utilises graph-based embeddings to distinguish articles/charges that are similar/intertwined. The summaries produced by the model are easy to interpret. The high computational cost coupled with the dataset's reliance only on Chinese legal corpora make it difficult to directly transfer to Indian Legal landscape. The model struggles with verdict and sentencing when 2 articles are applicable which have high descriptive similarity. Due to the Black box nature it lacks explainability/ interpretability.

A comparison between Word2vec and BERT models was performed while using UDA (Unsupervised Data Augmentation); combining labelled and un-labelled data to increase robustness \cite{b4}. The dataset is scarce and was sourced from the Brazilian Prosecutor's office's records, leading to overfitting in BERT. After implementing UDA for Data Augmentation, the Accuracy of both models jumped from 80.7\% to 92\%. The main drawback is the use of Synthetic dataset to augment the model; the resulting performance may not fully generalize to real-world legal scenarios. Small claims court have verbose case descriptions which are constrained heavily by the 512 token limit of BERT models.

BART, Random Forest and LIME(XAI) are used in conjunction with each other to help provide both Summarization, IPC Prediction and Verdict Prediction respectively for the Document. A summary of 150 tokens is generated from a document of max length 1024 tokens\cite{b5} resulting in an accuracy of ~97\%. LIME is incorporated to improve explainability and transparency. However, due to the use of both BART and LIME complexity and computational overhead increases. Using max length of 1024 tokens consumes large amounts of VRAM, whereas limiting to 512 tokens provides comparable performance.

An Ontology-driven knowledge-block summarization method for Chinese judgment document classification was proposed. Domain ontologies and top-level legal ontologies are merged to extract three core blocks: objective facts, subjective intent, and judgment results. The system\cite{b6} uses Word2Vec embeddings, JieBa tokenizer (for Mandarin only) and Word Mover's Distance (WMD) to compute similarities between extracted blocks, followed by a KNN classifier. Using specific blocks and not entire documents increases both accuracy and speed. However, it requires high quality Ontologies and is linguistically dependent on Chinese corpora only, limiting its transferability between jurisdictions such as India. WMD requires high computational overhead standard models such as Bag of Words, TF-IDF fail to capture document structure and legal semantics in depth.

In line with the UN’s Sustainable Development Goals (SDGs) an ensemble of SVM, Naïve Bayes and LSTM was used to create a system that can correctly label/classify court cases based on their type and what SDG they fall under**\cite{b7}**. Data augmentation and ensemble strategies were implemented to handle label imbalance between classes, however lack of data from the Brazilian Supreme Court, led to overfitting. All metrics such as Accuracy and F1 score peaked at a stable ~0.80. This model performs effectively when important keywords are present but cannot infer deeper contextual relationships due to the small and restrictive nature of the dataset. SDG model relies heavily on case law/precedents leading to higher complexity and low reliability on cases that do not match any precedents while also being limited to one language (Portuguese).

A text-importance similarity matching framework was proposed to improve long document legal case retrieval. A novel Unsupervised clustering and Contrastive learning approach that identified and preserved only the most critical and factual sentences was created**\cite{b8}**. These extracted facts were then fed into a BERT based encoder to find Similarity score between Legal documents. Cluster-center distance is used to quantify the impact of each extracted fact; allowing the system to surpass the 512 token limit of the BERT based models. Integrating triplet based contrastive learning and center loss to better differentiate between cases, a final accuracy of 75.08\% was achieved. The system works effectively on larger inputs but is restricted to Chinese Legal corpora and non-transferrable due to differences in Jurisdictions and Language as well as the traditional 512 token limit of BERT models, that cannot be overcome in Indian Legal corpus, unless Contrastive learning is applied. Lower quality of the initial unsupervised clusters can cascade into low accuracy, increase inference latency and event sequence disruption.

A Transformer based ECHR case classification framework was proposed to automate the detection of Human rights violations from extremely large Court judgements. A Sliding-window text sequence expansion technique is used to exceed the 512 token limit for BERT based models such as RoBERTa, Legal-BERT, BigBird, ELECTRA. RoBERTa performed the best in the Binary violation classification(F1 score of 86.7\%), whereas for multi-class classification BigBird outperformed all other models with an F1 score of 78.1\% **\cite{b9}**.Although, the Sliding window approach allows for the BERT based models to exceed their token limits and process larger documents, it incurs high computational overhead and is overdependent on English corpus with good metadata. Adding extra case features such as court branch, importance score leads to diminishing returns due to the text content dominating the feature space. DAPT on LegalBERT and BigBird remains as a point of improvement(unexplored here).

A large scale Bangla NLP Legal corpus named KUMono was created by Web scraping 1.3 million articles across 18 different categories. It contains 353 million word tokens and 1.68 million unique tokens to address the pressing need for a Bangla language corpus. This corpus was further enhanced by using TF-IDF for Article categorization. 6 NLP/ML models were utilised to classify the Court cases present; highest accuracy was achieved by Random Forest and Decision Tree Classifiers with performance metrics exceeding 0.98(Precision, Recall and F1 score) **\cite{b10}**. KUMono has a large scale, but it lacks context and depth due to dependence on web scraping. Transformer models such as BERT are superior for the purpose of summarization/comparison. The system is also limited to the Bangla corpus with minimal Arabic coverage. Dataset size is low despite Bangla being the 7th most spoken language worldwide.

An SCM (Similar Case Matching) system was developed to enhance long document parsing and similarity matching using a fine-tuned LegalBERT encoder combined with a Dual attention architecture. Local self-attention was used to extra important intra-sentence features, Global attention was used to extract broader context between multiple documents**\cite{b11}**. The dual attention mechanism allowed the system to outperform existing systems on Cosine, Manhattan and Jaccard metrics. Trained on CAIL+SCM datasets the system was found to have good recall and an accuracy of 89.5\% for Criminal cases and 90.2\% for Civil cases. The dual attention architecture and complex fine-tuning of LegalBERT(12 layers) led to significant computational overhead. The system is limited to Chinese corpora only. The network model lacks semantic depth interaction for Siamese. Usage of basic word frequency model leads to failure in capturing legal jargon and local key features in larger documents.

Legal NLP is classified into 3 main categories/tasks: Legal Search (retrieval, entailment, QA), Legal Document Review (NER, similarity, classification, summarization), and Legal Prediction—showing that domain-specific models like LEGAL-BERT, LamBERTa, BureauBERTo, ConfliBERT consistently outperform general LLMs such as ChatGPT**\cite{b13}**. Domain Adaptive Pre-Training (DAPT) on relevant Legal corpora allows smaller NLP models to outperform LLMs and achieve competitive performance for the above 3 tasks with an F1 gain of 7.2\%. DAPT on legal dataset provides an F1 gain ranging from 15.4\% to 18.2\% as compared to DAPT on generic dataset. The computational cost of specialized NLP models is lower than LLMs but are found to be inferior in long context/large input documents, generalization across jurisdictions (Indian, Chinese, EU, USA, etc) and dataset diversity/size.

A cross-domain LJP frameworks named JurisCTC was proposed to overcome data scarcity in Criminal law by transferring knowledge from Civil law datasets using Unsupervised Data Augmentation (UDA) and Contrastive learning **\cite{b14}**. A BERT encoder is combined with a class and domain classifier through a Gradient Reversal layer to optimize Maximum Mean Discrepancy (MMD). The system achieves a substantial accuracy of 76.59\% on Criminal law alongside a 78.83\% on Civil law by learning domain invariant representations, The system has very strong generalisation due to UDA, but it is limited only to Chinese Legal corpora and incurs high computational cost for adversarial BERT training. JurisCTC has a higher rate of false positives in the context of criminal cases and trails behind GPT4.0 (75.92\% vs. 83.00\%) for the same. It also requires substantial manual intervention for feature engineering.

Keynote highlights the rapid progress and evolution in NLP, explaining the range of subtasks from basic pre-processing to NER, text similarity, QA, summarization, sliding sequence window, etc. A notable example is the Multi-lingual Legal NLP model developed for Swiss Federal court that can handle 20 different Languages**\cite{b15}**. Domain pre-trained NLP models such as BERT can provide performance equivalent or surpassing general LLMs with exponentially higher compute power, provided the 5 components: architecture, hyperparameters, training data, model weights/checkpoints, and source code are kept fully open Source. Private firms have an advantage when it comes to powerful models that can handle multiple legal case types, the model in question here required an investment of \$30 million which is infeasible for individuals or smaller teams.

An LJP system named KEMCAN was proposed, utilising a multi-cross attention architecture. The system incorporates legal charge knowledge (definitions, subjective/objective elements, etc) with the fact description to better differentiate between the similar charges/penal codes mentioned in the text**\cite{b16}**. The system encodes both fact sentences and knowledge units using Bi-GRU + Attention mechanism, mapping each sentence with relevant legal information. The system was able to outperform models such as NeurJudge**\cite{b3}**, LADAN, BERT-Crime, etc, with a F1 score difference between +3.22\% to +6.5\% over these models. KEMCAN is effective at understanding context but requires a manually refined dataset while also being limited to Chinese criminal legal corpus. KEMCAN was focused primarily on applicable articles and charges, while ignoring the critical subtasks such as prison sentence.

LASG is a legal document summarization framework that was proposed to streamline judicial document analysis by incorporating CKIP transformers (Chinese BERT for sentence embedding) with a PageRank based re-ranking algorithm to extract most representative/important sentences from the documents **\cite{b17}**. Semantic similarity of extracted facts is computed via cosine similarity after which PageRank is applied to select the top/k-most important sentences to be part of the summary. LASG outperforms BERTSUM and vanilla CKIP transformers and achieves high performance metrics on ROGUE2 (12.72), ROUGE-L (18.33), etc. The system is efficient, lightweight and easy to deploy but is dependent on CKIP transformer. The summaries generated might not encompass the full depth of the corpus. The model lacks domain-specific customization/fine-tuning for different legal case types leading to lack of contextual depth due to being trained primarily on criminal cases. Hallucinations are also a major setback due to the abstractive LLMs.

A legal text classifier was developed to classify petitions for Brazil's Public Prosecutor's Office. This study compared TF-IDF, Word2Vec, SVM, Logistic Regression, Decision trees, CNNs, RNNs, and it was found that Word2Vec combined with LSTM encoder achieved the highest performance at 90.47\% Accuracy and F1 score of 85.49\%**\cite{b18**}, across 18 legal classes and 922,000 cases. This approach offered better semantic generalization relative to traditional bag-of-words models. However, TF-IDF was found to be more effective for simpler classifier models and smaller, more domain-specific datasets, albeit with limited context capacity than BERT based models. Random Under-Sampling (RUS), Over-Sampling were restricted due to computational constraints.

A systematic study of all NLP/LLM systems found that domain-specific transformer-based NLP models such as BERT can outperform general purpose LLMs while also requiring substantially lower computation resources**\cite{b19}**. By using DAPT**\cite{b13**}, Contrastive learning**\cite{b8} \cite{b14}**, Dual attention architectures**\cite{b11}**. the F1 score of specialised NLP models such as Legal-BERT can be increased by 8-15\% over a generic LLM. Using techniques such as Sliding sequence window, the 512 token limitation**\cite{b9}** of traditional BERT models can also be overcome. The models also require high cost expert annotated judgements for reference.

LegalRAG is a RAG (Retrieval Augmented Generation) based framework designed for low-resource legal documents for the Bangla corpus. It compares and utilizes Llama3.2(3B) and Llama3.1(8B) wherein the cosine similarity increases from 0.76 to 0.82 when transferring from the former to the latter**\cite{b20}**. The dataset is augmented by using RAG to add relevant data scraped from external web sources. Due to the scarcity of the Bangla corpus, synthetic data was used to augment the overall dataset. The system has high accuracy for Bangla/English corpus but is constrained due to the low dataset size and unavailability of relevant data to scrape. The system exhibits overfitting lack of DAPT and overall low resource nature of the dataset. The usage of synthetic data leads to poor out of context vulnerability, closed loop bias and computational latency trade-offs.

Pre-trained Language Models (PLMs) across 8 legal datasets were evaluated and it was observed that they outperformed non-PLM models by 4\%-35\% on most NLP tasks. Domain specific models such as LegalBERT were the only models that could surpass the performance of PLMs**\cite{b21}**, achieving marginal performance gains of 2\%-5\%. PLMs demonstrated strengths in handling legal terminology, complex reasoning and better recall for multi-label tasks. However, they underperformed by 5\% or more in regards to cross-domain transferability and limited to a 512 token length. Domain specific PLMs also exhibited limited transferability between different legal sub-domains. Diminishing returns in F1 score were exhibited when processing larger legal documents; 1.5\% gain in F1 score required 7 times longer training. PLM retrieval suffered from low accuracy due to difficulty in handling shared keywords that are legally irrelevant which lead to a gap in legal semantic matching.

HANOI-Legal is a parallel learning framework that adapts Pre-trained Language Models (PLMs) using Uniprompt; a unified QA style prompting scheme that reformulates diverse datasets into a single text-to-text format. Built on an encoder-decoder PLM(Randeng-T5-784M) **\cite{b22}**, the system performs unified prompt-based fine tuning yielding strong gains; +22.13\% F1 on CivilEE-CLS dataset and +46.35\% and +41.46\% on CivilEE-Args and CJRC datasets, respectively. However, the performance of the system is constrained by the relatively small size of the T5 model. HANOI performs best in data-scarce environments only, in resource rich environments other models surpass it. HANOI framework's scalability for larger models (100B+ parameters) is unpredictable.

An NLP model was created to predict the outcomes of Philippines SC corpora. The system incorporated bag of words n-grams with spectral clustering-based classifiers alongside popular classifiers such as SVM and Random Forest. The dataset was small and included approximately 6,500 cleaned and metadata tagged SC cases. SVM with n-grams had an accuracy of 45\%; improved to 55\% with topic-cluster features**\cite{b23}**. The best performance was provided by Random Forest classifier with topic-cluster features at 59\%. The models were simple and computationally light, but due to the small dataset and lack of standardized legal document format significantly hindered the performance of the models. Bag of words model is insufficient for extracting abstract legal reasoning due to courts focusing on “questions of law” rather than “questions of facts”.

An NLP summarization model was created for the Turkish Constitutional Court decisions that utilised an expertly annotated 1300 case dataset fed to a BERT2BERT model to produce summaries and verdict prediction was performed using XGBoost. The extractive-abstractive nature of the models enabled it to circumvent the 512 token limit of BERT. The XGBoost model was able to attain a 93.84\%**\cite{b24}** accuracy when fed full texts and 62.30\% accuracy when fed BERT generated summaries. The main advantage of this Hybrid approach was high accuracy of prediction and summarization with relatively low computational cost in part due to the smaller dataset. However, due to the small dataset and its need to be annotated by experts, scalability is challenging due to the computational overhead of BERT2BERT. The model is limited to Constitutional court, and does not generalize well for Criminal and Administrative laws. The models lack transparency and there is a need to introduce XAI to improve interpretability.

Transformer based models (BERT) were compared with LLMs (Llama) to evaluate their effectiveness for summarizing Portuguese legal documents. A highly annotated and expertly curated dataset of 2,373 documents was used as the evaluation base. LegalBERT and BERT-TRJ were compared with Llama3.1(70B) and Gemma2(27B) for the NER task. The fine-tuned BERT models had the highest F1 score lying between 0.74-0.96**\cite{b25}**; outperforming LLMs due to their higher token-level precision. Llama3.1 was tested in a zero-shot method and achieved a peak F1 score of 0.93. Due to the imbalance in dataset and small size; generalization was limited. The LLMs could not handle complex, multi-span legal entities. Confidentiality constraints surrounding source documents and expert annotations led to issues with reproducibility. Gemma2 extracted excessive amounts of irrelevant information and also suffered from hallucinations.

Tabular Format:

Title (Year)	Authors	Key Findings	Gaps / Limitations	Relevance / Context
A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges \cite{b1} (2023)	J. Cui, X. Shen, and S. Wen	Domain-specific PLMs and Longformer models (e.g., Lawformer) significantly outperform generic LLMs in LJP tasks, achieving up to 96.6\% accuracy in charge prediction by capturing long-distance dependencies.	Existing models lack training on Indian Legal corpora, suffer from "black box" interpretability issues, and show suboptimal performance in prison term prediction and multilingual support.	The digitization of century-old legal data has enabled Legal Judgement Prediction (LJP) to enhance judicial efficiency through automated summarization and verdict prediction.
A Legal News Summarisation Model Based on RoBERTa, T5 and Dilated Gated CNN \cite{b2} (2023)	W. Qin and X. Luo	A hybrid architecture using RoBERTa+DGCNN for extraction and T5-PEGASUS for abstraction effectively summarizes legal news using ROUGE as a primary metric.	Limited ability to extract deep context or complex arguments from longer inputs.	Employs a two-stage approach—vectorization and dense-layer embedding—to streamline the generation of concise summaries for legal news articles.
A Circumstance-Aware Neural Framework for Explainable Legal Judgment Prediction \cite{b3} (2024)	L. Yue, Q. Liu, B. Jin, H. Wu, and Y. An	NeurJudge utilizes Circumstances of Crime aware Fact Separation (CCFS) and graph-based embeddings (NeurJudge+) to accurately model decisions and distinguish between intertwined charges	The system incurs high computational costs, lacks interpretability due to its "black box" nature, and struggles when distinguishing between highly similar articles.	A circumstance-aware LJP framework designed to assist judicial decision-making by categorizing facts into adjudicating, statutory, and discretionary circumstances

Title (Year)	**Authors**	**Key Findings**	**Gaps / Limitations**	**Relevance / Context**
A Small Claims Court for the NLP: Judging Legal Text Classification Strategies With Small Datasets \cite{b4} (2023)	M. Noguti, E. Vellasques, and L. S. Oliveira	Implementing Unsupervised Data Augmentation (UDA) increased accuracy from 80.7\% to 92\% for small datasets.	Use of synthetic data may prevent generalization to real-world legal scenarios; constrained by BERT's 512-token limit.	Examines strategies for legal text classification when data is scarce.
AI-Driven Prediction of Indian Criminal Case Outcomes \cite{b5} (2024)	L. Boppana, H. Ranga, P. S. A. Pravallika, T. Thakre, and Y. Lakshmi	An ensemble approach utilizing BART and Random Forest achieves ~97\% accuracy in IPC and verdict prediction, generating 150-token summaries from 1024-token inputs.	High VRAM consumption and computational overhead occur due to BART and LIME integration; performance at 1024 tokens is comparable to 512 tokens.	Incorporates Explainable AI (LIME) to improve transparency and explainability in the multi-task prediction of Indian criminal case outcomes.
An Ontology Driven Knowledge Block Summarization Approach for Chinese Judgment Document Classification \cite{b6} (2018)	Y. Ma, P. Zhang, and J. Ma	Extracting specific "knowledge blocks" (facts, intent, results) via ontologies and Word Mover’s Distance (WMD) increases classification accuracy and processing speed.	High linguistic dependence on Chinese corpora, high computational overhead for WMD, and the requirement for high-quality ontologies limit transferability to jurisdictions like India.	Proposes an ontology-driven approach to capture document structure and legal semantics more deeply than traditional Bag of Words or TF-IDF models.
Automated Labelling of Judicial Controversies Before the Brazilian Supreme Court According to the	R. L. Canalli, A. F. de Menezes, E. R. de	An ensemble of SVM, Naïve Bayes, and LSTM achieved a stable ~0.80 F1 score by using data augmentation to	Overfitting due to limited data, inability to infer deep contextual relationships beyond keywords,	Aligns judicial case classification with UN Sustainable Development Goals (SDGs) but is currently limited to

Title (Year)	**Authors**	**Key Findings**	**Gaps / Limitations**	**Relevance / Context**
Sustainable Development Goals \cite{b7} (2023)	Alencar, L. J. G. Freitas, P. H. V. Moreira, and R. P. Picanço	manage label imbalance.	and heavy reliance on precedents restrict its reliability for unique cases.	the Portuguese language and Brazilian case law.
Chinese Legal Case Similarity Matching Based on Text Importance Extraction \cite{b8} (2025)	A. Fan, S. Wang, and Y. Wang	Uses unsupervised clustering and contrastive learning to identify critical sentences and surpass the 512-token limit.	Lower quality of initial clusters can cause low accuracy and increased inference latency. Currently restricted to Chinese corpora.	Proposes a text-importance similarity matching framework for long documents.
Classifying European Court of Human Rights Cases Using Transformer-Based Techniques \cite{b9} (2023)	A. S. Imran, H. Hodnefjeld, Z. Kastrati, N. Fatima, S. M. Daudpota, and M. A. Wani	Sliding-window technique allows BERT models to process large documents; RoBERTa achieved 86.7\% F1 in binary classification while BigBird excelled in multi-class tasks (78.1\% F1).	Sliding-window approach incurs high computational overhead; overdependent on English corpus metadata; adding non-text features (e.g., importance scores) yields diminishing returns.	Automates detection of human rights violations in extremely large court judgments by extending BERT models beyond the 512-token limit.
Compilation, Analysis and Application of a Comprehensive Bangla Corpus KUMono \cite{b10} (2022)	A. Akther, M. S. Islam, H. Sultana, A. K. Z. R.	Random Forest and Decision Tree classifiers achieved >0.98 F1 scores in classifying court cases within a large-scale corpus of 1.3 million scraped articles.	The corpus lacks contextual depth due to web-scraping reliance, remains limited to Bangla with minimal Arabic coverage, and is	Addresses the scarcity of Bangla legal resources by providing 353 million tokens across 18 categories to facilitate legal article

Title (Year)	**Authors**	**Key Findings**	**Gaps / Limitations**	**Relevance / Context**
	Rahman, S. Saha, K. M. Alam, and R. Debnath		considered small relative to the language's global rank.	categorization and classification.
Deep Text Understanding Model for Similar Case Matching \cite{b11} (2024)	J. Xiong and Y. Qiu	A Dual Attention architecture (Local and Global) combined with LegalBERT achieved ~90\% accuracy in civil and criminal cases, outperforming standard metrics.	The 12-layer fine-tuning incurs significant computational overhead, lacks semantic depth interaction, and is restricted to Chinese corpora.	Utilizes a hierarchical attention mechanism to extract both intra-sentence features and broader inter-document context for long legal document parsing.
Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches \cite{b13} (2025)	M. Siino, M. Falco, D. Croce, and R. Paolo	Domain-specific models (LegalBERT, etc.) using DAPT consistently outperform general LLMs like ChatGPT with an F1 gain of 18.2\%.	Specialized models are inferior in long context handling and generalization across diverse jurisdictions.	Categorizes Legal NLP into Search, Document Review, and Prediction, highlighting the efficiency of smaller, domain-adapted models over high-compute LLMs.
JurisCTC: Enhancing LJP via Cross-Domain Transfer and Contrastive Learning \cite{b14} (2025)	Z. Kang, H. Cai, X. Ji, J. Li, and N. Gu	Transfers knowledge from Civil to Criminal law using UDA and domain invariant representations, achieving 76.59\% accuracy on Criminal cases.	Higher rate of false positives in criminal cases; trails behind GPT-4 in specific tasks.	Proposes a cross-domain LJP framework to overcome data scarcity in Criminal law by transferring knowledge from Civil law datasets.

Title (Year)	**Authors**	**Key Findings**	**Gaps / Limitations**	**Relevance / Context**
Keynote - AI for the Public Sector and the Case of Legal NLP \cite{b15} (2023)	M. Stürmer	Open-source domain-specific models (like BERT) can match or surpass the performance of high-compute general LLMs when all training components are transparent.	High-end models from private firms require massive investment (e.g., \$30 million), making them infeasible for small teams.	Highlights the evolution of NLP and the importance of open-source in the public legal sector.
Knowledge-Enriched Multi-Cross Attention Network for Legal Judgment Prediction \cite{b16} (2023)	C. He, T. Tan, X. Zhang, and S. Xue	KEMCAN utilizes Bi-GRU and multi-cross attention to map facts to legal knowledge, outperforming models like NeurJudge by 3.22\% to 6.5\% in F1 score.	Requires a manually refined dataset and is limited to Chinese criminal law; ignores prison sentence subtasks.	Incorporates legal charge knowledge (definitions and elements) directly into the fact description to differentiate between similar or confusing penal codes.
LASG: Streamlining Legal Adjudication with AI-Enabled Summary Generation \cite{b17} (2024)	Y. Liu and Y. Lin	LASG outperforms BERTSUM by using CKIP transformers and PageRank re-ranking, achieving ROUGE-L scores of 18.33 through cosine similarity filtering.	Summaries may lack full depth; the model lacks domain-specific fine-tuning and suffers from abstractive hallucinations.	A lightweight, efficient summarization framework designed to streamline judicial analysis by extracting the most representative sentences from legal documents.
Legal Document Classification: An Application to Law Area Prediction of Petitions to Public Prosecution Service \cite{b18} (2020)	M. Y. Noguti, E. Vellasques, and L. S. Oliveira	Word2Vec with an LSTM encoder achieved 90.47\% and 85.49\% F1 score accuracy across 18 legal classes and 922,000 cases.	TF-IDF is better for simple models/small datasets but lacks the context capacity of BERT-based models.	Compares traditional ML and DL models for classifying petitions for the Brazilian Prosecution Office.

Title (Year)	Authors	Key Findings	Gaps / Limitations	Relevance / Context
Legal Natural Language Processing From 2015 to 2022: A Systematic Mapping Study of Advances and Applications\cite{b19} (2024)	E. Quevedo, T. Cerny, A. Rodriguez, P. Rivas, J. Yero, K. Sooksatra, A. Zhakubayev, and D. Taibi	Specialized NLP models (Legal-BERT) using DAPT and contrastive learning increase F1 by 8-15\% over generic LLMs.	These models still require high-cost expert-annotated judgments for training and reference.	Domain-specific BERT models outperform general LLMs with lower resources and can overcome token limits via sliding-window techniques.
LegalRAG: A Hybrid RAG System for Multilingual Legal Information Retrieval \cite{b20} (2025)	M. R. Kabir, R. M. Sultan, F. Rahman, M. R. Amin, S. Momen, N. Mohammed, and S. Rahman	Utilizing Llama 3.1 (8B) increased cosine similarity to 0.82, outperforming the 3B model in a RAG-based pipeline for low-resource languages.	Synthetic data usage leads to closed-loop bias, overfitting, poor out-of-context vulnerability and computational latency.	Addresses data scarcity in the Bangla legal domain through Retrieval Augmented Generation (RAG).
On the Effectiveness of Pre-Trained Language Models for Legal NLP: An Empirical Study \cite{b21} (2022)	D. Song, S. Gao, B. He, and F. Schilder	PLMs outperform non-PLM models by 4\%-35\% across most tasks; LegalBERT achieves additional marginal gains in legal terminology and reasoning.	PLMs suffer from poor cross-domain transferability, a 512-token limit, and a significant diminishing returns ratio where a 1.5\% F1 gain requires 7x longer training.	Evaluates PLM effectiveness across 8 legal datasets, noting strengths in complex reasoning and recall.
Parallel Learning for Legal Intelligence: A	Z. Song, M. Huang,	Utilizing the Randeng-T5-784M model with unified prompting	Constrained by small model size (T5); scalability	Proposes a parallel learning framework (HANOI) that uses

Title (Year)	Authors	Key Findings	Gaps / Limitations	Relevance / Context
HANOI Approach \cite{b22} (2024)	Q. Miao, and F. Wang	yielded massive F1 gains, specifically +46.35\% on CivilEE-Args and +41.46\% on CJRC datasets.	and performance on 100B+ parameter models is unpredictable.	"UniPrompt" to reformulate diverse legal datasets into a single text-to-text format.
Predicting Decisions of the Philippine Supreme Court using Natural Language Processing and Machine Learning \cite{b23} (2018)	M. B. L. Virtucio, J. A. Aborot, J. K. C. Abonita, R. S. Aviñante, R. J. B. Copino, M. P. Neverida, V. O. Osiana, E. C. Peramo, J. G. Syjuco, and G. B. A. Tan	Random Forest with topic-cluster features achieved the highest accuracy at 59\%, outperforming SVM with n-grams (45\%) on a dataset of 6,500 cases.	Bag-of-words models are insufficient for abstract legal reasoning, and performance is hindered by a lack of standardized document formats and small dataset size.	Explores outcome prediction in the Philippines SC corpora with limited data and use of computationally light classifiers.
Summarization, Prediction, and Analysis of Turkish CC Decisions with XAO and a Hybrid NLP method \cite{b24} (2025)	T. Turan and E. U. Küçüksılle	Hybrid BERT2BERT and XGBoost approach attained 93.84\% accuracy for verdict prediction and used an extractive-abstractive method to bypass the 512-token limits.	Scalability is challenging due to the need for expert annotation; limited to Constitutional court cases, lacks interpratibility.	Combines extractive-abstractive summarization with Explainable AI (XAI) for Turkish Constitutional Court decisions.
Using Language Models for	S. Vasquez	Fine-tuned BERT models (F1 0.74-0.96)	LLMs suffered from	Evaluates the effectiveness of

Title (Year)	**Authors**	**Key Findings**	**Gaps / Limitations**	**Relevance / Context**
Extracting Legal Decisions from Portuguese Consumer Law Texts \cite{b25} (2025)	, T. Carvalho, D. P. Ardila, J. Verçosa, E. Ramos, A. Celecia, I. Frajhof, A. Mangeth, M. J. Lima, K. Figueiredo, M. Nigri, and M. Vellasco	outperform LLMs like Llama3.1 and Gemma2 in NER tasks.	hallucinations and extracted excessive irrelevant information.	BERT-based models versus zero-shot LLMs (Llama 3.1, Gemma 2) using a curated dataset of 2,373 Portuguese legal documents.

3.1 Research Gap Analysis:

Based on the comprehensive literature review, several critical research gaps have been identified in the current state of VR-based driving simulation and training systems:

3.1.1 Hardware and Physical Interaction Fidelity

- **High VRAM and Computational Overhead**: Several high-performing models, such as those using 1024-token **\cite{b5}** inputs or dual attention architectures, consume excessive VRAM and require significant computational power.
- **Inference Latency**: Complex models and specific similarity measures like Word Mover's Distance (WMD) **\cite{b6}** suffer from high inference latency **\cite{b8}**, hindering real-time application.
- **Adversarial Training Costs**: Advanced frameworks like JurisCTC **\cite{b14}** incur high computational costs specifically for adversarial BERT training.
- **Scalability Infrastructure**: There is a lack of predictable infrastructure for scaling specialized legal models to larger (100B+ parameter) **\cite{b22}** architectures.

3.1.2 Methodological and Sample Limitations

- **Jurisdictional and Geographic Bias**: A critical gap exists for models trained on Indian Legal corpora, as the majority of current research focuses on US, EU, or Chinese datasets **\cite{b6}** **\cite{b8}** **\cite{b11}** **\cite{b16}**.
- **Data Scarcity and Overfitting**: Many systems suffer from overfitting **\cite{b4}** **\cite{b7}** **\cite{b20}** due to small, restrictive datasets, such as those sourced from specific prosecutor offices or Supreme Courts with limited case counts.
- **Synthetic Data Generalization**: The heavy reliance on synthetic data to augment small datasets **\cite{b4}** **\cite{b20}** leads to concerns that model performance will not generalize to real-world legal scenarios.
- **Imbalanced Data Distribution**: Challenges in data distribution, particularly label imbalance **\cite{b7}** **\cite{b25}**, lead to suboptimal results in niche tasks like prison term prediction.

3.1.3 Pedagogical and Training Effectiveness

- **Lack of Domain-Specific Customization**: Many models lack fine-tuning for specific legal sub-domains **\cite{b17} \cite{b21}** (e.g., transitioning from Criminal to Administrative law), leading to a lack of contextual depth.
- **Diminishing Returns in Training**: There is a significant efficiency gap where marginal gains in F1 score (e.g., 1.5\%) require exponentially longer training times **\cite{b9} \cite{b21}** (e.g., 7x).
- **Annotation Dependency**: The effectiveness of these models is highly dependent on expert-annotated judgments **\cite{b19} \cite{b24} \cite{b25}**, which are expensive and difficult to scale.
- **Failure in Abstract Reasoning**: Traditional training methods like "Bag of Words" fail to capture abstract legal reasoning **\cite{b6} \cite{b18} \cite{b23}**, focusing too much on "questions of fact" rather than "questions of law".

3.1.4 Technical and Realism Constraints

- **The 512-Token Limitation**: Standard BERT-based models are constrained by a 512-token limit **\cite{b4} \cite{b8} \cite{b9} \cite{b21} \cite{b24}**, which is insufficient for verbose legal documents and complex case descriptions.
- **Hallucinations and Reliability**: Abstractive LLMs used for summarization suffer from hallucinations**\cite{b17} \cite{b25}**, which is a major setback in a high-stakes legal environment.
- **Semantic Depth and Jargon**: Basic models often fail to capture complex legal jargon **\cite{b10} \cite{b11} \cite{b17}** required for deep semantic matching.
- **Black Box Nature**: A lack of transparency and explainability **\cite{b1} \cite{b3}** in many models prevents them from being used in a justification-based legal landscape.

3.1.5 Regulatory, Ethical, and Commercialization Gaps

- **Financial Entry Barriers**: The extreme cost of developing powerful multi-case legal models (e.g., \$30 million) **\cite{b15}** creates a commercialization gap for smaller teams and public sector entities.

- **Confidentiality and Reproducibility**: Constraints regarding source document confidentiality and protected expert annotations **\cite{b19} \cite{b24} \cite{b25}** often lead to significant issues with research reproducibility.

- **Multilingual Support Gaps**: There is a massive regulatory and accessibility gap, with support missing for 27 out of 36 official global languages **\cite{b1}**.

- **Closed-Loop Bias**: The use of RAG and synthetic augmentation can lead to closed-loop biases **\cite{b20}**, where models reinforce their own errors rather than learning from diverse, objective evidence.

3.1.6 Summary of Research Gaps

The identified gaps in legal AI research reveal five overarching themes that require focused investigation:

1. **Computational efficiency and scalability**, as many state-of-the-art legal models demand excessive VRAM, incur high inference latency, and rely on costly training pipelines.
2. **Jurisdictional and dataset limitations**, including heavy bias toward non-Indian legal systems, data scarcity, synthetic-data overreliance, and severe label imbalance.
3. **Limited legal reasoning and pedagogical depth**, marked by poor abstraction, overdependence on expert annotations, and weak transfer across legal sub-domains.
4. **Technical realism and trustworthiness deficits**, such as token-length constraints, hallucinations in abstractive models, inadequate handling of legal jargon, and black-box behavior.
5. **Regulatory, ethical, and accessibility barriers**, including high development costs, confidentiality-driven reproducibility issues, and insufficient multilingual support.

The proposed **Legal Precedent Assistant for Indian Family Courts**, built using **LegalBERT** for structured semantic representation and **LLaMA-3.1** for long-context reasoning and explanation, directly addresses **Gaps 2, 3, and 4**, and partially mitigates **Gaps 1 and 5**. By narrowing the domain to family-law jurisprudence, leveraging weak supervision and domain-adaptive pretraining on Indian judgments, and employing explainable, modular inference pipelines, the system enables cost-effective, jurisdiction-aware, and interpretable legal decision support. Subsequent chapters detail the architecture, training strategy, and evaluation framework designed to bridge these critical research gaps within realistic academic and infrastructural constraints.

3.2 Requirements Analysis:

3.2.1 Functional Requirements:

Based on the identified research gaps and legal domain needs, the system must fulfill the following functional requirements:

FR1: Legal Document Input:

- The legal documents/summaries from both sides should be given as input to the system alongside any pre-existing court judgements/summaries.
- The files can be uploaded as .docx/.pdf/.txt formats.

FR2: Text Cleaning and Preprocessing:

- The system pipeline will then use regex/BERT models to remove any unnecessary symbols and patterns to convert the input documents into text files for further processing.

FR3: Evidence Scrutinization

- Llama 3.1 should be used to perform evidence scrutinization and a consistency audit for both sides, to try and determine contradictions in both side's statements.

FR4: IPC Mapping

- If the document contains any articles/code/IPC violations they should be tracked either based on the codes or based on statements.

FR5: Precedent Retrieval

- Match the current court case documents with past similar court cases and find precedents from past judgements to aid in verdict/judgement of present case.

FR6: Verdict Prediction

- Predict the verdict using the previous 3 requirements.

FR6: Summarization

- Summarize the judgement and the case contents in 150 words or less.

3.2.2 Non-Functional Requirements:

Based on the identified research gaps and legal domain needs, the system must fulfill the following functional requirements:

NFR1: Cost effectiveness:

- Make use of freely available datasets and LLM models.

NFR2: Usability and Accessibility:

- The system should be easy to use and understand not just for Legal professionals, but also laymen.

NFR3: Reliability and Maintainability

- The system should be able to handle any cases in the Family courts domain reliably, and should be maintained and upgraded as per changes in Family law.

NFR5: Explainability and Interpretability

- The verdicts/judgements/summaries should be explained and sources for precedents should be cited.

NFR5: Scalability and Extensibility

- The model should be expandable to other domains such as IP and Copyright laws, Criminal laws, Civil suits, etc.

3.3 Constraints and Design Trade-offs:

3.3.1 Hardware Constraints:

Based on the identified research gaps and driving school needs, the system must fulfill the following functional requirements:

C1: GPU requirements:

- The current device uses a Notebook version of GTX 1650 gpu with 4Gb of VRAM.
- This is sufficient to train smaller BERT/LLM models with small batch size but takes longer times and has low efficiency.
- Solution: More advanced GPUs such as RTX series, or Ampere architecture based GPUs might be required based on the dataset procured.

C2: VRAM/Memory Requirements:

- The current device; GTX 1650 GPU has 4GB of VRAM, which is sufficient to train small to medium size BERT models provided Batching is used.
- Solution: However, for LLMs like Llama 3 and above, we might require over 12GB of VRAM which is consequently tied to the GPU.
- Solution: If powerful GPUs are not available, we can use CPU RAM as swap space/shared memory given that the project is being run on a Linux kernel.

3.3.2 Software and Performance Constraints

C3: 512 token limit on BERT:

- LegalBERT and its contemporaries have a strict 512 token limit.
- Solution: This limit can be circumvented by using Sliding sequence window technique.

C4: Overfitting:

- Overfitting happens often due to small size of datasets and low variety in data.
- Overfitting can be overcome by using DAPT, Contrastive Learning, Regularization and Early stopping.

C5: Synthetic Data:

- Synthetic data can help augment the size and balancer of dataset but can lead to outputs that are not representative of real life scenarios.

- Solution: Use minimal amounts of synthetic data and only for classes that have no instances at all.

C6: Closed loop bias:

- Closed loop bias happens in RAGs and models using Synthetic data as the model reuses its own output for training.
- Solution: Can be overcome by building/acquiring a larger dataset and preventing/minimizing use of model output for training

3.4 System Objectives:

Based on the requirements analysis and design trade-offs, the system objectives are:

- To design a modular, reliable and explainable Legal Precedent Assistant for Indian Family courts.
- To create a Pipeline that takes court documents from both the opposing sides as inputs, cleans and pre-processes them and passes them to later stages for creating outputs in line with the functional and non-functional requirements.
- To integrate LegalBERT for IPC mapping of any codes/articles/statements found in the input documents as well as retrieving legal precedents from past cases.
- To integrate Llama 3 for evidence scrutinization, verdict prediction and summarization of given court documents.
- To create a system that can be transferred between different legal domains and not be constrained to just Family courts.

3.5 System Components and Architecture

Parameter	**Proposed System**	**Juris-CTC**	**KUMONO**
Cost			
Model			

3.6 System Components and Architecture

3.6.1 Hardware Subsystem

- CPU

- GPU
- Storage

3.6.2 Software Subsystem

- Python 3.10
- PyTorch CUDA
-
-

3.6 System Architecture Diagram

4. Proposed Methodology:

Idk what to write, only made dataset.

Probably divide into ~4 parts

1. Dataset creation
    a. refinement using NER or TF-IDF or something
    b. IPC mapping
2. Verdict prediction based on IPCs mapped
3. Summarization/Precedent search/finding

1. Dataset Creation:

Text Extraction:

- The dataset was procured from the AWS Indian Supreme Court registry containing metadata and PDFs of the case documents in English as well as local languages of a particular state's jurisdiction. The metadata files contained basic information such as raw html content, citation year, scraping location, etc. The data downloaded ranged from the year 1990 to 2025, older documents were avoided in order to keep the context cleaner and more modern, as going back further would risk the introduction of antiquated language and IPC and laws that are not longer relevant in modern judicial context.

- All the English PDF files were collected and transformed into text files by using regex for preserving patterns. Fitz python library was used for extracting text content from the PDFs. Fitz is a python interface for PyMuPDF that is used to parse/convert documents. Fitz was used in conjunction with ThreadPoolExecutor to create multiple threads for converting multiple documents at once.
- Regex patterns for anchors, citations, page headers were used to preserve certain keywords such as "Arguments", "IPC", "Issues" and many more. While, certain patterns were used to remove redundant page numbers and footers. These patterns also helped to extract new more refined metadata/.json files which contained data such as bench/coram details, judgement headers, word count, paragraph count, first paragraph.

IPC Mapping using NER:

- All the extracted text files and their associated metadata files are collected for parsing and refinement. Regex was used alongside 2 models for extracting important legal information such as IPC tags, arguments, issues, judge names, party names, etc, presented in the case documents. The 2 models used were dslim-BERT-Base-NER and Babelscape/wikineural-multilingual-ner.
- Dslim is a BERT based model that is fine-tuned primarily for Named Entity Recognition (NER) in the English language only. Dslim, is not trained on Indian legal corpus, but due to its NER proficiency for generic English it can be used to detect other entity names relevant to the project's goals. It was trained on the CoNLL-2003(Conference on Natural Language Learning) dataset and is proficient at extracting names of locations, judges and people in the legal context.
- Babelscape is a multilingual NER detection model proficient in English, Hindi as well as Marathi. Combining babelscape with appropriate regex allows the system to extract as many IPC sections/numbers as possible; to help with the goal of summarization, verdict prediction and precedent finding and much more.

Parquet Dataset Building:

- Using the 2 models for NER and IPC tagging; a parquet dataset was created based on the metadata and text content of Supreme court cases ranging from 1990 to 2025.
- It is made up of 26,000 records/rows and 45 columns, each row representing a unique case.
- The columns consist of IPC list, arguments, issues, verdict label, coram/judge list, text content, duration/decade, act names, facts, conclusions, disposal nature, constitutional articles, etc.

2. Domain Adaptive Pre-Training (DAPT):

Domain Adaptive Pre-training allows models to learn about new domains by training them pre-emptively on a particular given dataset. In this project, 3 primarily models were augmented using DAPT and all of them saw an increase in F1 score ranging from 7\% in the smaller models to 15\% in the larger models.

Masked Language Modelling (MLM) was used to check the improvements in the models. MLM masks/hides a certain number of words and checks whether the model is able to predict it. Loss indicates whether the prediction of the masked words was correct or not (should be as low as possible). Perplexity indicates the level of confusion in the model and whether the model truly understands the context that it is being pre-trained on/fed to.

Since the 2 primary BERT models initially used (small and medium BERTs) were trained only on generic English data such as Wikipedia and free Book Corpus, they do not have the ability to recognize legal jargon. BERT-small and BERT-medium without DAPT had an F1 score of ~55\% and 59\% respectively, which eventually rose to 64\% and 69\% respectively after augmentation with DAPT. Legal BERT-uncased was the biggest improvement, without DAPT the F1 score was observed to be 64\%, with DAPT augmentation it increased significantly to ~76\%.

The max-length for DAPT for all the models was set to 128 tokens due to the limited 4GB VRAM of the GTX 1650 GPU. Head-tail tokenization was used, since in Indian Legal corpora the 1st half of a document contains basic context and entity names, whilst the 2nd half of a document contains the verdict. This tokenization was done to extract as much context as possible, given the limited amount of computational power available at 128 tokens. Batch size for BERT-small was set to 8, due to it being the smallest model and being able to fit easily in the available VRAM, whereas for the other 2 models Batch size was reduced to 4. Perplexity (confusion score) and Loss are visualized in the table below. Models with more layers and ability to understand deeper context were able to make the most out of DAPT denoted by the lower loss and Perplexity.

Legal BERT-uncased was initially trained on US and EU based legal corpus, thus it already had a strong baseline regarding legal jargon. DAPT allowed it to adapt to the lexicon used in Indian legal corpus which is proven by its low loss and perplexity scores.

During Domain-Adaptive Pre-Training (DAPT), the model rapidly learns the linguistic patterns and repetitive structures present in Indian legal judgements. Prolonged training causes the model to overfit, wherein it begins to memorize specific sentences, IPC combinations, and document-level patterns rather than learning generalizable representations. This leads to a reduction in performance despite continued decreases in training loss. To mitigate this effect, early stopping is employed after 2–3 epochs, as most domain adaptation gains occur within the initial iterations and subsequent training offers diminishing returns. Early stopping thus prevents overspecialization, preserves generalization to unseen cases, and ensures stable performance across all downstream tasks.

Model	Epochs	Perplexity	Loss	Batch Size
BERT-small	2	8.80	2.1746	8
BERT-medium	2	7.99	2.0783	4
LegalBERT-uncased	2	6.63	1.8922	4
IN-LegalBERT	2	5.12	1.6374	4

3. Fine-Tuning/Training:

All four transformer models underwent Domain-Adaptive Pre-Training (DAPT) on the Indian Legal Corpus prior to task-specific fine-tuning. DAPT enabled each architecture—BERT-small, BERT-medium, LegalBERT-uncased, and Indian-LegalBERT—to better internalize domain-specific linguistic patterns, IPC citation formats, and the structural characteristics of Indian judicial documents. Following DAPT, the models were fine-tuned for IPC extraction, verdict prediction, and summarization using task-supervised datasets.

To accommodate differences in model capacity and GPU memory consumption, each architecture was fine-tuned with a customized hyperparameter configuration. BERT-small utilized a preprocessing window of 120 words, a maximum sequence length of 128 tokens, and a batch size of 4. Training was performed for up to five epochs with the AdamW optimizer, a learning rate of 2×10^{-5}, and weight decay of 0.01.

BERT-medium and LegalBERT-uncased were trained with larger input windows of up to 2000 words and a maximum sequence length of 384 tokens. Both models used a batch size of 2 with 16-step gradient accumulation (effective batch size of 32) to manage memory constraints. Training was conducted for four epochs using AdamW with identical learning-rate and weight-decay settings. Due to class imbalance in IPC prediction and verdict labels, focal loss with parameters α=0.25 and γ=2.0 was applied to emphasize harder samples during optimization.

The Indian-LegalBERT model used the same fine-tuning hyperparameters as LegalBERT-uncased but was initialized from an extended DAPT checkpoint trained over a larger subset of Indian court records, enabling deeper adaptation to domain-specific terminology. All models employed truncation strategies for handling long judgments and incorporated validation-driven early stopping to mitigate overfitting.

Results:

All four models were fine-tuned on the supervised Indian Legal Corpus for the downstream tasks of IPC extraction, verdict prediction, and summarization. Each model was trained for three epochs using a validation-driven early stopping strategy to prevent overfitting. Table I summarizes the comparative performance across the evaluated architectures. The smaller general-domain models, BERT-small and BERT-medium, achieved validation accuracies of 67\% and 71\%, with F1 scores of 0.65 and 0.69, respectively, reflecting their limited capacity to capture legal-domain semantics despite consistent training convergence. LegalBERT-uncased, pre-trained on non-Indian legal corpora, demonstrated a notable improvement with 78\% accuracy and an F1 score of 0.76, benefiting from its prior exposure to legal terminology. The best performance was observed for the Indian-LegalBERT model, obtained through Domain-Adaptive Pre-Training (DAPT) on Indian judicial text, which achieved 85\% validation accuracy, an F1 score of 0.84, and the lowest validation loss of 0.025. These results confirm that domain adaptation substantially enhances downstream fine-tuning effectiveness, enabling the model to better capture IPC citation patterns, factual reasoning, and judgement structures characteristic of Indian court documents.

Model	Epochs	Accuracy	F1	Validation Loss
BERT-small	3	67%	0.65	0.6517
BERT-medium	3	71%	0.69	0.0351
LegalBERT-uncased	3	78%	0.76	0.0311
IN-LegalBERT	3	85%	0.84	0.025

Conclusions:

This work presents a domain-adapted transformer-based framework for IPC extraction, verdict prediction, summarization, and precedent retrieval using Indian Supreme Court judgments. The integration of Domain-Adaptive Pre-Training (DAPT) and task-specific fine-tuning significantly improved model performance, with IN-LegalBERT achieving the strongest overall results, demonstrating the value of adapting pretrained models to Indian legal language and citation patterns.

However, several limitations remain. The system primarily handles English judgments, and performance decreases for multilingual or mixed-language cases. GPU memory constraints restricted sequence lengths, limiting the model’s ability to capture long-range dependencies in extensive judicial documents. IPC detection also relies partly on regex and general-purpose NER models, which may miss complex statutory references.

Future work will focus on expanding multilingual coverage, enabling longer-context processing through efficient transformer architectures, and developing custom legal NER models specifically tailored for Indian statutes. Incorporating retrieval-augmented generation and larger encoder–decoder models may further enhance precedent retrieval and summarization quality. These improvements aim to create a more robust and generalizable Legal Precedent Assistant for widespread judicial support.

Bibtex References:


```
@ARTICLE{b1,
  author={Cui, Junyun and Shen, Xiaoyu and Wen, Shaochun},
  journal={IEEE Access},
  title={A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges},
  year={2023},
  volume={11},
  number={},
  pages={102050-102071},
```

```
  keywords={Law;Task analysis;Surveys;Benchmark testing;Natural language processing;Measurement;Biological system modeling;Benchmark testing;Neural networks;Legal judgment prediction;natural language processing;survey;benchmark datasets;neural network},
  doi={10.1109/ACCESS.2023.3317083}}


@INPROCEEDINGS{b2,
  author={Qin, Weijian and Luo, Xudong},
  booktitle={2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)},
  title={A Legal News Summarisation Model Based on RoBERTa, T5 and Dilated Gated CNN},
  year={2023},
  volume={},
  number={},
  pages={889-897},
  keywords={Training;Measurement;Law;Training data;Logic gates;Data models;Data mining;Legal news;Automatic text summarisation;Legal intelligence;Sentence granularity;Extractive model;Pretrained model},
  doi={10.1109/ICTAI59109.2023.00134}}


@ARTICLE{b3,
  author={Yue, Linan and Liu, Qi and Jin, Binbin and Wu, Han and An, Yanqing},
  journal={IEEE Transactions on Knowledge and Data Engineering},
  title={A Circumstance-Aware Neural Framework for Explainable Legal Judgment Prediction},
  year={2024},
  volume={36},
  number={11},
  pages={5453-5467},
```

keywords={Law;Task analysis;Semantics;Rendering (computer graphics);Predictive models;Mobile handsets;Feature extraction;Explainable legal document mining;fact separation;label embedding;rationalization;text classification},

doi={10.1109/TKDE.2024.3387580}}

@INPROCEEDINGS{b4,

author={Noguti, Mariana and Vellasques, Eduardo and Oliveira, Luiz S.},

booktitle={2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)},

title={A Small Claims Court for the NLP: Judging Legal Text Classification Strategies With Small Datasets},

year={2023},

volume={},

number={},

pages={1840-1845},

keywords={Support vector machines;Law;Text categorization;Data augmentation;Transformers;Data models;Task analysis},

doi={10.1109/SMC53992.2023.10394189}}

@INPROCEEDINGS{b5,

author={Boppana, Lakshmi and Ranga, Harshith and Pravallika, Potnuri Sri Anjali and Thakre, Tanushka and Lakshmi, Yellanki},

booktitle={TENCON 2024 - 2024 IEEE Region 10 Conference (TENCON)},

title={AI-Driven Prediction of Indian Criminal Case Outcomes},

year={2024},

volume={},

number={},

pages={1389-1393},

keywords={Performance evaluation;Analytical models;Accuracy;Text analysis;Law;Computational modeling;Manuals;Reliability;Artificial intelligence;IEEE Regions;Legal Assistant;AI;Machine Learning;Judgment Prediction;Case Summarization;Indian Penal Code;Punishment Prediction;Legal Efficiency;Legal Decision Making;Judicial System},

doi={10.1109/TENCON61640.2024.10902847}}

@ARTICLE{b6,

author={Ma, Yinglong and Zhang, Peng and Ma, Jiangang},

journal={IEEE Access},

title={An Ontology Driven Knowledge Block Summarization Approach for Chinese Judgment Document Classification},

year={2018},

volume={6},

number={},

pages={71327-71338},

keywords={Ontologies;Semantics;Law;Information retrieval;Computational modeling;Neural networks;Document classification;domain ontology;document similarity;WMD},

doi={10.1109/ACCESS.2018.2881682}}

@INPROCEEDINGS{b7,

author={Canalli, Rodrigo Lobo and de Menezes, Alexandre Ferreira and de Alencar, Euler Rodrigues and Freitas, Lucas José Gonçalves and Moreira, Paulo Henrique Vencio and Picanço, Regis Proença},

booktitle={2023 IEEE International Symposium on Technology and Society (ISTAS)},

title={Automated Labelling of Judicial Controversies Before the Brazilian Supreme Court According to the Sustainable Development Goals},

year={2023},

volume={},

number={},

pages={1-7},

keywords={Visualization;Machine learning algorithms;Law;Text categorization;Supervised learning;Tag clouds;Natural language processing;analysis of legal documents;artificial intelligence;machine learning;natural language processing},

doi={10.1109/ISTAS57930.2023.10305895}}

@ARTICLE{b8,

author={Fan, Aman and Wang, Shaoxi and Wang, Yanchuan},

journal={IEEE Access},

title={Chinese Legal Case Similarity Matching Based on Text Importance Extraction},

year={2025},

volume={13},

number={},

pages={118745-118758},

keywords={Law;Feature extraction;Encoding;Natural language processing;Bidirectional control;Semantics;Accuracy;Data mining;Training;Contrastive learning;Similarity case matching;deep learning;unsupervised clustering;contrastive learning},

doi={10.1109/ACCESS.2025.3585265}}

@ARTICLE{b9,

author={Imran, Ali Shariq and Hodnefjeld, Henrik and Kastrati, Zenun and Fatima, Noureen and Daudpota, Sher Muhammad and Wani, Mudasir Ahmad},

journal={IEEE Access},

title={Classifying European Court of Human Rights Cases Using Transformer-Based Techniques},

year={2023},

volume={11},

number={},

pages={55664-55676},

keywords={Law;Transformers;Bit error rate;Natural language processing;Legal aspects;Europe;Mathematical models;Legal documents classification;European court of human rights (ECHR) dataset;natural language processing;transformers;BERT;BigBird;ELECTRA;XLNet;legal-BERT},

doi={10.1109/ACCESS.2023.3279034}}

@ARTICLE{b10,

author={Akther, Aysha and Shymon Islam, Md. and Sultana, Hafsa and Rasel Rahman, A. K. Z. and Saha, Sujana and Alam, Kazi Masudul and Debnath, Rameswar},

journal={IEEE Access},

title={Compilation, Analysis and Application of a Comprehensive Bangla Corpus KUMono},

year={2022},

volume={10},

number={},

pages={79999-80014},

keywords={Linguistics;Buildings;Natural language processing;Annotations;Tagging;Statistical analysis;Standards;NLP;Bangla corpus;N-gram;Zipf's law;article categorization},

doi={10.1109/ACCESS.2022.3195236}}

@ARTICLE{b11,

author={Xiong, Jie and Qiu, Yihui},

journal={IEEE Access},

title={Deep Text Understanding Model for Similar Case Matching},

year={2024},

volume={12},

number={},

pages={109877-109885},

keywords={Feature extraction;Law;Accuracy;Artificial intelligence;Data mining;Vectors;Semantics;Similar case matching;text mining;similarity analysis;attention mechanism;feature extraction},

doi={10.1109/ACCESS.2024.3439775}}

@INPROCEEDINGS{b12,

author={Dan, JingPei and Zhang, TianYuan and Wang, YuMing},

booktitle={2023 International Joint Conference on Neural Networks (IJCNN)},

title={Double Granularity Graph Network For Chinese Legal Question Answering},

year={2023},

volume={},

number={},

pages={1-8},

keywords={Adaptation models;Law;Scalability;Benchmark testing;Question answering (information retrieval);Cognition;Graph neural networks;Graph neural network;Double-granularity;Legal Question Answering},

doi={10.1109/IJCNN54540.2023.10192015}}

@ARTICLE{b13,

author={Siino, Marco and Falco, Mariana and Croce, Daniele and Rosso, Paolo},

journal={IEEE Access},

title={Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches},

year={2025},

volume={13},

number={},

pages={18253-18276},

keywords={Law;Artificial intelligence;Transformers;Systematic literature review;Contracts;Attention mechanisms;Databases;Reliability;Question answering (information retrieval);Quality assessment;Natural language processing;law;AI for law;legal NLP;legal tech;GPT;transformers;literature review},

doi={10.1109/ACCESS.2025.3533217}}

@INPROCEEDINGS{b14,

author={Kang, Zhaolu and Cai, Hongtian and Ji, Xiangyang and Li, Jinzhe and Gu, Nanfei},

booktitle={2025 International Joint Conference on Neural Networks (IJCNN)},

title={JurisCTC: Enhancing Legal Judgment Prediction via Cross-Domain Transfer and Contrastive Learning},

year={2025},

volume={},

number={},

pages={1-8},

keywords={Adaptation models;Accuracy;Law;Large language models;Contrastive learning;Predictive models;Natural language processing;Criminal law;Logic;Knowledge transfer;Unsupervised domain Adaptation;Legal Judgment Prediction;Transfer Learning;Contrastive Learning;Large Language Models},

doi={10.1109/IJCNN64981.2025.11229207}}

@INPROCEEDINGS{b15,

author={Stürmer, Matthias},

booktitle={2023 Ninth International Conference on eDemocracy \& eGovernment (ICEDEG)},

title={Keynote - AI for the Public Sector and the Case of Legal NLP},

year={2023},

volume={},

number={},

pages={1-2},

keywords={},

doi={10.1109/ICEDEG58167.2023.10122084}}

@ARTICLE{b16,

author={He, Congqing and Tan, Tien-Ping and Zhang, Xiaobo and Xue, Sheng},

journal={IEEE Access},

title={Knowledge-Enriched Multi-Cross Attention Network for Legal Judgment Prediction},

year={2023},

volume={11},

number={},

pages={87571-87582},

keywords={Law;Task analysis;Predictive models;Knowledge engineering;Multitasking;Deep learning;Urban areas;Legal judgment prediction;legal charge knowledge;multi-cross attention;confusing charges and law articles},

doi={10.1109/ACCESS.2023.3305259}}

@INPROCEEDINGS{b17,

author={Liu, Yi-Hung and Lin, Yi-Fun},

booktitle={2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)},

title={LASG: Streamlining Legal Adjudication with AI-Enabled Summary Generation},

year={2024},

volume={},

number={},

pages={2242-2247},

keywords={Technological innovation;Text analysis;Law;Heuristic algorithms;Transformers;Solids;Software;Legal Adjudication;Summarization;BERT;CKIP-Transformers;PageRank},

doi={10.1109/COMPSAC61105.2024.00360}}

@INPROCEEDINGS{b18,

author={Noguti, Mariana Y. and Vellasques, Eduardo and Oliveira, Luiz S.},

booktitle={2020 International Joint Conference on Neural Networks (IJCNN)},

title={Legal Document Classification: An Application to Law Area Prediction of Petitions to Public Prosecution Service},

year={2020},

volume={},

number={},

pages={1-8},

keywords={Law;Task analysis;Recurrent neural networks;Feature extraction;Natural language processing;Natural Language Processing;Text Classification;Word Embeddings;Recurrent Neural Networks},

doi={10.1109/IJCNN48605.2020.9207211}}

@ARTICLE{b19,

author={Quevedo, Ernesto and Cerny, Tomas and Rodriguez, Alejandro and Rivas, Pablo and Yero, Jorge and Sooksatra, Korn and Zhakubayev, Alibek and Taibi, Davide},

journal={IEEE Access},

title={Legal Natural Language Processing From 2015 to 2022: A Comprehensive Systematic Mapping Study of Advances and Applications},

year={2024},

volume={12},

number={},

pages={145286-145317},

keywords={Law;Natural language processing;Task analysis;Systematics;Information retrieval;Surveys;Search problems;Deep learning;Systematic-mapping-study;legal-NLP;deep learning},

doi={10.1109/ACCESS.2023.3333946}}

@INPROCEEDINGS{b20,

author={Kabir, Muhammad Rafsan and Sultan, Rafeed Mohammad and Rahman, Fuad and Amin, Mohammad Ruhul and Momen, Sifat and Mohammed, Nabeel and Rahman, Shafin},

booktitle={2025 International Joint Conference on Neural Networks (IJCNN)},

title={LegalRAG: A Hybrid RAG System for Multilingual Legal Information Retrieval},

year={2025},

volume={},

number={},

pages={1-8},

keywords={Measurement;Law enforcement;Retrieval augmented generation;Pipelines;Government;Neural networks;Information retrieval;Natural language processing;Computational linguistics;Multilingual;Retrieval-Augmented Generation;Information retrieval;Computational linguistics;Natural language processing},

doi={10.1109/IJCNN64981.2025.11228374}}

@ARTICLE{b21,

author={Song, Dezhao and Gao, Sally and He, Baosheng and Schilder, Frank},

journal={IEEE Access},

title={On the Effectiveness of Pre-Trained Language Models for Legal Natural Language Processing: An Empirical Study},

year={2022},

volume={10},

number={},

pages={75835-75858},

keywords={Law;Task analysis;Feature extraction;Deep learning;Bit error rate;Support vector machines;Convolutional neural networks;Legal natural language processing;pre-trained language model;deep learning;machine learning},

doi={10.1109/ACCESS.2022.3190408}}

@ARTICLE{b22,

author={Song, Zhuoyang and Huang, Min and Miao, Qinghai and Wang, Fei-Yue},

journal={IEEE Transactions on Computational Social Systems},

title={Parallel Learning for Legal Intelligence: A HANOI Approach Based on Unified Prompting},

year={2024},

volume={11},

number={2},

pages={2765-2775},

keywords={Task analysis;Law;Artificial intelligence;Solid modeling;Linguistics;Adaptation models;Training;Natural language processing (NLP);parallel learning (PL);parallel systems;pretrained language model (PLM);prompt tuning},

doi={10.1109/TCSS.2023.3301400}}

@INPROCEEDINGS{b23,

author={Virtucio, Michael Benedict L. and Aborot, Jeffrey A. and Abonita, John Kevin C. and Aviñante, Roxanne S. and Copino, Rother Jay B. and Neverida, Michelle P. and Osiana, Vanesa O. and Peramo, Elmer C. and Syjuco, Joanna G. and Tan, Glenn Brian A.},

booktitle={2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)},

title={Predicting Decisions of the Philippine Supreme Court Using Natural Language Processing and Machine Learning},

year={2018},

volume={02},

number={},

pages={130-135},

keywords={Databases;Drugs;Natural language processing;Law;Metadata;Machine learning;NLP;machine-learning;artificial-intelligence;law;judiciary;Philippines},

doi={10.1109/COMPSAC.2018.10348}}

@ARTICLE{b24,

author={Turan, Tülay and Küçüksılle, Ecır Uğur},

journal={IEEE Access},

title={Summarization, Prediction, and Analysis of Turkish Constitutional Court Decisions With Explainable Artificial Intelligence and a Hybrid Natural Language Processing Method},

year={2025},

volume={13},

number={},

pages={59766-59779},

keywords={Law;Artificial intelligence;Natural language processing;Text summarization;Transformers;Predictive models;Mathematical models;Libraries;Encoding;Bidirectional control;Explainable artificial intelligence;hybrid document summarization;legal judgment prediction;natural language processing;Turkish constitutional court},

doi={10.1109/ACCESS.2025.3556725}}

@INPROCEEDINGS{b25,

author={Vasquez, Santiago and Carvalho, Thiago and Ardila, Diego Páez and Verçosa, João and Ramos, Eduardo and Celecia, Alimed and Frajhof, Isabella and Mangeth, Ana and Lima, Maria Julia and Figueiredo, Karla and Nigri, Meyer and Vellasco, Marley},

booktitle={2025 International Joint Conference on Neural Networks (IJCNN)},

title={Using Language Models for Extracting Legal Decisions from Portuguese Consumer Law Texts},

year={2025},

volume={},

number={},

pages={1-7},

keywords={Deep learning;Analytical models;Accuracy;Law;Large language models;Neural networks;Named entity recognition;Transformers;Distance measurement;Prompt engineering;Deep Learning;Transformer Networks;Large Language Models (LLM);Named Entity Recognition (NER);Prompt Engineering;Legal Portuguese domain},

doi={10.1109/IJCNN64981.2025.11227838}}