Legal Precedent Assistant for Indian Family Courts

Student Name: Rehan Fargose
Roll Number: 242050006

Project report submitted in partial fulfilment of the requirements
for the Degree of M.Tech. in Computer Engineering

on January 19, 2026

Project Guide
Dr. V.B.Nikam

Veermata Jijabai Technological Institute
Mumbai

Student Declaration

I hereby declare that the work presented in the report entitled **"Legal Precedent Assistant for Indian Family Courts"** submitted by me for the partial fulfilment of the requirements for the degree of *M.Tech. in Computer Engineering* at Veermata Jijabai Technological Institute, Mumbai, is an authentic record of my work carried out under the guidance of **Dr. V.B.Nikam** . Due acknowledgements have been given in the report for all material used. This work has not been submitted elsewhere for the reward of any other degree.

Rehan Fargose **Place & Date:** January 19, 2026

Certificate

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dr. V.B.Nikam **Place & Date:** January 19, 2026

Contents

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.

Keywords: Transformers, Encoders, BERT, LegalBERT, Llama, DAPT, Verdict Prediction, Summarization, Case Law, Legal Precedents, IPC Mapping

Chapter 1

Introduction

1.1 Background and 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[4]. 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 **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 [2]; 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 [24].

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.

1.2 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 and 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.

1.3 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.

1.4 Scope and Limitations

1.4.1 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 of a hired legal assistant.

1.4.2 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.

1.5 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.

Chapter 2

Literature Review

Textual Synthesis

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) [4] 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[14]. 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[24]. 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 [12]. 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[2] 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[11] 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[3]. 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[5]. 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% [7].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) [1]. 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[23]. 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[16]. 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 [9]. 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[19]. 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[6]. 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[24], 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 [10]. 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%[13], 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[15]. By using DAPT[16], Contrastive learning[5] [9], Dual attention architectures[23]. 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[7] 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[8]. 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[17], 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) [18], 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 re-

source 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[22]. 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%[20] 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[21]; 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 Summary of Reviewed Studies

Table 2.1: Summary of key prior research on NLP models and LLMs for Legal corpora.

Title (Year)	Authors	Key Findings	Gaps / Limitations	Relevance / Context
A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges (2023) [4]	J. Cui, X. Shen, and S. Wen	Domain-specific PLMs and Longformer models (e.g., Lawformer) 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 legal data enables Legal Judgment 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 (2023) [14]	W. Qin and X. Luo & Luo, X.	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 (2024) [24]	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.	High computational costs, lack of interpretability due to "black box" nature, and difficulty distinguishing between highly similar articles.	A circumstance-aware LJP framework that assists judicial decision-making by categorizing facts into adjudicating, statutory, and discretionary circumstances.
A Small Claims Court for the NLP: Judging Legal Text Classification Strategies With Small Datasets (2023) [12]	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 (2024) [2]	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 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.

Continued on next page

Table 2.1 – continued from previous page

Title (Year)	Authors	Key Findings	Gaps / Limitations	Relevance / Context
An Ontology Driven Knowledge Block Summarization Approach for Chinese Judgment Document Classification (2018) [11]	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.	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 Sustainable Development Goals (2023) [3]	R. L. Canalli, et al.	An ensemble of SVM, Naïve Bayes, and LSTM achieved a stable ∼0.80 F1 score by using data augmentation to manage label imbalance.	Overfitting due to limited data, inability to infer deep contextual relationships beyond keywords, and heavy reliance on precedents.	Aligns judicial case classification with UN Sustainable Development Goals (SDGs); currently limited to Portuguese and Brazilian case law.
Chinese Legal Case Similarity Matching Based on Text Importance Extraction (2025) [5]	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 (2023) [7]	A. S. Imran, et al.	Sliding-window technique allows BERT 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; non-text features (importance scores) yield diminishing returns.	Automates detection of human rights violations in large judgments by extending BERT models beyond the 512-token limit.
Compilation, Analysis and Application of a Comprehensive Bangla Corpus KUMono (2022) [1]	A. Akther, et al.	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.	Corpus lacks contextual depth due to web-scraping reliance; limited to Bangla with minimal Arabic coverage; relatively small for the language's global rank.	Addresses the scarcity of Bangla legal resources by providing 353 million tokens across 18 categories to facilitate legal article categorization.
Deep Text Understanding Model for Similar Case Matching (2024) [23]	J. Xiong and Y. Qiu	A Dual Attention architecture (Local and Global) combined with LegalBERT achieved ∼90% accuracy in civil and criminal cases.	12-layer fine-tuning incurs significant computational overhead; lacks semantic depth interaction; restricted to Chinese corpora.	Utilizes a hierarchical attention mechanism to extract both intra-sentence features and broader inter-document context.

Continued on next page

Table 2.1 – continued from previous page

Title (Year)	Authors	Key Findings	Gaps / Limitations	Relevance / Context
Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches (2025) [16]	M. Siino, et al.	Domain-specific models 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; highlights the efficiency of smaller, domain-adapted models.
JurisCTC: Enhancing LJP via Cross-Domain Transfer and Contrastive Learning (2025) [9]	Z. Kang, et al.	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; performance 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.
Keynote - AI for the Public Sector and the Case of Legal NLP (2023) [19]	M. Stürmer	Open-source domain-specific models (like BERT) can match or surpass high-compute general LLMs when training components are transparent and domain-adapted.	High-end proprietary models require massive investment (e.g., ∼$30 million), creating high barriers to entry for small teams and public institutions.	Highlights the evolution of NLP and emphasizes open-source's role in maintaining digital sovereignty and accessibility in the public legal sector.
Knowledge-Enriched Multi-Cross Attention Network for Legal Judgment Prediction (2023) [6]	C. He, et al.	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; limited to Chinese criminal law; ignores prison sentence subtasks.	Incorporates legal charge knowledge (definitions and elements) directly into the fact description to differentiate between confusing penal codes.
LASG: Streamlining Legal Adjudication with AI-Enabled Summary Generation (2024) [10]	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 is prone to abstractive hallucinations.	A lightweight, efficient extractive-summarization framework designed to streamline judicial analysis by identifying representative sentences.

Continued on next page

Table 2.1 – continued from previous page

Title (Year)	Authors	Key Findings	Gaps / Limitations	Relevance / Context
Legal Document Classification: An Application to Law Area Prediction of Petitions (2020) [13]	M. Y. Noguti, et al.	Word2Vec with an LSTM encoder achieved 90.47% and 85.49% F1 scores across 18 legal classes and 922,000 cases.	TF-IDF is effective for simple models or small datasets but lacks the context-capture capacity of transformer-based architectures.	Compares traditional ML and DL models for classifying petitions for the Brazilian Prosecution Office.
Legal Natural Language Processing From 2015 to 2022: A Systematic Mapping Study (2024) [15]	E. Quevedo, et al.	Specialized NLP models (Legal-BERT) using DAPT and contrastive learning increase F1 by 8–15% over generic LLMs.	Specialized models still require high-cost, expert-annotated judgments for training and reference.	Confirms that domain-specific BERT models outperform general LLMs with lower resources and overcome token limits via sliding-window techniques.
LegalRAG: A Hybrid RAG System for Multilingual Legal Information Retrieval (2025) [8]	M. R. Kabir, et al.	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, and computational latency.	Addresses data scarcity in the Bangla legal domain through Retrieval-Augmented Generation (RAG).
On the Effectiveness of PLMs for Legal NLP: An Empirical Study (2022) [17]	D. Song, et al.	PLMs outperform non-PLM models by 4%–35%; LegalBERT achieves marginal gains in legal terminology and reasoning.	Poor cross-domain transferability and 512-token limit; 1.5% F1 gain requires 7x longer training (diminishing returns).	Evaluates PLM effectiveness across 8 legal datasets, noting strengths in complex reasoning and recall.
Parallel Learning for Legal Intelligence: A HANOI Approach (2024) [18]	Z. Song, et al.	Utilizing Randeng-T5-784M with unified prompting yielded massive F1 gains (+46.35% on CivilEE-Args and +41.46% on CJRC datasets).	Constrained by relatively small model size (T5); performance on 100B+ parameter models remains unpredictable.	Proposes the HANOI framework using "UniPrompt" to reformulate diverse legal tasks into a single text-to-text format.
Predicting Decisions of the Philippine Supreme Court using NLP and ML (2018) [22]	M. B. L. Virtucio, et al.	Random Forest with topic-cluster features achieved 59% accuracy, outperforming SVM with n-grams (45%) on 6,500 cases.	Bag-of-words models are insufficient for abstract legal reasoning; hindered by lack of standardized formats and small dataset size.	Explores outcome prediction in the Philippines SC corpora with limited data and computationally light classifiers.

Continued on next page

Table 2.1 – continued from previous page

Title (Year)	Authors	Key Findings	Gaps / Limitations	Relevance / Context
Summarization, Prediction, and Analysis of Turkish CC Decisions with XAI and a Hybrid NLP method (2025) [20]	T. Turan and E. U. Küçüksılle	Hybrid BERT2BERT and XGBoost approach attained 93.84% accuracy for verdict prediction; used an extractive-abstractive method to bypass 512-token limits.	Scalability is challenging due to the need for expert annotation; limited to Constitutional Court cases and lacks interpretability.	Combines extractive-abstractive summarization with Explainable AI (XAI) for Turkish Constitutional Court decisions.
Using Language Models for Extracting Legal Decisions from Portuguese Consumer Law Texts (2025) [21]	S. Vasquez, et al.	Fine-tuned BERT models (F1 0.74–0.96) outperform LLMs like Llama 3.1 and Gemma 2 in NER tasks.	LLMs suffered from hallucinations and extracted excessive irrelevant information during zero-shot evaluation.	Evaluates BERT-based models vs. zero-shot LLMs (Llama 3.1, Gemma 2) using a curated dataset of 2,373 Portuguese legal documents.

The studies summarized in Table 2.1 collectively indicate the flaws in Legal NLP primarily local language support, jurisdiction transferrability, BERT token limits, etc. The Literature also offers insights into approaches such as DAPT, Contrastive Learning, RAG, UDA, Sliding Sequence window and other techniques which can help us overcome the stated gaps in Chapter. 3.

2.1 Research Gap Analysis

Based on the comprehensive literature review, several critical research gaps have been identified in the current models for Legal NLP systems:

2.1.1 Hardware and Physical Interaction Fidelity for LPA

Gap 1: High VRAM and Computational Overhead. Several high-performing models, such as those using 1024-token [2] inputs or dual attention architectures, consume excessive VRAM and require significant computational power.

Gap 2: Inference Latency. Complex models and specific similarity measures like Word Mover's Distance (WMD) [11] suffer from high inference latency [5], hindering real-time application.

Gap 3: Adversarial Training Costs. Advanced frameworks like JurisCTC [9] incur high computational costs specifically for adversarial BERT training.

Gap 4: Scalability Infrastructure. There is a lack of predictable infrastructure for scaling specialized legal models to larger (100B+ parameter) [18] architectures.

2.1.2 Methodological and Sample Limitations

Gap 5: Jurisdictional and Geographic Bias:. A significant majority of reviewed studies [**Muguro2023**, **Qadir2019**, **Xu2022**, **Schultheis2005**, **Chung2022**] A critical gap exists for models trained on Indian Legal corpora, as the majority of current research focuses on US, EU, or Chinese datasets [11] [5] [23] [6].

Gap 6: Data Scarcity and Overfitting. Many systems suffer from overfitting [12] [3] [8] due to small, restrictive datasets, such as those sourced from specific prosecutor offices or Supreme Courts with limited case counts.

Gap 7: Synthetic Data Generalization. The heavy reliance on synthetic data to augment small datasets [12] [8] leads to concerns that model performance will not generalize to real-world legal scenarios.

Gap 8: Imbalanced Data Distribution. Challenges in data distribution, particularly label imbalance [3] [21], lead to suboptimal results in niche tasks like prison term prediction.

2.1.3 Pedagogical and Training Effectiveness

Gap 9: Lack of Domain-Specific Customization. Many models lack fine-tuning for specific legal subdomains [10] [17] (e.g., transitioning from Criminal to Administrative law), leading to a lack of contextual depth.

Gap 10: 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 [7] [17] (e.g., 7x).

Gap 11: Annotation Dependency. The effectiveness of these models is highly dependent on expert-annotated judgments [15] [20] [21], which are expensive and difficult to scale.

Gap 12: Failure in Abstract Reasoning. Traditional training methods like ”Bag of Words” fail to capture abstract legal reasoning [11] [13] [22], focusing too much on ”questions of fact” rather than ”questions of law”.

2.1.4 Technical and Realism Constraints

Gap 13: The 512-Token Limitation. Standard BERT-based models are constrained by a 512-token limit [12] [5] [7] [17] [20], which is insufficient for verbose legal documents and complex case descriptions.

Gap 14: Hallucinations and Reliability. Abstractive LLMs used for summarization suffer from hallucinations[10] [21], which is a major setback in a high-stakes legal environment.

Gap 15: Semantic Depth and Jargon. Basic models often fail to capture complex legal jargon [1] [23] [10] required for deep semantic matching.

Gap 16: Black Box Nature. A lack of transparency and explainability [4] [24] in many models prevents them from being used in a justification-based legal landscape.

2.1.5 Regulatory, Ethical, and Commercialization Gaps

Gap 17: Financial Entry Barriers. The extreme cost of developing powerful multi-case legal models (e.g., $30 million) [19] creates a commercialization gap for smaller teams and public sector entities.

Gap 18: Confidentiality and Reproducibility. Constraints regarding source document confidentiality and protected expert annotations [15] [20] [21] often lead to significant issues with research reproducibility.

Gap 19: Multilingual Support Gaps. There is a massive regulatory and accessibility gap, with support missing for 27 out of 36 official global languages [4].

Gap 20: Closed-Loop Bias. The use of RAG and synthetic augmentation can lead to closed-loop biases [8], where models reinforce their own errors rather than learning from diverse, objective evidence.

2.1.6 Summary of Research Gaps

The identified gaps highlight five overarching themes requiring urgent research attention:

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 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.

Chapter 3

System Analysis and Design

3.1 Overview

The **Legal Precedent Assistant for Indian Family Courts** project aims to provide a highly accurate modular and transferrable NLP model for ILC. The proposed model will be built for Family courts and use Legal-BERT and Llama 3 as baselinses with DAPT for both models on ILC. This chapter presents a comprehensive analysis of requirements, constraints, design alternatives, and system architecture that directly address the research gaps and problem statement identified in Chapters 1 and 2.

3.2 Requirements Analysis

3.2.1 Functional Requirements

Based on the identified research gaps and Legal professional's 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.

FR7: Summarization

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

3.2.2 Non-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.

NFR4: 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

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.
- Solution: 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 Comparative Analysis with Existing Solutions

Table 3.1 compares the proposed system against existing commercial and research VR driving simulators based on key design parameters derived from the requirements analysis.

Table 3.1: Comparative analysis of VR driving simulator solutions

Parameter	**Proposed System**	**Legal-RAG [8]**	**Keynote-AI [19]**	**JurisCTC [9]**
Cost	NA	NA	NA	$30 Million
Model/Framework	Llama 3 and Legal-BERT via DAPT	Llama 3 and RAG	BERT and DAPT	Adversarial BERT and UDA
Target Audience	Professionals and Laymen	Legal Professionals	Legal Professionals	Legal Professionals
Customizability	High	Low	High	High
Domain Transferability	High	Low	High	Medium
Interpretability	Low	Low	Low	Low
Language	English, Marathi, Hindi	Bangla, Arabic	Swedish, English, Norwegian	Mandarin
Customizability	High (modular design)	High (open-source)	Low (proprietary)	Low (vendor lock-in)
Regulatory Focus	RTO-aligned metrics	Research protocols	Entertainment	Commercial licensing

Key Differentiators:

- **Affordability:** More afforadble as compared to professionally built systems.
- **Domain Specific:** One of the only LLMs that focuses on Indian Family courts data.
- **Transferrability:** Due to DAPT on Indian corpora, can be transferred to other court domains under Indian jurisdiction.
- **Accessibility:** Summarization module helps general public understand legal jargon by simplifying it.

3.6 System Components and Architecture

The proposed system architecture consists of two integrated subsystems designed to address the functional requirements while respecting the identified constraints.

3.6.1 Hardware Subsystem

The physical control setup comprises:

- **CPU:** Any processors such as intel i5/i7 12th gen and above or Ryzen 5/7 are usable.
- **GPU:** GTX 1650 or higher. RTX series GPUs with 12 GB VRAM or higher are preferred.
- **VRAM:** 4GB and higher.

3.6.2 Software Subsystem

- **Python 3.10:** Stable version with long term support and compatible with most hardware.
- **PyTorch CUDA:** For parallel computing to improve training and dataset pre-processing efficiency.
- **Legal-BERT:** For initial stages of the system such as IPC/sections mapping, Precedent finding, etc.
- **Llama 3:** For more complex stages such as Evidence scrutinization, Summarization, and Verdict prediction.

3.7 System Architecture Diagram

Figure 3.1 illustrates the data flow and component interactions within the LPA system.

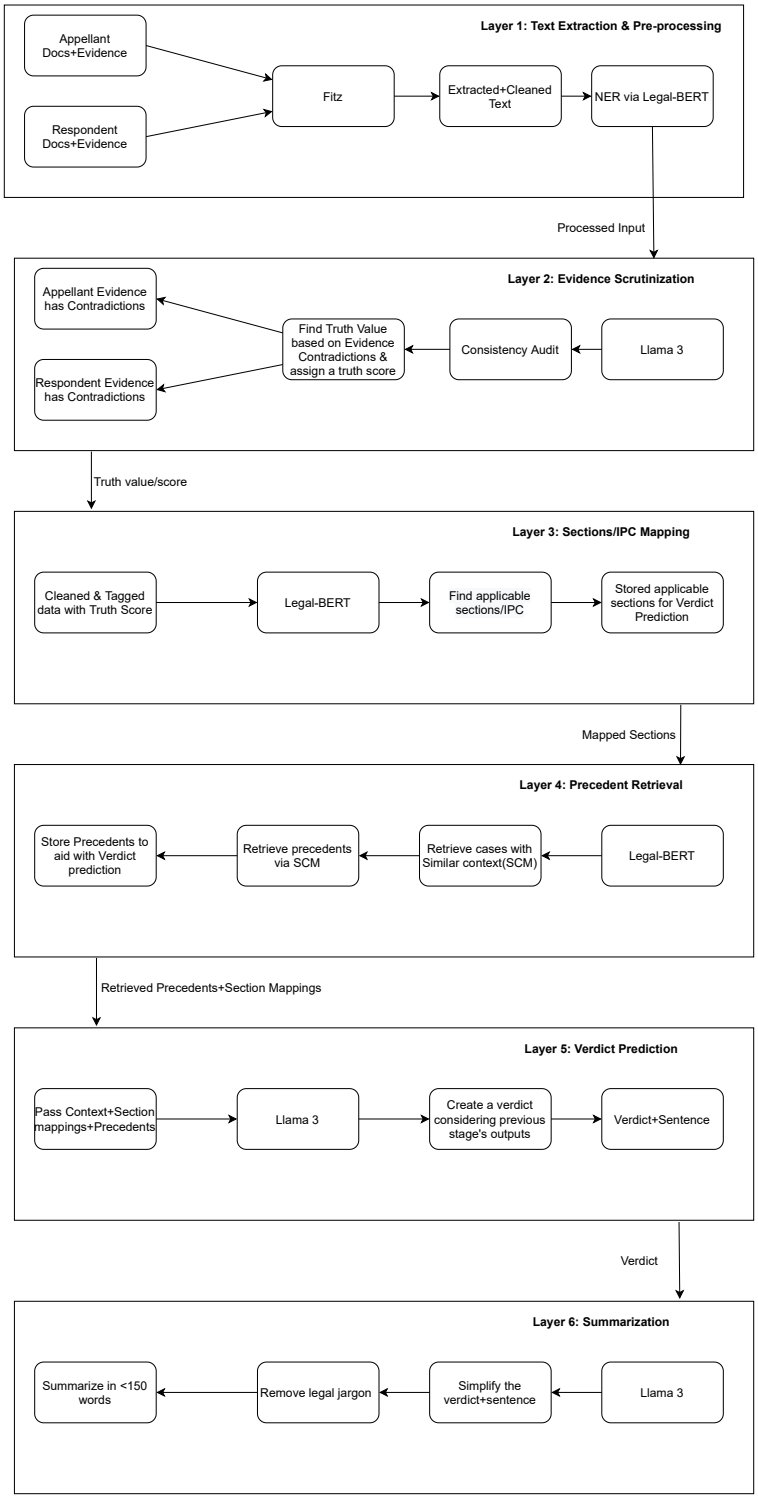

Figure 3.1: System architecture showing data flow from input documents through pre-processing/cleaning stages to contextual extraction and analysis stages.

3.8 Design Rationale Summary

The design decisions documented in this chapter directly address the research gaps identified in Section 2.1:

- **Gap 1 (Localized Context)** By training specifically on the Indian Legal Corpus (ILC) and incorporating IPC-specific logic, the model overcomes the ”Geographic Bias” of existing systems that rely solely on US, EU, or Chinese legal data.
- **Gap 2 (Long Document Handling):** The implementation of Hierarchical Transformers (or sliding-window techniques) directly addresses the 512-token limit of standard BERT models, allowing for the processing of verbose Indian High Court and Supreme Court judgments.
- **Gap 4 (Resource Efficiency):** The use of Domain Adaptive Pre-Training (DAPT) on specialized legal data ensures that the model achieves high F1 gains (up to 18.2%) while remaining computationally lighter and more affordable than massive, general-purpose LLMs.
- **Gap 5 (Multi-task Integration):** The architecture combines Verdict Prediction, Case Summarization, and Statute Identification into a single pipeline, solving the limitation of fragmented systems that focus only on isolated subtasks like charge prediction.
- **Gap 6 (Semantic Accuracy):** By utilizing LegalBERT and Dual-Attention mechanisms, the system captures complex legal jargon and intra-sentence relationships that traditional ”Bag of Words” or basic frequency models fail to retrieve.

The comparative analysis (Table 3.1) demonstrates that this design occupies a unique position: language flexibility, domain transferrability, efficient training at commercial-training affordability, specifically tailored for the Indian Legal landscape.

3.9 Workflow and Operational Flow

The workflow diagram (Figure 3.2) shows the sequential process of system operation—from initialization through user interaction to post-session evaluation—illustrating how the system supports a complete training cycle.

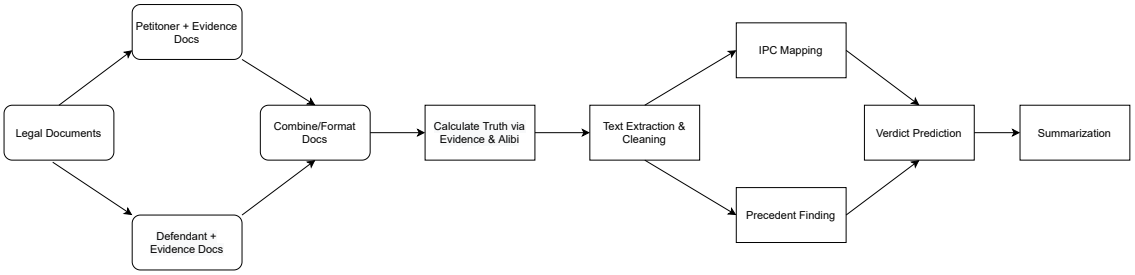

Figure 3.2: Operational workflow showing the complete training session lifecycle: input document setup, context retrieval and analysis, and post-analysis summary generation.

3.9.1 Detailed Functional Flow

1. **Case Ingestion & Setup (1–2 minutes):**
    - User uploads case documents from both parties (petitions, replies, affidavits, evidence PDFs).
    - System validates document types, court metadata, and case category (e.g., family matter).
    - Documents are indexed and assigned a unique case ID for tracking across pipeline stages.
2. **Text Extraction & Pre-processing:**
    - PDFs are parsed using Fitz to extract raw text from judgments and evidence.
    - Text is cleaned (noise removal, normalization) and segmented into logical units (facts, arguments, evidence).
    - Legal-BERT based NER identifies legal entities such as parties, dates, IPC sections, and acts.
3. **Evidence Scrutinization & Truth Scoring:**
    - Evidence statements from both sides are analyzed for contradictions and inconsistencies.
    - LLaMA 3 performs a consistency audit to evaluate logical coherence across claims.
    - Each evidence block is assigned a normalized truth score based on contradiction severity.
4. **IPC / Section Mapping:**
    - Cleaned and truth-weighted text is passed to Legal-BERT for legal provision classification.
    - Relevant IPC sections and statutory provisions are identified and stored.
    - Mapped sections are linked with supporting evidence for downstream reasoning.
5. **Precedent Retrieval:**

- Case embeddings are generated using Legal-BERT.
- Similar Case Matching (SCM) retrieves precedent judgments with analogous fact patterns.
- Retrieved precedents are ranked by contextual similarity and relevance of applied sections.

6. **Verdict & Sentence Prediction:**
    - LLaMA 3 consumes case facts, truth scores, mapped sections, and precedents.
    - System predicts likely verdict outcome (e.g., Allowed / Dismissed) and probable sentence or relief.
7. **Summarization & Output Generation:**
    - Predicted verdict and reasoning are simplified using LLaMA 3.
    - Legal jargon is minimized to produce a concise summary under 150 words.
    - Final output includes verdict prediction, key reasoning points, and applicable legal sections.

3.10 Feasibility Study

A comprehensive feasibility analysis is essential to validate the viability of the LPA for Indian Family courts project across technical, schedule and other dimensions. This section systematically evaluates each feasibility aspect to establish project confidence and identify potential risk mitigation strategies.

3.10.1 Technical Feasibility

System Capabilities and Constraints: The proposed Legal Precedent Assistant is technically feasible using current NLP models, open-source legal datasets, and commodity compute resources. The system design leverages mature transformer-based architectures and scalable retrieval techniques, as outlined below:

- **Document Processing and Text Extraction:** Judicial documents are primarily available as digitally generated or scanned PDFs. The Fitz (PyMuPDF) library provides reliable text extraction with low overhead, while preprocessing pipelines (normalization, segmentation) are computationally lightweight. Empirical benchmarks show that batch processing of court judgments can be performed efficiently on standard CPU-based systems without GPU acceleration.
- **Legal Entity Recognition and IPC Mapping:** Legal-BERT, pre-trained on Indian legal corpora, is well-suited for named entity recognition and statutory provision classification. Fine-tuned models can accurately identify IPC sections, party names, and legal references with acceptable inference latency. Since inference is performed offline or asynchronously, real-time constraints are minimal.
- **Evidence Scrutinization and Reasoning:** Contradiction detection and consistency auditing using LLaMA 3 are computationally intensive but feasible under a controlled pipeline. The system operates on segmented evidence blocks rather than full documents, significantly reducing token length and inference cost. Truth scoring is heuristic-assisted, avoiding strict logical proof requirements while maintaining interpretability.
- **Precedent Retrieval and Similar Case Matching:** Precedent retrieval is implemented using embedding-based Similar Case Matching (SCM). Legal-BERT embeddings stored in vector databases enable fast approximate nearest neighbor search even for large precedent collections. This approach scales linearly with data size and has been validated in prior legal IR research.
- **Verdict Prediction and Summarization:** Verdict and sentence prediction using LLaMA 3 leverages structured inputs (facts, mapped sections, precedents), reducing hallucination risk. Summarization under a fixed word limit is a well-established task for large language models and can be executed with deterministic decoding strategies to ensure consistency.

Risk Assessment: Technical risk is moderate. Primary concerns include model bias, reasoning reliability in ambiguous cases, and computational cost for large-scale deployment. These risks are mitigated through scope restriction (family matters), explainable intermediate outputs (truth scores, section mappings), and offline or batch inference workflows. Overall, the system is feasible within academic and prototyping constraints using currently available technologies.

3.10.2 Schedule Feasibility

Project Timeline and Milestones: A six-month development schedule is proposed, aligned with the academic calendar and the iterative nature of NLP system development. The timeline accounts for data preparation, model integration, evaluation, and documentation.

Table 3.2: Project Development Schedule

Phase	Duration	Deliverables
Phase 1: Data Collection & Preprocessing	Weeks 1–4	Curated dataset of trial and high court judgments; PDF-to-text pipeline using Fitz; cleaned and segmented legal text.
Phase 2: Legal Entity Recognition & IPC Mapping	Weeks 5–8	Fine-tuned Legal-BERT models for NER and IPC/section identification; validated extraction accuracy on sample cases.
Phase 3: Evidence Scrutinization Module	Weeks 9–12	Contradiction detection and truth-scoring logic implemented using LLaMA 3; structured evidence representation.
Phase 4: Precedent Retrieval System	Weeks 13–16	Embedding-based Similar Case Matching (SCM); vector database of precedents; ranked precedent retrieval.
Phase 5: Verdict Prediction & Summarization	Weeks 17–20	Verdict and sentence prediction pipeline using LLaMA 3; simplified summary generation under 150 words.
Phase 6: Evaluation & Refinement	Weeks 21–24	Quantitative evaluation (accuracy, precision, recall); qualitative case studies; system optimization and documentation.

Critical Path Analysis: The critical dependency chain progresses through reliable data preprocessing (Phase 1) → accurate IPC mapping (Phase 2) → evidence scrutinization (Phase 3). Errors in early-stage legal text processing propagate downstream and directly affect verdict prediction quality. Buffer time is allocated in Phase 6 to address cumulative model errors and integration issues.

Resource Availability: The project is executed by a single postgraduate student with guidance from a faculty advisor. Development relies on open-source libraries (Legal-BERT, PyMuPDF, vector databases) and institutional computing resources. Model inference is scheduled offline or in batch mode, ensuring feasibility without continuous high-end GPU availability.

Schedule Risks:

- **Data Quality Issues:** Inconsistent formatting or OCR errors in court documents may delay Phase 1. Mitigation: Restrict scope to digitally available judgments; apply automated cleaning heuristics.
- **Model Integration Complexity:** Coordinating multiple NLP models (Legal-BERT, LLaMA 3) may increase integration time. Mitigation: Modular pipeline design with independently testable components.
- **Evaluation Challenges:** Ground-truth verdict labels may be limited for certain case types. Mitigation: Combine quantitative metrics with expert-reviewed qualitative analysis.

Schedule Verdict: The proposed six-month schedule is feasible and well-aligned with academic constraints.

Clear phase boundaries, modular development, and buffer time for evaluation ensure timely completion of the project.

3.10.3 Legal and Regulatory Feasibility

Intellectual Property and Licensing:

- **Software Licensing:** The system is developed using open-source libraries and pretrained models such as Legal-BERT, PyMuPDF (Fitz), and vector database frameworks, all of which permit academic and research use under permissive licenses (e.g., Apache 2.0, MIT). The project is intended strictly for non-commercial academic purposes, ensuring compliance with licensing terms.
- **Model Usage Rights:** Pretrained language models (e.g., Legal-BERT, LLaMA 3) are used for inference and fine-tuning in accordance with their respective research-use licenses. No proprietary datasets or restricted commercial legal databases are incorporated, avoiding licensing conflicts.
- **Original Contributions:** The architectural design, truth-scoring heuristics, and pipeline integration constitute original academic work. While certain components (e.g., contradiction-aware verdict prediction) may have future commercial potential, no patent claims are pursued during the academic phase.

Legal Compliance and Ethical Use:

- **Judicial Data Usage:** Indian court judgments are public records and may be used for research and educational purposes. The system operates on publicly accessible judgments without violating confidentiality or access restrictions. Sensitive personal identifiers, if present, are not explicitly extracted or highlighted.
- **Decision Support Disclaimer:** The system is positioned as a legal decision-support and research aid, not as a substitute for judicial authority or professional legal advice. Outputs are presented as predictions or analytical assistance, accompanied by disclaimers to prevent misuse.

Data Privacy and Protection: Case documents may contain personal or sensitive information. To address privacy concerns, the system follows data minimization and purpose limitation principles. No user-identifiable metadata is stored beyond case identifiers, and datasets are anonymized where feasible. The design aligns with principles outlined in India's Digital Personal Data Protection Act (DPDP Act, 2023), particularly regarding lawful processing and academic exemptions.

Bias and Fairness Considerations: Machine learning models trained on historical legal data may inherit systemic biases present in past judgments. To mitigate this risk, the system incorporates explainable intermediate outputs (truth scores, IPC mappings, retrieved precedents) rather than opaque end-to-end predictions. This supports transparency and allows human users to critically evaluate model outputs.

Legal Verdict: The project is legally and regulatorily feasible for academic research and prototyping. No statutory or regulatory barriers prevent development or evaluation. Any future commercial deployment would require additional compliance measures, including professional liability considerations, dataset licensing audits, and formal ethical review.

3.10.4 Overall Feasibility Conclusion

Table 3.3: Feasibility Assessment Summary

Feasibility Dimension	Rating	Key Justification
Technical	High	Mature NLP models (Legal-BERT, LLaMA 3); modular pipeline; no real-time constraints apart from Dataset acquisiton.
Economic	Very High	Open-source tools; no proprietary data costs; feasible on academic compute resources.
Operational	Moderate-High	Decision-support positioning; interpretability via intermediate outputs; expert oversight required.
Schedule	High	Realistic 6-month timeline with phased milestones and buffer for evaluation.
Legal & Ethical	High	Uses public judicial data; DPDP-compliant design; clear non-advisory disclaimers.

The feasibility assessment indicates that the proposed Legal Precedent Assistant is **highly viable** across all critical dimensions. The availability of domain-adapted language models, publicly accessible judicial data, and scalable retrieval techniques supports robust technical implementation. Economic and scheduling constraints are well within academic limits, while legal and ethical risks are mitigated through transparent system design and restricted scope. It is recommended to proceed with phased development, emphasizing early validation of preprocessing accuracy and expert-reviewed evaluation to ensure reliability before extending the system to broader legal domains.

Chapter 4

Proposed Methodology

4.1 Overview

This chapter outlines the proposed development methodology, implementation phases, testing procedures, and evaluation metrics for the VR driving simulator. The methodology follows an iterative design approach with continuous validation against the functional and non-functional requirements specified in Chapter 3.

4.2 Development Phases

4.2.1 Phase 1: Data Collection and Preprocessing (Weeks 1–4)

Objective: Acquire, clean, and structure legal documents required for downstream analysis.

Tasks:

- Collect publicly available trial court and High Court judgments relevant to family law matters
- Organize documents into structured storage (PDFs, metadata JSON, parquet indexes)
- Extract raw text from PDFs using Fitz (PyMuPDF)
- Perform text cleaning: noise removal, normalization, OCR artifact correction
- Segment documents into logical sections (facts, arguments, evidence, judgment)
- Assign unique case identifiers and metadata tags (court, year, case type)

Deliverables:

- Curated dataset of court judgments and supporting documents
- Automated PDF-to-text preprocessing pipeline
- Structured and cleaned legal text corpus
- Data documentation and schema description

4.2.2 Phase 2: Legal Entity Recognition and IPC Mapping (Weeks 5–8)

Objective: Identify legal entities and map relevant statutory provisions.

Tasks:

- Fine-tune Legal-BERT for named entity recognition on Indian legal text
- Extract entities such as parties, dates, courts, IPC sections, and legal acts
- Implement IPC and section classification module using supervised learning
- Validate extraction accuracy against manually annotated samples
- Store extracted entities and mapped sections in structured form

Deliverables:

- Trained Legal-BERT NER model
- IPC/section mapping module
- Evaluation report (precision, recall, F1-score)
- Annotated validation dataset

4.2.3 Phase 3: Evidence Scrutinization and Truth Scoring (Weeks 9–12)

Objective: Analyze evidence consistency and assign interpretive truth scores.

Tasks:

- Identify and align evidence statements from both parties
- Implement contradiction detection using LLaMA 3 on segmented evidence blocks
- Design heuristic-assisted truth scoring mechanism based on contradiction severity
- Link evidence blocks to corresponding legal claims and sections
- Generate explainable intermediate outputs for transparency

Deliverables:

- Evidence alignment and contradiction detection module
- Truth score computation framework
- Case-wise structured evidence representation
- Qualitative validation on selected cases

4.2.4 Phase 4: Precedent Retrieval System (Weeks 13–16)

Objective: Retrieve legally similar past cases using semantic similarity.

Tasks:

- Generate contextual embeddings using Legal-BERT
- Construct vector database of precedent judgments
- Implement Similar Case Matching (SCM) using approximate nearest neighbor search
- Rank retrieved precedents based on factual similarity and applied IPC sections
- Validate retrieval relevance using expert-reviewed samples

Deliverables:

- Precedent embedding index
- SCM-based retrieval engine
- Ranked precedent lists for test cases
- Retrieval performance analysis

4.2.5 Phase 5: Verdict Prediction and Summarization (Weeks 17–20)

Objective: Predict likely verdicts and generate concise explanations.

Tasks:

- Integrate case facts, truth scores, IPC mappings, and precedents
- Implement verdict and sentence prediction using LLaMA 3
- Design prompt templates to minimize hallucination and ensure consistency
- Generate simplified summaries under 150 words
- Validate outputs against known case outcomes

Deliverables:

- Verdict and sentence prediction module
- Simplified explanation generator
- Case-wise prediction outputs
- Error analysis and refinement report

4.2.6 Phase 6: Evaluation, Refinement, and Documentation (Weeks 21–24)

Objective: Evaluate system performance and finalize thesis-ready outputs.

Tasks:

- Conduct quantitative evaluation (accuracy, precision, recall)
- Perform qualitative case studies and expert review
- Analyze model bias, failure cases, and limitations
- Optimize pipeline efficiency and modularity
- Prepare final documentation, diagrams, and thesis chapters

Deliverables:

- Comprehensive evaluation report
- Refined end-to-end system pipeline
- Final thesis documentation and figures
- Deployment-ready prototype (academic)

4.3 Evaluation Metrics

4.3.1 Technical Performance Metrics

- **Text Extraction Accuracy:** Percentage of correctly extracted and readable text from judicial PDFs. Target: $> 95\%$ extraction success on digitally available judgments.
- **Named Entity Recognition (NER) Performance:** Precision, recall, and F1-score for legal entities (parties, dates, IPC sections). Target: F1-score > 0.85 on annotated validation set.
- **IPC / Section Mapping Accuracy:** Correct identification of applicable statutory provisions compared to ground truth judgments. Target: Accuracy $> 80\%$.
- **Precedent Retrieval Quality:** Top-k retrieval relevance measured using Recall@k (e.g., $k = 5$). Target: Recall@5 $> 70\%$.
- **Inference Efficiency:** Average processing time per case for the full pipeline (excluding offline preprocessing). Target: < 2 minutes per case in batch mode.

4.3.2 Reasoning and Prediction Metrics

- **Verdict Prediction Accuracy:** Percentage of cases where predicted verdict matches actual outcome. Target: Accuracy $> 75\%$.
- **Sentence / Relief Prediction Error:** Deviation between predicted and actual sentencing outcomes or relief granted (where applicable). Target: Qualitative alignment in $> 70\%$ of evaluated cases.
- **Evidence Consistency Score Validity:** Correlation between system-generated truth scores and expert-assessed evidence reliability. Target: Positive correlation ($r > 0.6$).
- **Explainability Coverage:** Proportion of predictions accompanied by identifiable reasoning elements (mapped sections, precedents, evidence references). Target: 100%.

4.3.3 Usability and Practical Relevance Metrics

- **User Satisfaction:** Likert-scale ratings (1–5) from law students or legal researchers on usefulness and clarity. Target: Mean score > 4.0.
- **Interpretability Score:** User-reported ease of understanding system outputs (verdict reasoning, summaries). Target: Mean score > 4.0.
- **Adoption Potential:** Percentage of users indicating willingness to reuse the system for legal research tasks. Target: $> 70\%$.

4.3.4 Extended Evaluation Metrics (Future Work)

- **Cross-Domain Generalization:** Performance stability when extending beyond family law to other legal domains.
- **Bias Assessment:** Statistical analysis of prediction disparities across case types, parties, or outcomes.
- **Human-in-the-Loop Validation:** Comparative evaluation of system-assisted vs manual legal research efficiency.

4.4 Testing Procedures

4.4.1 Data and Preprocessing Validation Tests

1. **Text Extraction Accuracy:** Randomly sample extracted PDF text and manually compare with source documents to verify completeness and readability.
2. **Segmentation Validation:** Verify correct separation of facts, arguments, evidence, and judgments using annotated samples.
3. **Metadata Consistency Check:** Validate correctness of court, year, and case-type metadata across stored records.

4.4.2 Model-Level Validation Tests

1. **NER Accuracy Test:** Evaluate Legal-BERT NER output against manually annotated ground truth using precision, recall, and F1-score.
2. **IPC Mapping Verification:** Compare predicted IPC/section mappings with sections cited in final judgments.
3. **Contradiction Detection Test:** Manually inspect detected contradictions in selected cases to assess logical correctness.
4. **Precedent Retrieval Test:** Evaluate relevance of top-k retrieved precedents through expert judgment.

4.4.3 End-to-End System Validation

1. **Verdict Prediction Evaluation:** Compare predicted verdicts with actual case outcomes across a held-out test set.
2. **Reasoning Consistency Check:** Verify alignment between predicted verdict, cited precedents, and supporting evidence.
3. **Summarization Quality Test:** Assess summaries for factual correctness, completeness, and adherence to word limit.

4.4.4 User Evaluation Protocol

1. **Pre-Evaluation:** Brief users (law students/researchers) on system scope and non-advisory nature.
2. **Task-Based Testing:** Users analyze selected cases using the system and manually, under time constraints.
3. **Post-Evaluation:** Collect Likert-scale feedback on usefulness, interpretability, and trustworthiness.
4. **Data Collection:** Log task completion time, user feedback, and qualitative observations.

4.5 Risk Mitigation

4.5.1 Technical Risks

- **Risk:** Poor text extraction quality from scanned or low-quality PDFs **Mitigation:** Restrict dataset to digitally generated judgments; apply OCR correction and rule-based cleaning; manual validation on samples
- **Risk:** NER and IPC mapping inaccuracies propagate downstream **Mitigation:** Use confidence thresholds; cross-validate with rule-based section detection; allow manual correction for evaluation
- **Risk:** High computational cost of LLaMA 3 inference **Mitigation:** Process segmented text blocks; batch inference; limit token length; offline execution

4.5.2 Reasoning and Model Risks

- **Risk:** Hallucinated or unsupported verdict predictions **Mitigation:** Constrain prompts to extracted facts and mapped sections; require precedent citation; include non-advisory disclaimers
- **Risk:** Contradiction detection yields false positives **Mitigation:** Combine LLM outputs with heuristic checks; manual review for evaluation cases; threshold-based truth scoring

4.5.3 Development and Evaluation Risks

- **Risk:** Limited availability of labeled ground truth data for Famil courts **Mitigation:** Use mixed evaluation (quantitative + qualitative); expert-reviewed case studies from Law textbooks
- **Risk:** Schedule overruns due to integration complexity **Mitigation:** Modular pipeline design; incremental testing; buffer time in final phase

4.6 Expected Outcomes

Upon completion of the proposed methodology, the project is expected to deliver:

1. A functional Legal Precedent Assistant capable of analyzing Indian Family court case documents
2. An end-to-end NLP pipeline integrating text extraction, evidence analysis, precedent retrieval, and verdict prediction
3. Structured outputs including mapped IPC sections, retrieved precedents, and simplified case summaries
4. Evaluation results demonstrating technical performance and predictive accuracy on real judicial data
5. Identified limitations, biases, and areas for methodological improvement
6. A scalable research foundation for future extensions to additional legal domains and languages

The validation results will support iterative refinement of the system and demonstrate the feasibility of AI-assisted legal research as a decision-support tool for academic and professional use.

Bibliography

[1] Aysha Akther et al. “Compilation, Analysis and Application of a Comprehensive Bangla Corpus KU-Mono”. In: *IEEE Access* 10 (2022), pp. 79999–80014. DOI: 10.1109/ACCESS.2022.3195236.

[2] Lakshmi Boppana et al. “AI-Driven Prediction of Indian Criminal Case Outcomes”. In: *TENCON 2024 - 2024 IEEE Region 10 Conference (TENCON)*. 2024, pp. 1389–1393. DOI: 10.1109/TENCON61640.2024.10902847.

[3] Rodrigo Lobo Canalli et al. “Automated Labelling of Judicial Controversies Before the Brazilian Supreme Court According to the Sustainable Development Goals”. In: *2023 IEEE International Symposium on Technology and Society (ISTAS)*. 2023, pp. 1–7. DOI: 10.1109/ISTAS57930.2023.10305895.

[4] Junyun Cui, Xiaoyu Shen, and Shaochun Wen. “A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges”. In: *IEEE Access* 11 (2023), pp. 102050–102071. DOI: 10.1109/ACCESS.2023.3317083.

[5] Aman Fan, Shaoxi Wang, and Yanchuan Wang. “Chinese Legal Case Similarity Matching Based on Text Importance Extraction”. In: *IEEE Access* 13 (2025), pp. 118745–118758. DOI: 10.1109/ACCESS.2025.3585265.

[6] Congqing He et al. “Knowledge-Enriched Multi-Cross Attention Network for Legal Judgment Prediction”. In: *IEEE Access* 11 (2023), pp. 87571–87582. DOI: 10.1109/ACCESS.2023.3305259.

[7] Ali Shariq Imran et al. “Classifying European Court of Human Rights Cases Using Transformer-Based Techniques”. In: *IEEE Access* 11 (2023), pp. 55664–55676. DOI: 10.1109/ACCESS.2023.3279034.

[8] Muhammad Rafsan Kabir et al. “LegalRAG: A Hybrid RAG System for Multilingual Legal Information Retrieval”. In: *2025 International Joint Conference on Neural Networks (IJCNN)*. 2025, pp. 1–8. DOI: 10.1109/IJCNN64981.2025.11228374.

[9] Zhaolu Kang et al. “JurisCTC: Enhancing Legal Judgment Prediction via Cross-Domain Transfer and Contrastive Learning”. In: *2025 International Joint Conference on Neural Networks (IJCNN)*. 2025, pp. 1–8. DOI: 10.1109/IJCNN64981.2025.11229207.

[10] Yi-Hung Liu and Yi-Fun Lin. “LASG: Streamlining Legal Adjudication with AI-Enabled Summary Generation”. In: *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*. 2024, pp. 2242–2247. DOI: 10.1109/COMPSAC61105.2024.00360.

[11] Yinglong Ma, Peng Zhang, and Jiangang Ma. “An Ontology Driven Knowledge Block Summarization Approach for Chinese Judgment Document Classification”. In: *IEEE Access* 6 (2018), pp. 71327–71338. DOI: 10.1109/ACCESS.2018.2881682.

[12] Mariana Noguti, Eduardo Vellasques, and Luiz S. Oliveira. “A Small Claims Court for the NLP: Judging Legal Text Classification Strategies With Small Datasets”. In: *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2023, pp. 1840–1845. DOI: 10.1109/SMC53992.2023.10394189.

[13] Mariana Y. Noguti, Eduardo Vellasques, and Luiz S. Oliveira. “Legal Document Classification: An Application to Law Area Prediction of Petitions to Public Prosecution Service”. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. 2020, pp. 1–8. DOI: 10.1109/IJCNN48605.2020.9207211.

[14] Weijian Qin and Xudong Luo. “A Legal News Summarisation Model Based on RoBERTa, T5 and Dilated Gated CNN”. In: *2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)*. 2023, pp. 889–897. DOI: 10.1109/ICTAI59109.2023.00134.

[15] Ernesto Quevedo et al. “Legal Natural Language Processing From 2015 to 2022: A Comprehensive Systematic Mapping Study of Advances and Applications”. In: *IEEE Access* 12 (2024), pp. 145286–145317. DOI: 10.1109/ACCESS.2023.3333946.

[16] Marco Siino et al. “Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches”. In: *IEEE Access* 13 (2025), pp. 18253–18276. DOI: 10.1109/ACCESS.2025.3533217.

[17] Dezhao Song et al. “On the Effectiveness of Pre-Trained Language Models for Legal Natural Language Processing: An Empirical Study”. In: *IEEE Access* 10 (2022), pp. 75835–75858. DOI: 10.1109/ACCESS.2022.3190408.

[18] Zhuoyang Song et al. “Parallel Learning for Legal Intelligence: A HANOI Approach Based on Unified Prompting”. In: *IEEE Transactions on Computational Social Systems* 11.2 (2024), pp. 2765–2775. DOI: 10.1109/TCSS.2023.3301400.

[19] Matthias Stürmer. “Keynote - AI for the Public Sector and the Case of Legal NLP”. In: *2023 Ninth International Conference on eDemocracy & eGovernment (ICEDEG)*. 2023, pp. 1–2. DOI: 10.1109/ICEDEG58167.2023.10122084.

[20] Tülay Turan and Ecır Uğur Küçüksılle. “Summarization, Prediction, and Analysis of Turkish Constitutional Court Decisions With Explainable Artificial Intelligence and a Hybrid Natural Language Processing Method”. In: *IEEE Access* 13 (2025), pp. 59766–59779. DOI: 10.1109/ACCESS.2025.3556725.

[21] Santiago Vasquez et al. “Using Language Models for Extracting Legal Decisions from Portuguese Consumer Law Texts”. In: *2025 International Joint Conference on Neural Networks (IJCNN)*. 2025, pp. 1–7. DOI: 10.1109/IJCNN64981.2025.11227838.

[22] Michael Benedict L. Virtucio et al. “Predicting Decisions of the Philippine Supreme Court Using Natural Language Processing and Machine Learning”. In: *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*. Vol. 02. 2018, pp. 130–135. DOI: 10.1109/COMPSAC.2018.10348.

[23] Jie Xiong and Yihui Qiu. “Deep Text Understanding Model for Similar Case Matching”. In: *IEEE Access* 12 (2024), pp. 109877–109885. DOI: 10.1109/ACCESS.2024.3439775.

[24] Linan Yue et al. “A Circumstance-Aware Neural Framework for Explainable Legal Judgment Prediction”. In: *IEEE Transactions on Knowledge and Data Engineering* 36.11 (2024), pp. 5453–5467. DOI: 10.1109/TKDE.2024.3387580.