Vichara: Appellate Judgment Prediction and Explanation for the Indian Judicial System

Pavithra PM Nair, Preethu Rose Anish

TCS Research
pavithra.nair@tcs.com, preethu.rose@tcs.com

Abstract

In jurisdictions like India, where courts face an extensive backlog of cases, artificial intelligence offers transformative potential for legal judgment prediction. A critical subset of this backlog comprises appellate cases, which are formal decisions issued by higher courts reviewing the rulings of lower courts. To this end, we present **Vichara**, a novel framework tailored to the Indian judicial system that predicts and explains appellate judgments. Vichara processes English-language appellate case proceeding documents and decomposes them into *decision points*. Decision points are discrete legal determinations that encapsulate the legal issue, deciding authority, outcome, reasoning, and temporal context. The structured representation isolates the core determinations and their context, enabling accurate predictions and interpretable explanations. Vichara's explanations follow a structured format, inspired by the IRAC (Issue-Rule-Application-Conclusion) framework and adapted for Indian legal reasoning. This enhances interpretability, allowing legal professionals to assess the soundness of predictions efficiently. We evaluate Vichara on two datasets, PredEx and the expert-annotated subset of the Indian Legal Documents Corpus (ILDC_expert), using four large language models (GPT-4o mini, Llama-3.1-8B, Mistral-7B, Qwen2.5-7B). Vichara surpasses existing judgment prediction benchmarks on both datasets, with GPT-4o mini achieving the highest performance (F1: 81.5 on PredEx, 80.3 on ILDC_expert), followed by Llama-3.1-8B. Human evaluation of the generated explanations across *Clarity*, *Linking*, and *Usefulness* metrics highlights GPT-4o mini's superior interpretability.

1 Introduction

Legal Judgment Prediction (LJP) aims to algorithmically forecast judicial outcomes based on case texts. The integration of artificial intelligence (AI) into LJP systems presents a promising avenue for enhancing efficiency and transparency in judicial processes. This is particularly relevant in India, where the judiciary is burdened by a massive backlog of cases. According to data from the National Judicial Data Grid, as published on the Open Government Data Platform of India[1], as of 21 March 2025, there were approximately

[1]https://data.gov.in/resource/court-wise-number-cases-pending-various-courts-information-available-national-judicial-1

45.51 million cases pending before District and Subordinate Courts, 6.25 million cases before High Courts, and 81,598 cases before the Supreme Court. Altogether, more than 51 million cases remain unresolved across various levels of the judiciary. While the majority of pending cases are at the first-instance level, appellate cases, which are cases reviewed by higher courts following lower-court rulings, represent a critical subset. These cases are particularly important because they set legal precedent and ensure consistency across lower courts, making timely resolution essential (Breyer 2006).

AI-driven systems for appellate judgment prediction (AJP) can assist in prioritizing appeals, evaluating legal reasoning, and generating interpretable explanations to support judicial decision-making. However, building effective AJP systems remains challenging due to the complexity and domain-specific nature of legal language and reasoning. Beyond predictive accuracy, AJP systems must generate interpretable explanations to ensure transparency and trustworthiness. Appellate judgments carry high-stakes consequences, and without structured, comprehensive explanations, AI-generated outputs are difficult to validate, contest, or rely upon in practice.

We present Vichara, a framework for appellate judgment prediction and explanation tailored to the Indian judicial context. Named after the Sanskrit word for *deliberation* or *reasoned consideration*, Vichara reflects the analytical rigor inherent in judicial decision-making. The framework comprises six stages: rhetorical role classification, case context construction, decision point extraction, present court ruling generation, judgment prediction, and explanation generation. In the first stage, each sentence in the case proceeding document is classified according to its rhetorical role (Bhattacharya et al. 2019), which refers to the function it serves within the legal discourse, such as stating facts, presenting arguments, citing precedents, or delivering rulings. From the sentences identified as facts, the framework then constructs the case context, capturing the core legal issue, the court deciding the appeal, the parties, and their stances. Next, the framework extracts structured decision points that encode the individual legal issues under consideration, the deciding authority, the outcome, the underlying reasoning, and any temporal information. The decision points where the deciding authority is the present court are retained to generate the present court ruling. Finally, the judgment outcome is de-

rived by comparing this ruling with the appellant's stance, and an explanation is produced using the case context, decision points, present court ruling, and predicted judgment outcome. Vichara currently operates exclusively on English-language case documents.

A central contribution of Vichara is its generation of structured explanations for predicted judgments. Rather than producing free-form text, Vichara outputs explanations in a standardized format comprising sections such as *Facts of the Case*, *Legal Issues Presented*, *Applicable Law and Precedents*, *Reasoning*, and *Conclusion*. This format is inspired by the widely adopted IRAC (Issue–Rule–Application–Conclusion) framework (Metzler 2002) and adapted to reflect the organization of judicial reasoning in Indian court documents. The structured format enhances interpretability, allowing legal professionals to quickly assess the legal soundness of predictions.

We evaluate Vichara on two datasets: PredEx (Nigam et al. 2024a) and the expert-annotated subset of the Indian Legal Documents Corpus (ILDC_expert) (Malik et al. 2021). Our experiments use four large language models (LLMs): GPT-4o mini (OpenAI 2024), Llama-3.1-8B (Grattafiori et al. 2024), Mistral-7B (Jiang et al. 2023), and Qwen2.5-7B (Yang et al. 2024). Each model is assessed on two axes: prediction performance and explanation quality. GPT-4o mini achieves the highest prediction performance (F1: 81.5 on PredEx, 80.3 on ILDC_expert), followed by Llama-3.1-8B (F1: 76.7 on PredEx, 78.5 on ILDC_expert). For explanation quality, we conduct human evaluation using three metrics: *Clarity*, *Linking*, and *Usefulness*, with GPT-4o mini again receiving the highest ratings, followed by Mistral-7B.

Our work makes the following key contributions:

- We propose Vichara, a novel framework for appellate judgment prediction and explanation, centered on decision point extraction from case proceedings.

- We introduce a structured explanation format grounded in legal reasoning conventions, enhancing interpretability and alignment with judicial logic.

- We conduct an empirical evaluation of Vichara across two datasets and four LLMs, demonstrating strong performance in both prediction and explanation quality. Vichara outperforms existing judgment prediction benchmarks on both datasets.

The rest of the paper is organized as follows. Section 2 discusses related work in legal judgment prediction. Section 3 and 4 discuss the datasets used and the task formulation, respectively. Section 5 describes the Vichara framework in detail. Section 6 presents the results and analysis. Section 7 concludes with a discussion of implications and future directions, and Section 8 presents the limitations of Vichara. For the sake of reproducibility, we have made the code accessible via a GitHub link[2].

---

[2]https://github.com/pavithranair/Vichara

2 Related Work

Legal Judgment Prediction (LJP) has advanced considerably in recent years, driven by the growing need to automate legal outcome prediction and ease the burden of increasing caseloads on judicial systems. The field has been shaped by several foundational studies that established core tasks, datasets, and modeling approaches. (Aletras et al. 2016) first demonstrated that court decisions could be predicted from textual case descriptions using traditional feature-based models. (Zhong et al. 2018) extended this by introducing TopJudge, a multi-task framework that jointly predicts legal charges, applicable statutes, and sentence lengths, supported by the large-scale CAIL2018 dataset (Xiao et al. 2018). In the European legal context, (Chalkidis, Androutsopoulos, and Aletras 2019) advanced the field with neural architectures that improved performance over prior baselines on judgment prediction tasks. (Şulea et al. 2017) applied text classification methods to predict case outcomes, law areas, and ruling periods for French Supreme Court cases, while analyzing the impact of temporal context and masking judges' motivations. (Medvedeva, Vols, and Wieling 2020) demonstrate the application of large-scale statistical analysis and machine learning to European Court of Human Rights case texts to predict judicial decisions.

In the Indian legal domain, research on judgment prediction has evolved rapidly, with an increasing focus on both explainability and realism in predictive settings. Foundational datasets such as ILDC (Malik et al. 2021) and PredEx (Nigam et al. 2024a) introduced large, expertly annotated collections of Indian court cases that serve as benchmarks for evaluating models on judgment prediction and explanation tasks. Fact-driven approaches have gained prominence through studies like (Nigam et al. 2024b) and (Nigam and Deroy 2023), which restrict inputs to case facts or limited procedural history, reflecting constraints faced in real-world litigation support systems. More recently, the NyayaAnumana dataset (Nigam et al. 2025) has set a new benchmark in scale and coverage, comprising over 700,000 cases from across the Indian judiciary. Accompanied by INLegalLlama, a domain-specialized generative model, this work demonstrates substantial gains in both predictive accuracy and the coherence of generated explanations.

Despite recent advances, most existing approaches either focus primarily on prediction, where performance still lags behind human experts, or attempt to generate explanations that fall into two categories. Some are *extractive* (Prasad, Boughanem, and Dkaki 2023; Yamada et al. 2024), simply highlighting relevant case facts, while others are *abstractive* (Nigam et al. 2024b, 2025), producing free-form summaries. In both cases, explanations are usually unstructured and offer limited insight into how facts and legal arguments lead to the court's decisions. In contrast, Vichara leverages LLMs to deliver strong predictive performance while producing structured, interpretable explanations that allow legal professionals to trace reasoning from evidence to outcome. By explicitly modeling the legal reasoning process, Vichara supports both practitioners and the wider public in navigating complex judicial environments, while paving the way for more transparent and accountable AI-driven legal tools.
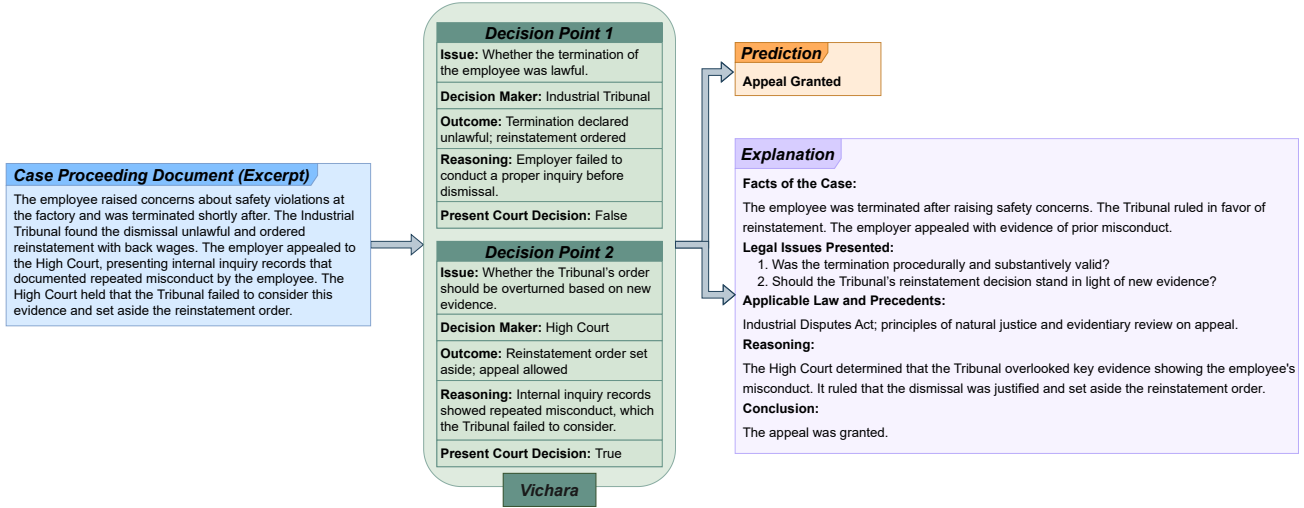
Figure 1: Vichara's appellate judgment prediction and explanation for a sample case excerpt.

3 Dataset

We evaluate our framework using two benchmark datasets from the Indian legal domain:

- PredEx: PredEx (Nigam et al. 2024a) is the largest publicly available source for joint judgment prediction and explanation tasks in the Indian context. We use the test split of this dataset, which contains 3,044 appellate cases from the Supreme Court of India and various high courts, annotated with binary outcome labels (Appeal Granted or Dismissed) and expert-annotated explanations.

- ILDC_expert: ILDC_expert is the expert-annotated subset of the Indian Legal Documents Corpus (Malik et al. 2021). It consists of 56 Supreme Court of India appellate cases, each annotated with binary outcome labels (Appeal Granted or Dismissed) and corresponding expert provided explanations.

All case proceedings and explanations in both datasets are written in the English language.

4 Vichara Task Formulation

Vichara consists of a unified pipeline with two interconnected components:

4.1 Judgment Prediction Component

Given a segment from a Supreme Court of India or high court appellate case proceeding document, Vichara predicts whether the court ruled in favor of the appellant. The output is a binary label: $\{1, 0\}$, where 1 denotes that the appeal was granted, and 0 indicates dismissal. While real cases can involve mixed outcomes, Vichara abstracts these into a single binary decision to focus on the core outcome.

4.2 Explanation Generation Component

Alongside predicting the outcome, Vichara produces a structured rationale explaining the prediction. The explanation captures the key facts, applicable laws, and legal reasoning, following an IRAC-inspired structure adapted to Indian legal discourse.

Figure 1 illustrates Vichara's task formulation.

5 Methodology

Vichara comprises six sequential stages: rhetorical role classification, case context construction, decision point extraction, present court ruling generation, judgment prediction, and explanation generation.

The process begins by classifying each sentence in the input case proceeding document according to its rhetorical role (Bhattacharya et al. 2019), identifying whether it states facts, presents arguments, cites precedents, or issues rulings. All subsequent stages in the pipeline operate purely through prompting without any fine-tuning of the underlying LLMs. The case context is extracted exclusively from sentences labeled as facts, capturing the key legal issue, the court deciding the appeal, the parties, and their respective stances. Vichara extracts decision points that summarize the court's legal determinations, filtering them to retain only those attributable to the present court. These filtered decision points are then used to generate the present court's ruling, ensuring it reflects solely the outcome of the appeal under consideration. The predicted judgment is derived by comparing this ruling with the appellant's stance. Finally, Vichara produces an IRAC-style (Metzler 2002) explanation grounded in the extracted facts, laws, and reasoning. The overall architecture is illustrated in Figure 2, with detailed prompt templates provided in Table 5 in the appendix. Next, we explain each of the steps in the methodology.

5.1 Rhetorical Role Classification

Rhetorical role labelling of sentences in a legal document refers to identifying the semantic function each sentence
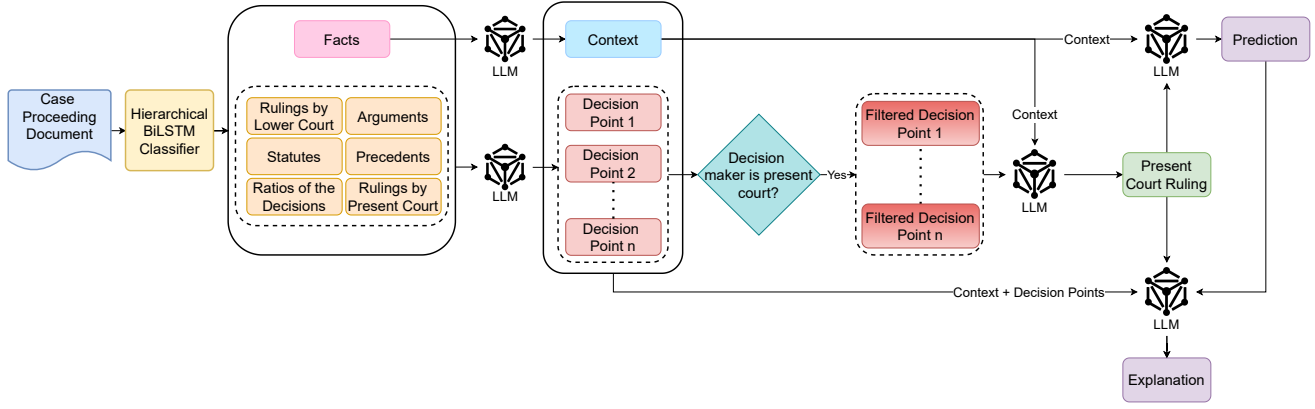
Figure 2: Vichara flow diagram

serves, such as stating the facts of the case, presenting arguments of the parties, or delivering the court's judgment (Bhattacharya et al. 2019).

The input case proceeding document is segmented into sentences using the SAT-12L (Segment Any Text) pretrained model, fine-tuned via Low-Rank Adaptation (LoRA) for accurate sentence boundary detection in legal texts (Frohmann et al. 2024). (Bhattacharya et al. 2019) define seven categories of rhetorical roles, namely 'Facts', 'Ruling by Lower Court', 'Argument', 'Statute', 'Precedent', 'Ratio of the Decision', and 'Ruling by Present Court'. Each sentence in our dataset is classified into one of these seven categories by a hierarchical BiLSTM-CRF model trained to assign rhetorical roles, following the schema introduced by (Bhattacharya et al. 2019).

5.2 Case Context Construction

To generate case-level context, an LLM is applied to the subset of sentences labeled as 'Facts' during rhetorical role classification (Section 5.1). The LLM extracts six key fields:

- Appellants: The individuals or entities filing the appeal.
- Respondents: The opposing parties in the case.
- Issue: The main legal or factual question under consideration.
- Appellant's Stance: The relief sought or position taken by the appellant.
- Respondent's Stance: The response or objections raised by the respondent.
- Present Court: The court deciding the appeal.

This structured context summarizes the parties, their positions, the present court and the core dispute in the case.

5.3 Decision Point Extraction

To extract decision points, the case proceeding document is divided into manageable segments for prompting. When available, explicit bullet-point lists in the document are used, which often enumerate key issues or rulings. If bullet-point structure is absent, the document is divided into 1000-token

segments. This chunk size is chosen based on empirical observations that longer segments tend to degrade the quality of decision point extraction.

These segments are passed on to an LLM to extract decision points. Decision points are discrete legal determinations made at different stages throughout the case. Each decision point is represented as a structured unit consisting of the following six fields:

- Issue: The specific legal or factual question under consideration.
- Decision Maker: The authority that issued the decision.
- Outcome: The court's resolution of the issue, such as whether a claim was upheld, dismissed, or partially granted.
- Reasoning (optional): The rationale provided by the decision maker for the determination on that issue.
- Time (optional): Any explicit reference to when the decision was made, especially relevant in multi-phase cases or appeals.
- Present Court Decision: A Boolean flag indicating whether the decision maker is the present court deciding the appeal.

5.4 Present Court Ruling Generation

To ensure that the LLM focuses solely on the reasoning of the present court, the extracted decision points are filtered to retain only those where the deciding authority is the present court. This prevents the framework from incorporating rulings made by lower courts, which may appear earlier in the case document and introduce confusion. We also extract the final sentence of the case document with the rhetorical role 'Ruling by Present Court', referred to as the *final statement*.

Using the structured case context, the filtered decision points, and the final statement, an LLM is prompted to generate a detailed summary of the present court's ruling. The output includes specific reliefs granted or denied, any orders or directions issued, the court's reasoning and key considerations, and relevant timelines or compliance instructions. This step captures only the outcome of the current appeal,

ignoring lower court decisions or earlier findings unless explicitly referenced by the present court.

5.5 Judgment Prediction

To determine the binary outcome of the appeal, either granted or dismissed, an LLM compares the present court ruling (Section 5.4) with the appellant's stance extracted during context construction (Section 5.2). If the ruling fully or partially grants the relief sought by the appellant, the framework outputs a prediction of 1 (Appeal Granted). Otherwise, if the relief is entirely denied, the prediction is 0 (Appeal Dismissed).

5.6 Structured Explanation Generation

The final stage of the pipeline involves generating a structured explanation for the predicted outcome of the appeal. An LLM is used to synthesize the explanation. The inputs to the LLM include the case context (Section 5.2), extracted decision points (Section 5.3), present court ruling (Section 5.4), and the predicted outcome (Section 5.5).

The LLM is prompted to organize the explanation into five predefined sections:

- Facts of the Case: A brief overview of the dispute, parties involved, and procedural background.
- Legal Issue(s) Presented: The central legal questions that the court addressed.
- Applicable Law and Precedents: Statutes, constitutional provisions, or prior case law cited or relied upon.
- Analysis / Reasoning: The logical application of the law to the facts, incorporating relevant decision points to justify the predicted outcome.
- Predicted Conclusion: A summary conclusion reflecting the model's judgment (e.g., "Appeal Granted" or "Appeal Dismissed").

This structured format ensures that explanations are interpretable and legally grounded.

6 Results and Analysis

In this section, we present the evaluation of Vichara on both judgment prediction and explanation quality, as well as an ablation study to assess the necessity of each stage of the Vichara framework. We evaluate the framework using four LLMs, GPT-4o mini (OpenAI 2024), Llama-3.1-8B[3], Mistral-7B[4] and Qwen2.5-7B[5]. All experiments were repeated across 5 independent random seeds to account for the non-deterministic nature of LLMs. Reported results for judgment prediction and automated evaluation of explanations correspond to the mean performance across seeds, with standard deviations included to indicate robustness.

6.1 Results for Judgment Prediction

We evaluate the binary outcome of judgment prediction using standard classification metrics: accuracy, precision, recall, and F1-score. To ensure balanced evaluation across

---

[3]https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct
[4]https://huggingface.co/mistralai/Mistral-7B-Instruct-v0.3
[5]https://huggingface.co/Qwen/Qwen2.5-7B-Instruct

both classes (Appeal Granted and Dismissed), we report macro-averaged scores. Results are presented in Table 1.

Among the LLMs evaluated in our experiments within the Vichara framework, GPT-4o mini achieves the highest scores on both datasets, followed by Llama-3.1-8B, Qwen2.5-7B, and Mistral-7B. Compared to the state-of-the-art baseline, INLegalLlama (Nigam et al. 2025), all LLMs evaluated in our experiments surpass its performance on the ILDC_expert dataset, while on PredEx, GPT-4o mini and Llama-3.1-8B outperform INLegalLlama.

6.2 Results for Explanation

To evaluate explanation quality, we rely on human assessments conducted by experts from the legal domain. We use three metrics: Clarity, Linking, and Usefulness, which together capture essential aspects of interpretability, legal alignment, and practical value. These metrics are particularly important in our setting, as the structured explanations produced by our framework differ significantly in form and granularity from the gold standard references in existing datasets (Nigam et al. 2024a; Malik et al. 2021), making reference-based automatic metrics less reliable.

Following (Nigam et al. 2024b), we adopt their definitions of Clarity and Linking, and introduce a new metric, Usefulness, as defined below:

- Clarity: Measures how well-structured, readable, and logically coherent the explanation is.
- Linking: Assesses the degree to which the explanation offers a justifiable connection between the facts and the predicted outcome.
- Usefulness: Evaluates how informative and practically helpful the explanation is for a legal professional seeking to understand the court's reasoning and apply it in real-world legal work.

For the evaluation, we recruited three legal experts, all practicing advocates with 7-8 years of professional experience. Each expert was provided with the case proceeding document, the predicted judgment outcome, and the corresponding explanation generated by the model. They rated 25 explanations per model across four LLMs, resulting in 100 annotated examples. Each explanation was rated independently on a 5-point Likert scale (1 = very poor, 5 = excellent) for all three criteria. Detailed guidelines provided to the evaluators are included in Appendix A.

Evaluating explanation quality across multiple random seeds was not performed due to the substantial human effort that would be required for assessing outputs from five separate runs. The evaluation was conducted on the explanation outputs of a single run. Inter-annotator agreement, measured using Fleiss' Kappa (Fleiss 1971), was substantial (Clarity: 0.66, Linking: 0.70, Usefulness: 0.63). Table 2 reports the average scores per LLM per metric. Among the models, GPT-4o mini achieved the highest scores across all three dimensions, followed by Mistral-7B, which also demonstrated strong performance in Clarity and Linking.

While our primary focus is on human evaluation due to its greater reliability, we also report standard reference-based automatic metrics (ROUGE-1, ROUGE-2, ROUGE-

Model	PredEx				ILDC_expert			
	Accuracy	Macro Precision	Macro Recall	Macro F1	Accuracy	Macro Precision	Macro Recall	Macro F1
GPT-4o mini	**81.62 ± 0.41**	**81.57 ± 0.43**	**81.45 ± 0.40**	**81.50 ± 0.42**	**80.36 ± 0.37**	**81.25 ± 0.34**	**80.65 ± 0.36**	**80.30 ± 0.35**
Llama-3.1-8B	76.93 ± 0.49	76.75 ± 0.51	76.60 ± 0.48	76.66 ± 0.50	78.57 ± 0.42	78.71 ± 0.40	78.42 ± 0.39	78.46 ± 0.41
Mistral-7B	69.41 ± 0.55	69.21 ± 0.53	69.00 ± 0.52	69.04 ± 0.54	75.00 ± 0.46	75.48 ± 0.44	75.22 ± 0.43	74.97 ± 0.45
Qwen2.5-7B	71.73 ± 0.50	70.06 ± 0.52	71.97 ± 0.48	72.10 ± 0.51	76.79 ± 0.44	78.48 ± 0.42	77.20 ± 0.45	76.60 ± 0.43
INLegalLlama (Nigam et al. 2025)	76.05	76.23	76.05	76.01	72.23	73.01	72.23	71.98

Table 1: Judgment prediction performance comparison of models on PredEx and ILDC_expert datasets, reported as mean ± standard deviation across 5 random seeds. The LLMs listed (GPT-4o mini, Llama-3.1-8B, Mistral-7B, Qwen2.5-7B) are used within the Vichara framework. The best-performing model for each dataset is highlighted in bold. We have also included the state-of-the-art baseline, INLegalLlama (Nigam et al. 2025), for comparison.

	Clarity	Linking	Usefulness
GPT-4o mini	**4.57 ± 0.22**	**4.96 ± 0.21**	**4.37 ± 0.18**
Llama-3.1-8B	3.43 ± 0.30	3.77 ± 0.25	3.29 ± 0.28
Mistral-7B	4.11 ± 0.27	4.44 ± 0.22	3.85 ± 0.24
Qwen2.5-7B	3.44 ± 0.26	3.29 ± 0.31	3.33 ± 0.29

Table 2: Expert evaluation results for the explanation task. Values are reported as mean ± standard deviation across three evaluators. Bold values indicate the highest score for each metric.

L, BLEU, METEOR, BERTScore, and BLANC) for completeness. We use the ground truth explanations available in the PredEx and ILDC_expert datasets as reference texts. The results are provided in Table 3. Qwen2.5-7B records the highest scores on most automatic metrics, including ROUGE-1, ROUGE-L, BLEU, and METEOR. Although automatic metrics favor Qwen2.5-7B, human evaluation results favor GPT-4o-mini. This contrast highlights the limitations of automatic metrics in fully capturing the explanation quality.

6.3   Ablation Study

To assess the necessity and contribution of each component within the Vichara framework, we conduct a systematic ablation study across four stages: (1) rhetorical role classification, (2) case context construction, (3) decision point extraction, and (4) present court ruling generation.

Experimental Setup   We create a series of ablated variants of the full pipeline by selectively removing individual stages while holding all other configurations constant. When a stage is removed, subsequent prompts are minimally adjusted to maintain coherence and ensure functionality without referencing missing inputs. All evaluations are performed using the GPT-4o mini model on the ILDC_expert dataset. We report accuracy and macro-F1 for the predictive component (judgment prediction), and Clarity, Linking, and Usefulness scores for the human evaluation of explanations. Judgment prediction results are reported as the mean of five independent runs with different random seeds, along with the standard deviation reflecting variation across seeds.

The explanation evaluation was conducted on 15 explanation outputs from a single run, as assessing multiple runs would require substantial additional human effort. It was carried out by the same three legal experts who evaluated the explanations in Section 6.2. For explanation quality, we report the mean score across the three evaluators, with the standard deviation representing variation among evaluators.

Ablation Configurations   The ablation study consisted of five configurations:

1. **Full Vichara (All Stages)**: The complete pipeline as described in Section 5.
2. **Without Rhetorical Role Classification**: Sentence-level segmentation and rhetorical role labeling are omitted. The entire case proceeding document is passed as input for Case Context Construction (Section 5.2). During Present Court Ruling Generation (Section 5.4), the final statement—the last sentence originally labeled as 'Ruling by Present Court'—is excluded from the input prompt.
3. **Without Case Context Construction**: Case context is not generated and therefore not provided as input to the stages of Present Court Ruling Generation (Section 5.4), Judgment Prediction (Section 5.5), or Structured Explanation Generation (Section 5.6).
4. **Without Decision Point Extraction**: The model directly generates the present court ruling from the input case proceeding document, the case context, and the final statement, without relying on decision points. Decision points are also not provided as input during Structured Explanation Generation (Section 5.6).
5. **Without Present Court Ruling Generation**: The present court ruling is not generated, and therefore not provided as input to the stages of Judgment Prediction (Section 5.5), or Structured Explanation Generation (Section 5.6). Instead, the framework predicts the final judgment directly by comparing the case context with the extracted decision points.

Results and Discussion   The results of the ablation study are presented in Table 4. Each stage of the Vichara pipeline demonstrably contributes to both predictive performance and explanation quality. The full configuration achieves the

Models	Lexical Based Evaluation (%)					Semantic Evaluation (%)	
	Rouge-1	Rouge-2	Rouge-L	BLEU	METEOR	BERTScore(F1)	BLANC
PredEx							
GPT-4o mini	38.60 ± 0.42	14.52 ± 0.33	19.35 ± 0.29	3.24 ± 0.15	21.87 ± 0.31	82.79 ± 0.38	13.48 ± 0.21
Llama-3.1-8B	32.14 ± 0.40	14.10 ± 0.30	17.49 ± 0.25	3.06 ± 0.12	19.06 ± 0.28	81.55 ± 0.36	12.44 ± 0.20
Mistral-7B	35.75 ± 0.41	15.53 ± 0.32	18.80 ± 0.27	2.74 ± 0.13	20.61 ± 0.30	**82.90 ± 0.37**	13.41 ± 0.22
Qwen2.5-7B	**38.70 ± 0.43**	**15.80 ± 0.34**	**19.49 ± 0.30**	**7.06 ± 0.25**	**23.77 ± 0.35**	81.66 ± 0.38	**13.59 ± 0.23**
ILDC_expert							
GPT-4o mini	33.90 ± 0.38	15.18 ± 0.31	17.02 ± 0.27	1.25 ± 0.10	14.60 ± 0.25	82.69 ± 0.35	12.65 ± 0.18
LLaMa-3.1-8B	31.17 ± 0.36	15.79 ± 0.32	16.43 ± 0.26	1.91 ± 0.11	13.91 ± 0.24	82.18 ± 0.33	12.02 ± 0.17
Mistral-7B	32.12 ± 0.37	**16.45 ± 0.33**	16.48 ± 0.27	1.01 ± 0.09	14.13 ± 0.26	**82.77 ± 0.34**	**13.01 ± 0.19**
Qwen2.5-7B	**36.13 ± 0.39**	15.67 ± 0.32	**17.41 ± 0.28**	**4.16 ± 0.18**	**17.45 ± 0.30**	82.23 ± 0.34	12.17 ± 0.18

Table 3: Explanation performance comparison of various models across automatic evaluation metrics. Results are reported as mean ± standard deviation. The highest scores are in bold.

Configuration	Accuracy	Macro F1	Clarity	Linking	Usefulness
Full Vichara (All Stages)	**80.36 ± 0.37**	**81.25 ± 0.34**	**4.57 ± 0.22**	**4.96 ± 0.21**	**4.37 ± 0.18**
Without Rhetorical Role Classification	79.12 ± 0.45	78.85 ± 0.41	4.32 ± 0.28	4.66 ± 0.24	4.21 ± 0.25
Without Case Context Construction	77.84 ± 0.52	77.40 ± 0.47	4.21 ± 0.31	4.55 ± 0.26	4.18 ± 0.20
Without Decision Point Extraction	71.65 ± 0.61	70.92 ± 0.57	3.80 ± 0.35	4.10 ± 0.28	3.95 ± 0.33
Without Present Court Ruling Generation	73.22 ± 0.54	72.86 ± 0.49	3.95 ± 0.30	4.20 ± 0.27	4.00 ± 0.29

Table 4: Ablation study results on the *ILDC_expert* dataset. Each configuration removes one stage from the Vichara pipeline. Reported values are mean ± standard deviation. Bold values indicate the highest performance for each metric.

highest prediction performance, as well as top human evaluation scores for explanation quality. Removing individual components results in consistent declines across all metrics, confirming that the multi-stage architecture is essential for producing legally coherent and interpretable outcomes. Eliminating Decision Point Extraction (Configuration 4) produces the largest drop in predictive performance, reducing macro-F1 by nearly 11 points. Eliminating Rhetorical Role Classification (Configuration 2) produces the smallest drop in predictive performance and explanation quality.

7 Conclusions

We introduced Vichara, a framework for appellate judgment prediction and explanation tailored to the Indian judicial system. By representing legal documents as sequences of decision points, Vichara enables accurate outcome prediction and interpretable explanation generation. Our structured explanation format, grounded in legal reasoning conventions, supports transparency and practical usability. Through experiments on two Indian legal datasets, PredEx and ILDC_expert, we demonstrated that Vichara not only achieves strong performance across both proprietary and open-weight LLMs, but also surpasses existing benchmark results on these datasets for judgment prediction. Notably, smaller models such as Llama-3.1-8B, Qwen2.5-7B, and Mistral-7B, when used within Vichara, achieved results comparable to larger models like GPT-4o mini, offering a viable path for resource-constrained deployments. Human evaluation validated the quality of the generated explanations, showing high scores across Clarity, Linkage, and Use-

fulness. Vichara contributes a step forward in the development of explainable AI systems for judicial applications.

Future work will explore methods to reduce computational overhead and adapt the framework to other case types and legal jurisdictions.

8 Limitations

While Vichara demonstrates strong performance in both judgment prediction and structured explanation generation, several limitations remain.

Our evaluation is restricted to the Indian judiciary, with a primary focus on appellate-level cases from the Supreme Court and selected High Courts. Although the core methodology may be applicable to other legal systems, transferring the approach to jurisdictions with different procedural structures, legal doctrines, or language conventions would require substantial adaptation and validation.

Vichara relies on prompt-based querying of LLMs at multiple stages of the pipeline. This approach introduces variability due to the non-deterministic nature of model outputs and the sensitivity of results to prompt phrasing.

The human evaluation of explanation quality was conducted on a limited sample consisting of 25 explanations per language model, reviewed by three advocates. While this provides useful qualitative insights, broader evaluation involving a more diverse pool of legal professionals and a wider range of case types is necessary to establish generalizability and practical relevance.

Finally, the multi-stage architecture of Vichara, which includes multiple LLM calls, may impose computational and

deployment challenges. These constraints could be particularly limiting in environments with restricted resources or strict latency requirements.

Future research will focus on investigating methods to reduce computational overhead and expanding evaluation to include additional legal domains and jurisdictions. This includes exploring strategies such as prompt optimization and model distillation to improve scalability while maintaining performance in real-world legal applications. To broaden evaluation, we plan to assess the framework on a wider range of case types, court levels, and jurisdiction-specific datasets, enabling a more comprehensive understanding of its effectiveness across diverse legal contexts.

Ethical Statement

In conducting this research, we adhered to ethical standards in both data usage and human evaluation. All legal case documents used in our experiments are publicly available and drawn from established open-access legal datasets. No private, confidential, or sensitive information was accessed or utilized at any stage. For the human evaluation of model-generated explanations, we engaged three practicing advocates with formal legal training and courtroom experience. Their participation was entirely voluntary. The evaluation was conducted with informed consent, and participants were briefed on the purpose and scope of the study.

References

Aletras, N.; Tsarapatsanis, D.; Preoţiuc-Pietro, D.; and Lampos, V. 2016. Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ computer science*, 2: e93.

Bhattacharya, P.; Paul, S.; Ghosh, K.; Ghosh, S.; and Wyner, A. 2019. Identification of rhetorical roles of sentences in indian legal judgments. In *Legal knowledge and information systems*, 3–12. IOS Press.

Breyer, S. G. 2006. Reflections on the Role of Appellate Courts: A View from the Supreme Court. *J. App. Prac. & Process*, 8: 91.

Chalkidis, I.; Androutsopoulos, I.; and Aletras, N. 2019. Neural Legal Judgment Prediction in English. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4317–4323.

Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5): 378.

Frohmann, M.; Sterner, I.; Vulić, I.; Minixhofer, B.; and Schedl, M. 2024. Segment Any Text: A Universal Approach for Robust, Efficient and Adaptable Sentence Segmentation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 11908–11941.

Grattafiori, A.; Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Vaughan, A.; et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Jiang, A. Q.; Sablayrolles, A.; Mensch, A.; Bamford, C.; Chaplot, D. S.; de las Casas, D.; Bressand, F.; Lengyel, G.; Lample, G.; Saulnier, L.; Lavaud, L. R.; Lachaux, M.-A.;

Stock, P.; Scao, T. L.; Lavril, T.; Wang, T.; Lacroix, T.; and Sayed, W. E. 2023. Mistral 7B. arXiv:2310.06825.

Malik, V.; Sanjay, R.; Nigam, S. K.; Ghosh, K.; Guha, S. K.; Bhattacharya, A.; and Modi, A. 2021. ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 4046–4062.

Medvedeva, M.; Vols, M.; and Wieling, M. 2020. Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, 28(2): 237–266.

Metzler, J. 2002. The importance of IRAC and legal writing. *U. Det. Mercy L. Rev.*, 80: 501.

Nigam, S.; Sharma, A.; Khanna, D.; Shallum, N.; Ghosh, K.; and Bhattacharya, A. 2024a. Legal Judgment Reimagined: PredEx and the Rise of Intelligent AI Interpretation in Indian Courts. In *Findings of the Association for Computational Linguistics ACL 2024*, 4296–4315.

Nigam, S. K.; Balaramamahanthi, D. P.; Mishra, S.; Shallum, N.; Ghosh, K.; and Bhattacharya, A. 2025. NYAYAANUMANA and INLEGALLLAMA: The Largest Indian Legal Judgment Prediction Dataset and Specialized Language Model for Enhanced Decision Analysis. In *Proceedings of the 31st International Conference on Computational Linguistics*, 11135–11160.

Nigam, S. K.; and Deroy, A. 2023. Fact-based court judgment prediction. In *Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation*, 78–82.

Nigam, S. K.; Deroy, A.; Maity, S.; and Bhattacharya, A. 2024b. Rethinking Legal Judgement Prediction in a Realistic Scenario in the Era of Large Language Models. In *Proceedings of the Natural Legal Language Processing Workshop 2024*, 61–80.

OpenAI. 2024. GPT-4o mini. https://platform.openai.com/docs/models/gpt-4o-mini.

Phoenix, J.; and Taylor, M. 2024. *Prompt engineering for generative AI*. " O'Reilly Media, Inc.".

Prasad, N.; Boughanem, M.; and Dkaki, T. 2023. A Hierarchical Neural Framework for Classification and its Explanation in Large Unstructured Legal Documents. In *The 3rd International Workshop on Mining and Learning in the Legal Domain*.

Şulea, O.-M.; Zampieri, M.; Vela, M.; and van Genabith, J. 2017. Predicting the Law Area and Decisions of French Supreme Court Cases. In Mitkov, R.; and Angelova, G., eds., *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, 716–722. Varna, Bulgaria: INCOMA Ltd.

Xiao, C.; Zhong, H.; Guo, Z.; Tu, C.; Liu, Z.; Sun, M.; Feng, Y.; Han, X.; Hu, Z.; Wang, H.; et al. 2018. Cail2018: A large-scale legal dataset for judgment prediction. *arXiv preprint arXiv:1807.02478*.

Yamada, H.; Tokunaga, T.; Ohara, R.; Tokutsu, A.; Takeshita, K.; and Sumida, M. 2024. Japanese tort-case

dataset for rationale-supported legal judgment prediction. *Artificial Intelligence and Law*, 1–25.

Yang, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Li, C.; Liu, D.; Huang, F.; Wei, H.; et al. 2024. Qwen2. 5 technical report. *arXiv preprint*.

Zhong, H.; Guo, Z.; Tu, C.; Xiao, C.; Liu, Z.; and Sun, M. 2018. Legal judgment prediction via topological learning. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, 3540–3549.

A Human Evaluation Guidelines

The following instructions were provided to human evaluators for assessing the quality of the generated explanations. Each explanation was evaluated independently along three dimensions using a 5-point Likert scale.

A.1 Clarity

Measures how clearly the explanation is written and how well it communicates the rationale behind the decision.

1: The explanation is confusing or incoherent. The rationale is difficult to follow.
2: Some parts are understandable, but the reasoning is vague or underdeveloped.
3: The explanation is moderately clear but may lack smooth flow or sufficient detail.
4: The rationale is clearly presented and easy to follow. Terminology is appropriate.
5: The explanation is very well-written, logically structured, and highly understandable.

A.2 Linking

Measures how well the explanation connects the facts and legal reasoning to the final predicted outcome (e.g., Appeal Granted or Dismissed).

1: The explanation does not connect to the outcome at all or is highly inconsistent.
2: Weak or unclear linkage between the reasoning and the final decision.
3: Some linkage exists, but gaps or ambiguities are present.
4: Clear and logical connection to the judgment, with minimal gaps.
5: Strong and coherent justification that clearly supports the predicted decision.

A.3 Usefulness

Measures how useful the explanation would be for a human reader, particularly a legal practitioner, trying to understand the reasoning behind the AI's decision.

1: The explanation is not helpful or usable in any practical sense.
2: Limited usefulness; lacks essential detail or context.
3: Somewhat helpful, but may miss key points or feel generic.

4: Offers clear value in understanding the reasoning and potential legal implications.
5: Highly informative and usable; effectively mirrors real-world legal reasoning.

B Prompts for Each Processing Step

Table 5 contains the prompts used for each stage of the processing pipeline. The prompts were developed through an iterative prompt engineering process, inspired by the principles on prompt design (Phoenix and Taylor 2024). Figure 3 illustrates the outputs generated at each processing step using GPT-4o mini, for an example court case.

Step	Prompt
Case Context Construction	You are a legal assistant helping summarize appeal case details. Given the following facts from an appeal case document, extract the following information about the current appeal only: 1. Appellants – the persons or entities filing the current appeal. If their name is not mentioned, write what they are referred to as in the text (e.g., "the petitioner", "the appellant"). 2. Respondents – the persons or entities against whom the current appeal is filed. If their name is not mentioned, write what they are referred to as in the text (e.g., "respondent 1", "the respondent-Management"). 3. Issue – the main legal or factual issue being disputed in the current appeal. 4. Appellant's Stance (in the current appeal) – clearly state what the appellant is arguing for or seeking in the present appeal. 5. Respondent's Stance (in the current appeal) – clearly state what the respondent is arguing for or seeking in the present appeal. 6. Present Court – the court deciding the present appeal (e.g., Supreme Court of India, High Court of Bombay). ### Important instructions: - Do NOT assume the appellant is the party introduced first. Carefully check who has filed the current appeal. - Do NOT summarize or include opinions or findings of lower courts unless those are being specifically challenged in this appeal. - Focus on the actual parties to the legal dispute. - Do NOT invent names or facts. If something is not mentioned, leave it as an empty string. - Use only the output format specified below. ### Output Format: {{ "appellants": "\<name or description of appellant\>", "respondents": "\<name or description of respondent\>", "issue": "\<brief summary of the legal/factual issue\>", "appellant_stance": "\<stance of the appellant\>", "respondent_stance": "\<stance of the respondent\>", "present_court": "\<name of the court currently deciding the appeal\>" }} ### Facts: \<{facts}\>

| Decision Point Extraction | You are a legal assistant tasked with extracting **all decision points** from an excerpt of a court case proceeding. Decision points are discrete legal determinations made at different stages throughout the case.

Given the following text and the present court, extract **all identifiable decision points** and output them in strict JSON format as a list of objects.

Each decision point object should include:
- "issue": the legal issue or question being addressed
- "decision_maker": the court or authority that made the decision (e.g., Supreme Court, Trial Court, High Court)
- "outcome": the result or resolution of the issue
- "time": (optional) the date or timeframe of the decision if mentioned
- "reasoning": (optional) summary of the Court's reasoning, including references to statutes, arguments, facts, or precedents
- "present_court_decision": true if the "decision_maker" is the same as the present court provided, otherwise false

Important instructions:
- Do not include any extra text or explanation.
- Do not assume or hallucinate decision makers.
- Do NOT include triple backticks ("'").

Output Format:

[{
"issue": "<string>",
"decision_maker": "<string>",
"outcome": "<string>",
"time": "<string or null>",
"reasoning": "<string or null>",
"present_court_decision": <true or false>
}]

Input:
Present Court: "<{present_court}>"
Text: <{group_text}> |

Present Court Ruling Generation	Your goal is to identify the **final ruling of the present court** in this appeal — that is, what the present court ultimately decided and ordered.
	### Case Context:
	Below is the context of the case, which clearly identifies: - Who the **appellant** is (the party who filed the appeal)
	- Who the **respondent** is (the party defending against the appeal)
	- What the **main issue** of the appeal is.
	- Appellant's Stance (in the current appeal) – What the appellant is arguing for or seeking **in the present appeal**
	- Respondent's Stance (in the current appeal) – What the respondent is arguing for or seeking **in the present appeal**
	Please pay close attention to this information, it overrides any assumptions you might make from the decision points. If the appellants, respondents or the issue of the appeal are not mentioned in the context, ONLY then infer these from the decision points.
	<{context}>
	—
	### Decision Points:
	Decision points are key legal determinations made at different stages throughout the case. The below decision points collectively summarize the key determinations the present court made throughout the case.
	<{present_court_points}>
	—
	### Final Statements from the Present Court:
	This section contains the last official statements or conclusions made by the present court in this appeal. These are the most authoritative and conclusive indication of the court's final position and must be treated as such.
	<{final_statement}>
	—
	### Your Task:
	1. Provide a comprehensive explanation of the final ruling, including:
	- Specific reliefs granted or denied
	- Any orders or directions issued
	- The court's reasoning and key factors considered
	- Relevant timelines or compliance expectations
	2. Focus on the decision points to determine what the present court considered during this appeal.

Present Court Ruling Generation	3. Use the **Final Statements from the Present Court** to determine what the court ultimately ruled. — ### Important: - Do not confuse appellants and respondents. Use the parties as stated in the case context. - Respond clearly and concisely. - Do NOT state whether the appeal was granted or dismissed. Your response should only describe the final ruling of the present court. — ### Output Format: Final Ruling: <Provide a detailed and comprehensive explanation of the present court's final decision in the appeal.>
Judgment Prediction	You are a legal assistant helping to analyze the outcome of an appeal. Your task is to determine whether the **present court's final ruling** aligns with what the **appellant** was seeking in this appeal. ### Case Context: The context below includes: - Appellants: The **appellant** (party who filed the appeal) - Respondents: The **respondent** (party defending the appeal) - Issue: The **main issue** - Appellant's Stance: What the **appellant is seeking** in the current appeal - Respondent's Stance: What the **respondent is seeking** in the current appeal - Present Court: The court deciding the present appeal. <{context}> — ### Final Court Ruling: <{court_ruling}> — ### Your Task: - If the **court fully or partially granted what the appellant was seeking**, output: 'Prediction: 1' - If the **court did not grant what the appellant was seeking**, output: 'Prediction: 0' — ### Output Format: Prediction: <0 or 1>

Structured Explanation Generation	You are a legal assistant. Your task is to generate a structured legal explanation for the court's predicted decision in this appeal case. You are given: - The **case context** including appellant, respondent, issue, stances, and the court deciding the appeal - The **final court ruling** from the present court - A set of **decision points** extracted from the case. A **decision point** refers to a key moment in the case where a specific issue was considered, a responsible authority or decision-maker evaluated it, and a determination or outcome was reached. - The **predicted outcome** of the appeal Generate a structured explanation with the following sections: — Facts of the Case: [A brief summary of the background, parties involved, and what led to the appeal.] Legal Issue(s) Presented: [The legal question(s) the court had to decide.] Applicable Law and Precedents: [Key statutes, constitutional provisions, or case law relied on.] Analysis / Reasoning: [A logical application of law to facts, showing why the court ruled the way it did.] Predicted Conclusion: [Restate the predicted outcome using legal terminology (e.g., 'Appeal Allowed' or 'Dismissed').] — ### Case Context: <{context}> ### Final Court Ruling: <{court_ruling}> ### Decision Points: <{decision_points_text}> ### Predicted Outcome: <{predicted_outcome}>

Table 5: Prompts for each processing step

Processing Steps

Input Case Proceeding Document

Case Proceeding Document

1. Appeal No. 1878 of 1967 was dismissed by this Court on July 17, 1969. The Court held that the sale of standing trees is not a sale of goods chargeable to sales tax under the Bombay Sales Tax Act, 1959.

2. The State later filed a petition for review, contending that before the appeal was heard, the Bombay Sales Tax Act, 1959 had been amended by the Maharashtra Act 15 of 1967, with retrospective effect. The amendment included "standing timber" in the definition of goods, but counsel for the State had failed to bring this to the Court's attention. As the judgment was claimed to suffer from an error apparent on the record, review was granted.

3. Under Section 2(13) of the Bombay Sales Tax Act, 1959, the definition of "goods" was altered by Maharashtra Act 15 of 1967 to include: *"… and all standing timber which is agreed to be severed before sale or under the contract of sale."* This definition applied retrospectively.

4. The expression "sale of goods" in Entry 54, List II of the Constitution has the same meaning as in the Sale of Goods Act, 1930. In *State of Madras v. Gannon Dunkerley & Co.* (1959 SCR 379; AIR 1958 SC 560), this Court held that "sale of goods" was a term of well-recognized legal import and must be interpreted in the same way under the Constitution. The State Legislature may not extend the meaning to cover transactions not falling within the Sale of Goods Act.

5. By Article 366(12) of the Constitution, "goods" is inclusively defined as "all materials, commodities and articles." However, the Sale of Goods Act, 1930 defines goods more precisely as "every kind of moveable property other than actionable claims and money; and includes stock and shares, growing crops, grass, and things attached to or forming part of the land which are agreed to be severed before sale or under the contract of sale."

6. In the present case, the contract expressly provided that the timber agreed to be sold would be severed. The timber was therefore "goods" under Section 2(7) of the Sale of Goods Act. Accordingly, the sale of timber agreed to be severed under the contract qualified as a "sale of goods" for the purposes of Entry 54, List II of the Constitution.

Case Context Construction

Case Context

Appellants: The State

Respondents: Not explicitly mentioned, referred to as "the Court"

Issue: Whether the sale of standing timber is chargeable to sales tax under the Bombay Sales Tax Act, 1959, considering the retrospective amendment that includes standing timber in the definition of goods.

Appellant's Stance: The State is seeking to have the previous judgment reviewed and to assert that the sale of standing timber is subject to sales tax due to the retrospective amendment of the definition of goods.

Respondent's Stance: Not explicitly stated, but implied that the previous judgment dismissing the appeal should be upheld.

Present Court: This Court

Decision Point Extraction

Decision Point 1

Issue: Whether the sale of standing trees constitutes a sale of goods chargeable to sales tax under the Bombay Sales Tax Act following the amendment by Maharashtra Act 15 of 1967.

Decision Maker: This Court.

Outcome: The appeal was dismissed, and the sale of standing trees was held not to be chargeable to sales tax.

Time: July 17, 1969.

Present Court Decision: True

Decision Point 2

Issue: Whether the definition of "goods" in the Bombay Sales Tax Act was altered by the Maharashtra Act 15 of 1967.

Decision Maker: This Court.

Outcome: The definition of "goods" was confirmed to include standing timber as per the retrospective amendment.

Reasoning: The Court highlighted that the amendment provided that "standing timber which is agreed to be severed before sale or under the contract of sale" was to be considered as goods. This change was deemed to have retrospective operation, thereby affecting the interpretation of sales tax liability.

Present Court Decision: True

Decision Point 3

Issue: The interpretation of the expression "sale of goods" in the context of the Constitution and its alignment with the Sale of Goods Act.

Decision Maker: This Court.

Outcome: The Court affirmed that the expression "sale of goods" in Entry 54, List II of Sch. VII of the Constitution has the same connotation as in the Sale of Goods Act.

Reasoning: The Court referenced previous decisions, establishing that the term "sale of goods" was well-recognized in legal practice and must be interpreted consistently across legislative frameworks. The Court concluded that the State Legislature could not extend the definition to impose tax on transactions that do not qualify as sales of goods under the Sale of Goods Act.

Present Court Decision: True

Decision Point 4

Issue: The interpretation of the term "goods" as defined in the Sale of Goods Act and its application to timber agreed to be severed under a contract of sale.

Decision Maker: This Court.

Outcome: The timber was determined to be "goods" within the meaning of Section 2 (7) of the Sale of Goods Act.

Reasoning: The Court concluded that since the contract expressly provided for the severance of timber before sale, it qualified as "goods" under the Sale of Goods Act. The expression "sale of goods" in the Constitution was found to have the same meaning as in the Sale of Goods Act, thus allowing the sale of timber to be regarded as a sale of goods.

Present Court Decision: True

Present Court Ruling Generation

Present Court Ruling

The present court ruled that the sale of standing timber, when agreed to be severed before sale or under the contract of sale, constitutes a sale of goods chargeable to sales tax under the Bombay Sales Tax Act, 1959. The court confirmed that the retrospective amendment to the definition of "goods" in the Act, as introduced by Maharashtra Act 15 of 1967, includes standing timber, thereby affirming the taxability of such sales.

In its final statements, the court emphasized that while standing timber may not ordinarily be regarded as "goods," the inclusive definition provided in Section 2 (7) of the Sale of Goods Act allows for items attached to land to be considered goods if they are to be severed before sale or under the terms of the contract. The court noted that the contract in question explicitly stated that the timber was to be severed, thus qualifying it as "goods" under the Sale of Goods Act.

The court's reasoning was grounded in the consistent legal interpretation of the term "sale of goods" across various legislative contexts, asserting that the definition of "goods" in the Constitution aligns with that in the Sale of Goods Act. The court concluded that the State Legislature could not extend the definition to impose tax on transactions that do not qualify as sales of goods under the Sale of Goods Act.

As a result, the court ordered that the sale of standing timber is indeed subject to sales tax, thereby granting the State's appeal for the retrospective application of the amendment. There were no specific timelines or compliance requirements mentioned in the ruling, but the court's decision effectively reinstates the tax liability on the sale of standing timber as defined by the amended law.

Judgment Prediction and Structured Explanation Generation

Judgment Prediction

Prediction: 1 (Appeal Granted)

Explanation

Facts of the Case:
The case involves an appeal by the State regarding the taxability of the sale of standing timber under the Bombay Sales Tax Act, 1959. The State contends that a retrospective amendment to the definition of "goods" in the Act, introduced by Maharashtra Act 15 of 1967, includes standing timber, thereby making its sale subject to sales tax. The previous judgment dismissed the appeal, asserting that standing timber was not chargeable to sales tax. The State seeks to overturn this ruling.

Legal Issue(s) Presented:
The primary legal question is whether the sale of standing timber is chargeable to sales tax under the Bombay Sales Tax Act, considering the retrospective amendment that includes standing timber in the definition of goods.

Applicable Law and Precedents:
Key statutes and legal principles involved include:
- The Bombay Sales Tax Act, 1959, particularly the definition of "goods" as amended by Maharashtra Act 15 of 1967.
- The Sale of Goods Act, specifically Section 2 (7), which defines "goods" and includes items attached to land that are agreed to be severed before sale.
- Relevant constitutional provisions regarding the definition of "sale of goods" in Entry 54, List II of Schedule VII.

Analysis / Reasoning:
The court's analysis focused on the retrospective nature of the amendment to the definition of "goods" in the Bombay Sales Tax Act. It determined that standing timber, when agreed to be severed before sale or under the terms of the contract, qualifies as "goods" under the Sale of Goods Act. The court emphasized that the term "sale of goods" has a consistent legal interpretation across various legislative contexts, and the retrospective amendment effectively imposes tax liability on the sale of standing timber. The court also noted that the previous judgment failed to consider the amendment, which was crucial in determining the taxability of such sales.

Predicted Conclusion:
Appeal Granted. The court is expected to rule that the sale of standing timber is subject to sales tax under the Bombay Sales Tax Act, affirming the retrospective application of the amendment to the definition of "goods."

Figure 3: Outputs at each processing step for an example court case.