

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

Anonymous ACL submission

## 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<sup>1</sup>, as of 21<sup>st</sup> March 2025, there were approximately 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*,

<sup>1</sup>Court-wise Number of Cases Pending in Various Courts as per Information Available on National Judicial Data Grid

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 derived 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. For the sake of reproducibility, we have made the code accessible via a GitHub link<sup>2</sup>.

## 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. (Aletas 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,

<sup>2</sup><https://anonymous.4open.science/r/Vichara-9706>

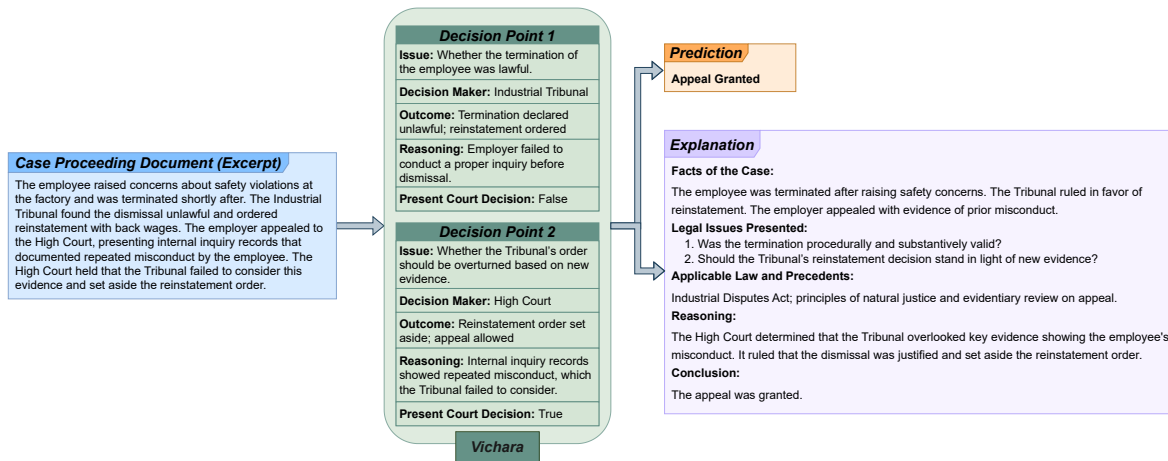


Figure 1: Vichara’s appellate judgment prediction and explanation for a sample case excerpt.

applicable statutes, and sentence lengths, supported by the large-scale CAIL2018 dataset (Xiao et al., 2018). In the European legal context, (Chalkidis et al., 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 et al., 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 pre-

dictive 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 one of two categories. Some are *extractive* (Prasad et al., 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.

### 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 out-

247	come labels (Appeal Granted or Dismissed)	the key legal issue, the court deciding the appeal,	294
248	and expert-annotated explanations.	the parties, and their respective stances. Vichara	295
249	• <b>ILDC_expert:</b> ILDC_expert is the expert-	extracts decision points that summarize the court’s	296
250	annotated subset of the Indian Legal Docu-	legal determinations, filtering them to retain only	297
251	ments Corpus (Malik et al., 2021). It con-	those attributable to the present court. These fil-	298
252	sists of 56 Supreme Court of India appellate	tered decision points are then used to generate the	299
253	cases, each annotated with binary outcome	present court’s ruling, ensuring it reflects solely the	300
254	labels (Appeal Granted or Dismissed) and cor-	outcome of the appeal under consideration. The	301
255	responding expert provided explanations.	predicted judgment is derived by comparing this	302
256	All case proceedings and explanations in both	ruling with the appellant’s stance. Finally, Vichara	303
257	datasets are written in English.	produces an IRAC-style (Metzler, 2002) explan-	304
258	<b>4 Vichara Task Formulation</b>	ation grounded in the extracted facts, laws, and rea-	305
259	Vichara consists of a unified pipeline with two in-	soning. The overall architecture is illustrated in	306
260	terconnected components:	Figure 2, with detailed prompt templates in Table 5	307
261	<b>4.1 Judgment Prediction Component</b>	in Appendix B. Next, we explain each of the steps	308
262	Given a segment from a Supreme Court of India	in the methodology.	309
263	or high court appellate case proceeding document,	<b>5.1 Rhetorical Role Classification</b>	310
264	Vichara predicts whether the court ruled in favor	Rhetorical role labelling of sentences in a legal docu-	311
265	of the appellant. The output is a binary label: {1,	ment refers to identifying the semantic function	312
266	0}, where 1 denotes that the appeal was granted,	each sentence serves, such as stating the facts of	313
267	and 0 indicates dismissal. While real cases can	the case, presenting arguments of the parties, or de-	314
268	involve mixed outcomes, Vichara abstracts these	livering the court’s judgment (Bhattacharya et al.,	315
269	into a single binary decision to focus on the core	2019).	316
270	outcome.	The input case proceeding document is seg-	317
271	<b>4.2 Explanation Generation Component</b>	mented into sentences using the SAT-12L (Seg-	318
272	Alongside predicting the outcome, Vichara pro-	ment Any Text) pretrained model, fine-tuned via	319
273	duces a structured rationale explaining the predic-	Low-Rank Adaptation (LoRA) for accurate sen-	320
274	tion. The explanation captures the key facts, ap-	tence boundary detection in legal texts (Frohmann	321
275	plicable laws, and legal reasoning, following an	et al., 2024). (Bhattacharya et al., 2019) define	322
276	IRAC-inspired structure adapted to Indian legal	seven categories of rhetorical roles, namely ‘Facts’,	323
277	discourse.	‘Ruling by Lower Court’, ‘Argument’, ‘Statute’,	324
278	Figure 1 illustrates Vichara’s task formulation.	‘Precedent’, ‘Ratio of the Decision’, and ‘Ruling	325
279	<b>5 Methodology</b>	by Present Court’. Each sentence in our dataset is	326
280	Vichara comprises six sequential stages: rhetori-	classified into one of these seven categories by a	327
281	cal role classification, case context construction,	hierarchical BiLSTM-CRF model trained to assign	328
282	decision point extraction, present court ruling gen-	rhetorical roles, following the schema introduced	329
283	eration, judgment prediction, and explanation gen-	by (Bhattacharya et al., 2019).	330
284	eration.	<b>5.2 Case Context Construction</b>	331
285	The process begins by classifying each sentence	To generate case-level context, an LLM is applied	332
286	in the input case proceeding document according	to the subset of sentences labeled as ‘Facts’ during	333
287	to its rhetorical role (Bhattacharya et al., 2019),	rhetorical role classification (Section 5.1). The	334
288	identifying whether it states facts, presents argu-	LLM extracts six key fields:	335
289	ments, cites precedents, or issues rulings. All subse-	• Appellants: The individuals or entities filing	336
290	quent stages in the pipeline operate purely through	the appeal.	337
291	prompting without any fine-tuning of the under-	• Respondents: The opposing parties in the	338
292	lying LLMs. The case context is extracted exclu-	case.	339
293	sively from sentences labeled as facts, capturing	• Issue: The main legal or factual question un-	340
		der consideration.	341

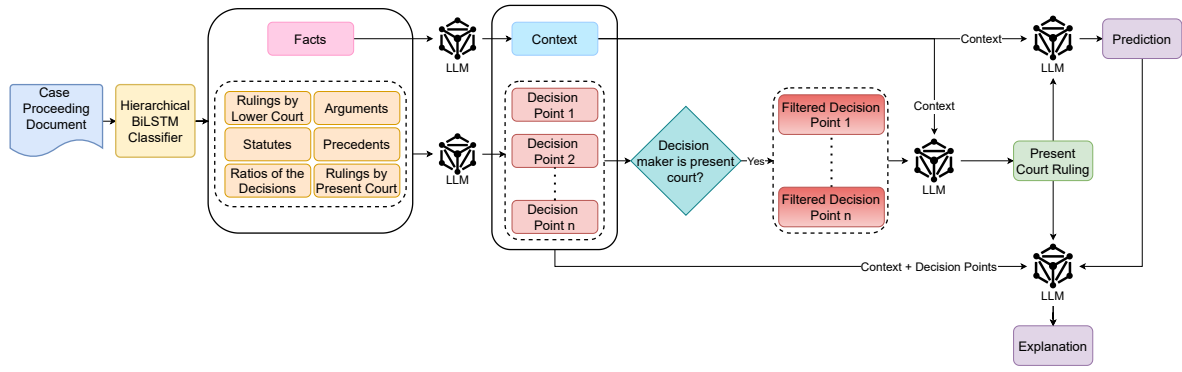


Figure 2: Vichara flow diagram

- 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 model from incorporating rulings made by lower courts, which may appear earlier in the 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

411 outputs a prediction of 1 (Appeal Granted). Other-  
412 wise, if the relief is entirely denied, the prediction  
413 is 0 (Appeal Dismissed).

## 414 5.6 Structured Explanation Generation

415 The final stage of the pipeline involves generating a  
416 structured explanation for the predicted outcome of  
417 the appeal. An LLM is used to synthesize the expla-  
418 nation. The inputs to the LLM include the case con-  
419 text (Section 5.2), extracted decision points (Sec-  
420 tion 5.3), present court ruling (Section 5.4) and the  
421 predicted outcome (Section 5.5).

422 The LLM is prompted to organize the explana-  
423 tion into five predefined sections:

- 424 • Facts of the Case: A brief overview of the  
425 dispute, parties involved, and procedural back-  
426 ground.
- 427 • Legal Issue(s) Presented: The central legal  
428 questions that the court addressed.
- 429 • Applicable Law and Precedents: Statutes, con-  
430 stitutional provisions, or prior case law cited  
431 or relied upon.
- 432 • Analysis / Reasoning: The logical application  
433 of the law to the facts, incorporating relevant  
434 decision points to justify the predicted out-  
435 come.
- 436 • Predicted Conclusion: A summary conclusion  
437 reflecting the model’s judgment (e.g., “Appeal  
438 Granted” or “Appeal Dismissed”).

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

## 441 6 Results and Analysis

442 In this section, we present the evaluation of Vichara  
443 on both judgment prediction and explanation qual-  
444 ity, as well as an ablation study to assess the neces-  
445 sity of each stage of the Vichara framework. We  
446 evaluate the framework using four LLMs, GPT-  
447 4o mini (OpenAI, 2024), Llama-3.1-8B<sup>3</sup>, Mistral-  
448 7B<sup>4</sup> and Qwen2.5-7B<sup>5</sup>. All experiments were re-  
449 peated across 5 independent random seeds to ac-  
450 count for the non-deterministic nature of LLMs.  
451 Reported results for judgment prediction and auto-  
452 mated evaluation of explanations correspond to the  
453 mean performance across seeds, with standard devi-  
454 ations included to indicate robustness. The ablation  
455 study, to assess the contribution of each stage of  
456 the Vichara framework is provided in Appendix C.

<sup>3</sup>Llama-3.1-8B

<sup>4</sup>Mistral-7B

<sup>5</sup>Qwen2.5-7B

## 457 6.1 Results for Judgment Prediction

458 We evaluate the binary outcome of judgment predic-  
459 tion using standard classification metrics: accuracy,  
460 precision, recall, and F1-score. To ensure balanced  
461 evaluation across both classes (Appeal Granted and  
462 Dismissed), we report macro-averaged scores. Re-  
463 sults are presented in Table 1.

464 Among the LLMs evaluated in our experi-  
465 ments within the Vichara framework, GPT-4o mini  
466 achieves the highest scores on both datasets, fol-  
467 lowed by Llama-3.1-8B, Qwen2.5-7B, and Mistral-  
468 7B. Compared to the state-of-the-art baseline, IN-  
469 LegalLlama (Nigam et al., 2025), all LLMs evalu-  
470 ated in our experiments surpass its performance on  
471 the ILDC\_expert dataset, while on PredEx, GPT-  
472 4o mini, and Llama-3.1-8B outperform INLegalL-  
473 lama.

## 474 6.2 Results for Explanation

475 To evaluate explanation quality, we rely on human  
476 assessments conducted by experts from the legal  
477 domain. We use three metrics: Clarity, Linking,  
478 and Usefulness, which together capture essential  
479 aspects of interpretability, legal alignment, and practi-  
480 cal value. These metrics are particularly important  
481 in our setting, as the structured explanations pro-  
482 duced by our framework differ significantly in form  
483 and granularity from the gold standard references  
484 in existing datasets, making reference-based auto-  
485 matic metrics less reliable.

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

- 489 • Clarity: Measures how well-structured, read-  
490 able, and logically coherent the explanation  
491 is.
- 492 • Linking: Assesses the degree to which the  
493 explanation offers a justifiable connection be-  
494 tween the facts and the predicted outcome.
- 495 • Usefulness: Evaluates how informative and  
496 practically helpful the explanation is for a  
497 legal professional seeking to understand the  
498 court’s reasoning and apply it in real-world  
499 legal work.

500 For the evaluation, we recruited three legal ex-  
501 perts, all practicing advocates with 7-8 years of  
502 professional experience. Each expert was provided  
503 with the case proceeding document, the predicted

Model	PredEx				ILDC_expert			
	Accuracy	Macro Precision	Macro Recall	Macro F1	Accuracy	Macro Precision	Macro Recall	Macro F1
<b>GPT-4o mini</b>	<b>81.62 ± 0.41</b>	<b>81.57 ± 0.43</b>	<b>81.45 ± 0.40</b>	<b>81.50 ± 0.42</b>	<b>80.36 ± 0.37</b>	<b>81.25 ± 0.34</b>	<b>80.65 ± 0.36</b>	<b>80.30 ± 0.35</b>
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
<b>GPT-4o mini</b>	<b>4.57 ± 0.22</b>	<b>4.96 ± 0.21</b>	<b>4.37 ± 0.18</b>
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.

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-L, BLEU, METEOR, BERTScore, and BLANC) for completeness. We use the ground truth explanations avail-

able 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.

## 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 Usefulness. 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.

Models	Lexical Based Evaluation (%)					Semantic Evaluation (%)	
	Rouge-1	Rouge-2	Rouge-L	BLEU	METEOR	BERTScore(F1)	BLANC
<b>PredEx</b>							
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	<b>82.90 ± 0.37</b>	13.41 ± 0.22
<b>Qwen2.5-7B</b>	<b>38.70 ± 0.43</b>	<b>15.80 ± 0.34</b>	<b>19.49 ± 0.30</b>	<b>7.06 ± 0.25</b>	<b>23.77 ± 0.35</b>	81.66 ± 0.38	<b>13.59 ± 0.23</b>
<b>ILDC_expert</b>							
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	<b>16.45 ± 0.33</b>	16.48 ± 0.27	1.01 ± 0.09	14.13 ± 0.26	<b>82.77 ± 0.34</b>	<b>13.01 ± 0.19</b>
<b>Qwen2.5-7B</b>	<b>36.13 ± 0.39</b>	15.67 ± 0.32	<b>17.41 ± 0.28</b>	<b>4.16 ± 0.18</b>	<b>17.45 ± 0.30</b>	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.

## 568 Limitations

569 While Vichara demonstrates strong performance in  
570 both judgment prediction and structured explana-  
571 tion generation, several limitations remain.

572 Our evaluation is restricted to the Indian judi-  
573 ciary, with a primary focus on appellate-level cases  
574 from the Supreme Court and selected High Courts.  
575 Although the core methodology may be applicable  
576 to other legal systems, transferring the approach to  
577 jurisdictions with different procedural structures,  
578 legal doctrines, or language conventions would re-  
579 quire substantial adaptation and validation.

580 Vichara relies on prompt-based querying of  
581 LLMs at multiple stages of the pipeline. This  
582 approach introduces variability due to the non-  
583 deterministic nature of model outputs and the sen-  
584 sitivity of results to prompt phrasing.

585 The human evaluation of explanation quality was  
586 conducted on a limited sample consisting of 25 ex-  
587 planations per language model, reviewed by three  
588 advocates. While this provides useful qualitative in-  
589 sights, broader evaluation involving a more diverse  
590 pool of legal professionals and a wider range of  
591 case types is necessary to establish generalizability  
592 and practical relevance.

593 Finally, the multi-stage architecture of Vichara,  
594 which includes multiple LLM calls, may impose  
595 computational and deployment challenges. These  
596 constraints could be particularly limiting in envi-  
597 ronments with restricted resources or strict latency  
598 requirements.

599 Future research will focus on investigating meth-  
600 ods to reduce computational overhead and expand-  
601 ing evaluation to include additional legal domains  
602 and jurisdictions. This includes exploring strate-  
603 gies such as prompt optimization and model dis-

604 tillation to improve scalability while maintaining  
605 performance in real-world legal applications. To  
606 broaden evaluation, we plan to assess the frame-  
607 work on a wider range of case types, court levels,  
608 and jurisdiction-specific datasets, enabling a more  
609 comprehensive understanding of its effectiveness  
610 across diverse legal contexts.

## Ethical Considerations

611 In conducting this research, we adhered to ethical  
612 standards in both data usage and human evalua-  
613 tion. All legal case documents used in our exper-  
614 iments are publicly available and drawn from es-  
615 tablished open-access legal datasets. No private,  
616 confidential, or sensitive information was accessed  
617 or utilized at any stage. For the human evalua-  
618 tion of model-generated explanations, we engaged  
619 three practicing advocates with formal legal train-  
620 ing and courtroom experience. The advocates were  
621 recruited through the authors’ professional and aca-  
622 demic networks based on their domain expertise.  
623 Their participation was entirely voluntary and un-  
624 compensated. The evaluation was conducted with  
625 informed consent, and participants were briefed on  
626 the purpose and scope of the study. AI-based writ-  
627 ing assistants were used for minor language editing  
628 and clarity improvements in writing this paper.  
629

## References

- 630 Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel  
631 Preoṭiuc-Pietro, and Vasileios Lampos. 2016. Pre-  
632 dicting judicial decisions of the european court of  
633 human rights: A natural language processing per-  
634 spective. *PeerJ computer science*, 2:e93.  
635  
636 Paheli Bhattacharya, Shounak Paul, Kripabandhu  
637 Ghosh, Saptarshi Ghosh, and Adam Wyner. 2019.

638	Identification of rhetorical roles of sentences in indian legal judgments. In <i>Legal knowledge and information systems</i> , pages 3–12. IOS Press.	692
639		693
640		694
641	Stephen G Breyer. 2006. Reflections on the role of appellate courts: A view from the supreme court. <i>J. App. Prac. &amp; Process</i> , 8:91.	695
642		696
643		697
644	Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. 2019. Neural legal judgment prediction in english. In <i>Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics</i> , pages 4317–4323.	698
645		699
646		700
647		701
648		702
649	Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. <i>Psychological bulletin</i> , 76(5):378.	703
650		704
651		705
652	Markus Frohmann, Igor Sterner, Ivan Vulić, Benjamin Minixhofer, and Markus Schedl. 2024. Segment any text: A universal approach for robust, efficient and adaptable sentence segmentation. In <i>Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing</i> , pages 11908–11941.	706
653		707
654		708
655		709
656		710
657		711
658	Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. <i>arXiv preprint arXiv:2407.21783</i> .	712
659		713
660		714
661		715
662		716
663	Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, L��lio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timoth��e Lacroix, and William El Sayed. 2023. <i>Mistral 7b</i> . Preprint, arXiv:2310.06825.	717
664		718
665		719
666		720
667		721
668		722
669		723
670		724
671	Vijit Malik, Rishabh Sanjay, Shubham Kumar Nigam, Kripabandhu Ghosh, Shouvik Kumar Guha, Arnab Bhattacharya, and Ashutosh Modi. 2021. Ildc for cjpe: Indian legal documents corpus for court judgment prediction and explanation. In <i>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)</i> , pages 4046–4062.	725
672		726
673		727
674		728
675		729
676		730
677		731
678		732
679		733
680	Masha Medvedeva, Michel Vols, and Martijn Wieling. 2020. Using machine learning to predict decisions of the european court of human rights. <i>Artificial Intelligence and Law</i> , 28(2):237–266.	734
681		735
682		736
683		737
684	Jeffrey Metzler. 2002. The importance of irac and legal writing. <i>U. Det. Mercy L. Rev.</i> , 80:501.	738
685		739
686		740
687	Shubham Nigam, Anurag Sharma, Danush Khanna, Noel Shallum, Kripabandhu Ghosh, and Arnab Bhattacharya. 2024a. Legal judgment reimagined: PreDEX and the rise of intelligent ai interpretation in indian courts. In <i>Findings of the Association for Computational Linguistics ACL 2024</i> , pages 4296–4315.	741
688		742
689		743
690		744
691		745
	Shubham Kumar Nigam, Deepak Patnaik Balaramamahanthi, Shivam Mishra, Noel Shallum, Kripabandhu Ghosh, and Arnab Bhattacharya. 2025. Nyayaanumana and inlegalllama: The largest indian legal judgment prediction dataset and specialized language model for enhanced decision analysis. In <i>Proceedings of the 31st International Conference on Computational Linguistics</i> , pages 11135–11160.	692
		693
		694
		695
		696
		697
		698
		699
	Shubham Kumar Nigam and Aniket Deroy. 2023. Fact-based court judgment prediction. In <i>Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation</i> , pages 78–82.	700
		701
		702
		703
	Shubham Kumar Nigam, Aniket Deroy, Subhankar Maity, and Arnab Bhattacharya. 2024b. Rethinking legal judgement prediction in a realistic scenario in the era of large language models. In <i>Proceedings of the Natural Language Processing Workshop 2024</i> , pages 61–80.	704
		705
		706
		707
		708
		709
	OpenAI. 2024. GPT-4o mini. <a href="https://platform.openai.com/docs/models/gpt-4o-mini">https://platform.openai.com/docs/models/gpt-4o-mini</a> .	710
		711
		712
	James Phoenix and Mike Taylor. 2024. <i>Prompt engineering for generative AI</i> . " O'Reilly Media, Inc."	713
		714
	Nishchal Prasad, Mohand Boughanem, and Taoufiq Dkaki. 2023. A hierarchical neural framework for classification and its explanation in large unstructured legal documents. In <i>The 3rd International Workshop on Mining and Learning in the Legal Domain</i> .	715
		716
		717
		718
		719
	Octavia-Maria ��ulea, Marcos Zampieri, Mihaela Vela, and Josef van Genabith. 2017. Predicting the law area and decisions of French Supreme Court cases. In <i>Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017</i> , pages 716–722, Varna, Bulgaria. INCOMA Ltd.	720
		721
		722
		723
		724
		725
		726
	Chaojun Xiao, Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Yansong Feng, Xi-anpei Han, Zhen Hu, Heng Wang, and 1 others. 2018. CAIL2018: A large-scale legal dataset for judgment prediction. <i>arXiv preprint arXiv:1807.02478</i> .	727
		728
		729
		730
		731
	Hiroaki Yamada, Takenobu Tokunaga, Ryutaro Ohara, Akira Tokutsu, Keisuke Takeshita, and Mihoko Sumida. 2024. Japanese tort-case dataset for rationale-supported legal judgment prediction. <i>Artificial Intelligence and Law</i> , pages 1–25.	732
		733
		734
		735
		736
	A Yang, Baosong Yang, B Zhang, B Hui, B Zheng, B Yu, Chengpeng Li, D Liu, F Huang, H Wei, and 1 others. 2024. Qwen2. 5 technical report. <i>arXiv preprint</i> .	737
		738
		739
		740
	Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018. Legal judgment prediction via topological learning. In <i>Proceedings of the 2018 conference on empirical methods in natural language processing</i> , pages 3540–3549.	741
		742
		743
		744
		745

746	<b>A Human Evaluation Guidelines</b>		
747	The following instructions were provided to human		
748	evaluators for assessing the quality of the generated		
749	explanations. Each explanation was evaluated inde-		
750	pendently along three dimensions using a 5-point		
751	Likert scale.		
752	<b>A.1 Clarity</b>		
753	Measures how clearly the explanation is written		
754	and how well it communicates the rationale behind		
755	the decision.		
756	1: The explanation is confusing or incoherent.		
757	The rationale is difficult to follow.		
758	2: Some parts are understandable, but the reason-		
759	ing is vague or underdeveloped.		
760	3: The explanation is moderately clear but may		
761	lack smooth flow or sufficient detail.		
762	4: The rationale is clearly presented and easy to		
763	follow. Terminology is appropriate.		
764	5: The explanation is very well-written, logically		
765	structured, and highly understandable.		
766	<b>A.2 Linking</b>		
767	Measures how well the explanation connects the		
768	facts and legal reasoning to the final predicted out-		
769	come (e.g., Appeal Granted or Dismissed).		
770	1: The explanation does not connect to the out-		
771	come at all or is highly inconsistent.		
772	2: Weak or unclear linkage between the reason-		
773	ing and the final decision.		
774	3: Some linkage exists, but gaps or ambiguities		
775	are present.		
776	4: Clear and logical connection to the judgment,		
777	with minimal gaps.		
778	5: Strong and coherent justification that clearly		
779	supports the predicted decision.		
780	<b>A.3 Usefulness</b>		
781	Measures how useful the explanation would be for		
782	a human reader, particularly a legal practitioner,		
783	trying to understand the reasoning behind the AI's		
784	decision.		
785	1: The explanation is not helpful or usable in any		
786	practical sense.		
	2: Limited usefulness; lacks essential detail or		787
	context.		788
	3: Somewhat helpful, but may miss key points		789
	or feel generic.		790
	4: Offers clear value in understanding the reason-		791
	ing and potential legal implications.		792
	5: Highly informative and usable; effectively		793
	mirrors real-world legal reasoning.		794
	<b>B Prompts for Each Processing Step</b>		795
	Table 5 contains the prompts used for each stage of		796
	the processing pipeline. The prompts were devel-		797
	oped through an iterative prompt engineering pro-		798
	cess, inspired by the principles on prompt design		799
	(Phoenix and Taylor, 2024). Figure 3 illustrates		800
	the outputs generated at each processing step using		801
	GPT-4o mini, for an example court case.		802
	<b>C Ablation Study</b>		803
	To assess the necessity and contribution of each		804
	component within the Vichara framework, we con-		805
	duct a systematic ablation study across four stages:		806
	(1) rhetorical role classification, (2) case context		807
	construction, (3) decision point extraction, and (4)		808
	present court ruling generation.		809
	<b>C.1 Experimental Setup</b>		810
	We create a series of ablated variants of the full		811
	pipeline by selectively removing individual stages		812
	while holding all other configurations constant.		813
	When a stage is removed, subsequent prompts are		814
	minimally adjusted to maintain coherence and en-		815
	sure functionality without referencing missing in-		816
	puts. All evaluations are performed using the GPT-		817
	4o mini model on the ILDC_expert dataset. We re-		818
	port accuracy and macro-F1 for the predictive com-		819
	ponent (judgment prediction), and Clarity, Linking,		820
	and Usefulness scores for the human evaluation		821
	of explanations. Judgment prediction results are		822
	reported as the mean of five independent runs with		823
	different random seeds, along with the standard		824
	deviation reflecting variation across seeds. The		825
	explanation evaluation was conducted on 15 ex-		826
	planation outputs from a single run, as assessing		827
	multiple runs would require substantial additional		828
	human effort. It was carried out by the same three		829
	legal experts who evaluated the explanations in		830
	Section 6.2. For explanation quality, we report the		831
	mean score across the three evaluators, with the		832

Configuration	Accuracy	Macro F1	Clarity	Linking	Usefulness
<b>Full Vichara (All Stages)</b>	<b>80.36 ± 0.37</b>	<b>81.25 ± 0.34</b>	<b>4.57 ± 0.22</b>	<b>4.96 ± 0.21</b>	<b>4.37 ± 0.18</b>
– Rhetorical Role Classification	79.12 ± 0.45	78.85 ± 0.41	4.32 ± 0.28	4.66 ± 0.24	4.21 ± 0.25
– Case Context Construction	77.84 ± 0.52	77.40 ± 0.47	4.21 ± 0.31	4.55 ± 0.26	4.18 ± 0.20
– Decision Point Extraction	71.65 ± 0.61	70.92 ± 0.57	3.80 ± 0.35	4.10 ± 0.28	3.95 ± 0.33
– 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  $\pm$  standard deviation. Bold values indicate the highest performance for each metric.

standard deviation representing variation among evaluators.

## C.2 Ablation Configurations

The ablation study consisted of five configurations:

1. **Full Vichara (All Stages)**: The complete pipeline as described in Section 5.
2. **– 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. **– 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. **– 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. **– 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.

## C.3 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 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.

Step	Prompt
Case Context Construction	<p>You are a legal assistant helping summarize appeal case details.</p> <p>Given the following facts from an appeal case document, extract the following information about the current appeal only:</p> <ol style="list-style-type: none"> <li>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").</li> <li>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").</li> <li>3. Issue – the main legal or factual issue being disputed in the current appeal.</li> <li>4. Appellant’s Stance (in the current appeal) – clearly state what the appellant is arguing for or seeking in the present appeal.</li> <li>5. Respondent’s Stance (in the current appeal) – clearly state what the respondent is arguing for or seeking in the present appeal.</li> <li>6. Present Court – the court deciding the present appeal (e.g., Supreme Court of India, High Court of Bombay).</li> </ol> <p>### Important instructions:</p> <ul style="list-style-type: none"> <li>- Do NOT assume the appellant is the party introduced first. Carefully check who has filed the current appeal.</li> <li>- Do NOT summarize or include opinions or findings of lower courts unless those are being specifically challenged in this appeal.</li> <li>- Focus on the actual parties to the legal dispute.</li> <li>- Do NOT invent names or facts. If something is not mentioned, leave it as an empty string.</li> <li>- Use only the output format specified below.</li> </ul> <p>### Output Format:</p> <pre> {{ "appellants": "&lt;name or description of appellant&gt;", "respondents": "&lt;name or description of respondent&gt;", "issue": "&lt;brief summary of the legal/factual issue&gt;", "appellant_stance": "&lt;stance of the appellant&gt;", "respondent_stance": "&lt;stance of the respondent&gt;", "present_court": "&lt;name of the court currently deciding the appeal&gt;" }}</pre> <p>### Facts: &lt;{facts}&gt;</p>

Decision Point Ex- traction	<p>You are a legal assistant tasked with extracting <b>**all decision points**</b> from an excerpt of a court case proceeding. Decision points are discrete legal determinations made at different stages throughout the case.</p> <p>Given the following text and the present court, extract <b>**all identifiable decision points**</b> and output them in strict JSON format as a list of objects.</p> <p>Each decision point object should include:</p> <ul style="list-style-type: none"> <li>- "issue": the legal issue or question being addressed</li> <li>- "decision_maker": the court or authority that made the decision (e.g., Supreme Court, Trial Court, High Court)</li> <li>- "outcome": the result or resolution of the issue</li> <li>- "time": (optional) the date or timeframe of the decision if mentioned</li> <li>- "reasoning": (optional) summary of the Court's reasoning, including references to statutes, arguments, facts, or precedents</li> <li>- "present_court_decision": true if the "decision_maker" is the same as the present court provided, otherwise false</li> </ul> <p>### Important instructions:</p> <ul style="list-style-type: none"> <li>- Do not include any extra text or explanation.</li> <li>- Do not assume or hallucinate decision makers.</li> <li>- Do NOT include triple backticks ("").</li> </ul> <p>### Output Format:</p> <pre>[[   {     "issue": "&lt;string&gt;",     "decision_maker": "&lt;string&gt;",     "outcome": "&lt;string&gt;",     "time": "&lt;string or null&gt;",     "reasoning": "&lt;string or null&gt;",     "present_court_decision": &lt;true or false&gt;   } ]]</pre> <p>### Input:</p> <p>Present Court: "&lt;{present_court}&gt;"</p> <p>Text: &lt;{group_text}&gt;</p>
--------------------------------	--

Present Court Ruling Generation	<p>Your goal is to identify the <b>final ruling of the present court</b> in this appeal — that is, what the present court ultimately decided and ordered.</p> <p><b>Case Context:</b></p> <p>Below is the context of the case, which clearly identifies: - Who the <b>appellant</b> is (the party who filed the appeal)</p> <ul style="list-style-type: none"> <li>- Who the <b>respondent</b> is (the party defending against the appeal)</li> <li>- What the <b>main issue</b> of the appeal is.</li> <li>- Appellant’s Stance (in the current appeal) – What the appellant is arguing for or seeking <b>in the present appeal</b></li> <li>- Respondent’s Stance (in the current appeal) – What the respondent is arguing for or seeking <b>in the present appeal</b></li> </ul> <p>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, <b>ONLY</b> then infer these from the decision points.</p> <p>&lt;{context}&gt;</p> <hr/> <p><b>Decision Points:</b></p> <p>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.</p> <p>&lt;{present_court_points}&gt;</p> <hr/> <p><b>Final Statements from the Present Court:</b></p> <p>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.</p> <p>&lt;{final_statement}&gt;</p> <hr/> <p><b>Your Task:</b></p> <ol style="list-style-type: none"> <li>1. Provide a comprehensive explanation of the final ruling, including: <ul style="list-style-type: none"> <li>- Specific reliefs granted or denied</li> <li>- Any orders or directions issued</li> <li>- The court’s reasoning and key factors considered</li> <li>- Relevant timelines or compliance expectations</li> </ul> </li> <li>2. Focus on the decision points to determine what the present court considered during this appeal.</li> </ol>
---------------------------------	--

Present Court Ruling Generation	<p>3. Use the <b>Final Statements from the Present Court</b> to determine what the court ultimately ruled.</p> <hr/> <p><b>### Important:</b></p> <ul style="list-style-type: none"> <li>- Do not confuse appellants and respondents. Use the parties as stated in the case context.</li> <li>- Respond clearly and concisely.</li> <li>- Do NOT state whether the appeal was granted or dismissed. Your response should only describe the final ruling of the present court.</li> </ul> <hr/> <p><b>### Output Format:</b></p> <p>Final Ruling:  &lt;Provide a detailed and comprehensive explanation of the present court’s final decision in the appeal.&gt;</p>
Judgment Prediction	<p>You are a legal assistant helping to analyze the outcome of an appeal. Your task is to determine whether the <b>present court’s final ruling</b> aligns with what the <b>appellant</b> was seeking in this appeal.</p> <p><b>### Case Context:</b>  The context below includes:</p> <ul style="list-style-type: none"> <li>- Appellants: The <b>appellant</b> (party who filed the appeal)</li> <li>- Respondents: The <b>respondent</b> (party defending the appeal)</li> <li>- Issue: The <b>main issue</b></li> <li>- Appellant’s Stance: What the <b>appellant is seeking</b> in the current appeal</li> <li>- Respondent’s Stance: What the <b>respondent is seeking</b> in the current appeal</li> <li>- Present Court: The court deciding the present appeal.</li> </ul> <p>&lt;{context}&gt;</p> <hr/> <p><b>### Final Court Ruling:</b>  &lt;{court_ruling}&gt;</p> <hr/> <p><b>### Your Task:</b></p> <ul style="list-style-type: none"> <li>- If the <b>court fully or partially granted</b> what the appellant was seeking, output: ‘Prediction: 1’</li> <li>- If the <b>court did not grant</b> what the appellant was seeking, output: ‘Prediction: 0’</li> </ul> <hr/> <p><b>### Output Format:</b>  Prediction: &lt;0 or 1&gt;</p>

Structured Explanation Generation	<p>You are a legal assistant. Your task is to generate a structured legal explanation for the court’s predicted decision in this appeal case.</p> <p>You are given:</p> <ul style="list-style-type: none"> <li>- The <b>case context</b> including appellant, respondent, issue, stances, and the court deciding the appeal</li> <li>- The <b>final court ruling</b> from the present court</li> <li>- A set of <b>decision points</b> extracted from the case. A <b>decision point</b> 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.</li> <li>- The <b>predicted outcome</b> of the appeal</li> </ul> <p>Generate a structured explanation with the following sections:</p> <hr/> <p>Facts of the Case: [A brief summary of the background, parties involved, and what led to the appeal.]</p> <p>Legal Issue(s) Presented: [The legal question(s) the court had to decide.]</p> <p>Applicable Law and Precedents: [Key statutes, constitutional provisions, or case law relied on.]</p> <p>Analysis / Reasoning: [A logical application of law to facts, showing why the court ruled the way it did.]</p> <p>Predicted Conclusion: [Restate the predicted outcome using legal terminology (e.g., 'Appeal Allowed' or 'Dismissed').]</p> <hr/> <p>### Case Context: &lt;{context}&gt;</p> <p>### Final Court Ruling: &lt;{court_ruling}&gt;</p> <p>### Decision Points: &lt;{decision_points_text}&gt;</p> <p>### Predicted Outcome: &lt;{predicted_outcome}&gt;</p>
-----------------------------------	---

Table 5: Prompts for each processing step

Input Case Proceeding Document	<p><b>Case Proceeding Document</b></p> <ol style="list-style-type: none"> <li>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.</li> <li>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.</li> <li>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.</li> <li>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 <i>State of Madras v. Gannon Dunkerley &amp; Co.</i> (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.</li> <li>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."</li> <li>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.</li> </ol>								
Case Context Construction	<p><b>Case Context</b></p> <p><b>Appellants:</b> The State</p> <p><b>Respondents:</b> Not explicitly mentioned, referred to as "the Court"</p> <p><b>Issue:</b> 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.</p> <p><b>Appellant's Stance:</b> 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.</p> <p><b>Respondent's Stance:</b> Not explicitly stated, but implied that the previous judgment dismissing the appeal should be upheld.</p> <p><b>Present Court:</b> This Court</p>								
Decision Point Extraction	<table border="1"> <thead> <tr> <th data-bbox="319 896 566 929">Decision Point 1</th> <th data-bbox="566 896 821 929">Decision Point 2</th> <th data-bbox="821 896 1069 929">Decision Point 3</th> <th data-bbox="1069 896 1311 929">Decision Point 4</th> </tr> </thead> <tbody> <tr> <td data-bbox="319 929 566 1146"> <p><b>Issue:</b> 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.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> The appeal was dismissed, and the sale of standing trees was held not to be chargeable to sales tax.</p> <p><b>Time:</b> July 17, 1969.</p> <p><b>Present Court Decision:</b> True</p> </td> <td data-bbox="566 929 821 1146"> <p><b>Issue:</b> Whether the definition of "goods" in the Bombay Sales Tax Act was altered by the Maharashtra Act 15 of 1967.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> The definition of "goods" was confirmed to include standing timber as per the retrospective amendment.</p> <p><b>Reasoning:</b> 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.</p> <p><b>Present Court Decision:</b> True</p> </td> <td data-bbox="821 929 1069 1146"> <p><b>Issue:</b> The interpretation of the expression "sale of goods" in the context of the Constitution and its alignment with the Sale of Goods Act.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> 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.</p> <p><b>Reasoning:</b> 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.</p> <p><b>Present Court Decision:</b> True</p> </td> <td data-bbox="1069 929 1311 1146"> <p><b>Issue:</b> 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.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> The timber was determined to be "goods" within the meaning of Section 2 (7) of the Sale of Goods Act.</p> <p><b>Reasoning:</b> 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.</p> <p><b>Present Court Decision:</b> True</p> </td> </tr> </tbody> </table>	Decision Point 1	Decision Point 2	Decision Point 3	Decision Point 4	<p><b>Issue:</b> 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.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> The appeal was dismissed, and the sale of standing trees was held not to be chargeable to sales tax.</p> <p><b>Time:</b> July 17, 1969.</p> <p><b>Present Court Decision:</b> True</p>	<p><b>Issue:</b> Whether the definition of "goods" in the Bombay Sales Tax Act was altered by the Maharashtra Act 15 of 1967.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> The definition of "goods" was confirmed to include standing timber as per the retrospective amendment.</p> <p><b>Reasoning:</b> 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.</p> <p><b>Present Court Decision:</b> True</p>	<p><b>Issue:</b> The interpretation of the expression "sale of goods" in the context of the Constitution and its alignment with the Sale of Goods Act.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> 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.</p> <p><b>Reasoning:</b> 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.</p> <p><b>Present Court Decision:</b> True</p>	<p><b>Issue:</b> 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.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> The timber was determined to be "goods" within the meaning of Section 2 (7) of the Sale of Goods Act.</p> <p><b>Reasoning:</b> 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.</p> <p><b>Present Court Decision:</b> True</p>
Decision Point 1	Decision Point 2	Decision Point 3	Decision Point 4						
<p><b>Issue:</b> 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.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> The appeal was dismissed, and the sale of standing trees was held not to be chargeable to sales tax.</p> <p><b>Time:</b> July 17, 1969.</p> <p><b>Present Court Decision:</b> True</p>	<p><b>Issue:</b> Whether the definition of "goods" in the Bombay Sales Tax Act was altered by the Maharashtra Act 15 of 1967.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> The definition of "goods" was confirmed to include standing timber as per the retrospective amendment.</p> <p><b>Reasoning:</b> 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.</p> <p><b>Present Court Decision:</b> True</p>	<p><b>Issue:</b> The interpretation of the expression "sale of goods" in the context of the Constitution and its alignment with the Sale of Goods Act.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> 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.</p> <p><b>Reasoning:</b> 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.</p> <p><b>Present Court Decision:</b> True</p>	<p><b>Issue:</b> 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.</p> <p><b>Decision Maker:</b> This Court.</p> <p><b>Outcome:</b> The timber was determined to be "goods" within the meaning of Section 2 (7) of the Sale of Goods Act.</p> <p><b>Reasoning:</b> 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.</p> <p><b>Present Court Decision:</b> True</p>						
Present Court Ruling Generation	<p><b>Present Court Ruling</b></p> <p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p>								
Judgment Prediction and Structured Explanation Generation	<p><b>Judgment Prediction</b></p> <p>Prediction: 1 (Appeal Granted)</p> <p><b>Explanation</b></p> <p><b>Facts of the Case:</b> 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.</p> <p><b>Legal Issue(s) Presented:</b> 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.</p> <p><b>Applicable Law and Precedents:</b> 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.</p> <p><b>Analysis / Reasoning:</b> 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.</p> <p><b>Predicted Conclusion:</b> 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."</p>								

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