JurisAgents: A Multi-Agent Framework for Legal Judgment Prediction

Anonymous ACL submission

Abstract

Legal Judgment Prediction (LJP) aims to predict the outcomes of legal cases based on factual descriptions, serving as a fundamental task to advance the development of legal systems. Traditional approaches to LJP often rely on statistical analyses of past legal judgments or propose an agent-based framework from a role-playing perspective. However, the existing framework struggles to handle multiple allegations and diverse forms of evidence. Additionally, simplistic courtroom simulations often lead to one-sided decisions and insufficient adaptability. In this paper, we introduce ***JurisAgents***, a novel framework for LJP that effectively decomposes trial tasks, standardizes processes, and organizes them into distinct stages. Furthermore, considering the dynamic nature and real-time updates of legal statutes, we propose ***JurisMM***. It comprises more than 50,000 recent legal case records derived from Chinese judicial proceedings. It includes both unimodal textual data and multimodal data that combine video and text, allowing for a comprehensive examination of the capabilities of our framework. We validate the capability of our framework on both *JurisMM* and the widely used legal benchmark LawBench, innovatively explore the impact of multimodal data, and achieve state-of-the-art results in multiple designed experiments. These results indicate that our framework is effective not only for LJP but also for a broader range of legal applications, offering new perspectives for the development of future legal methods and datasets.

1 Introduction

The legal field has emerged as a particularly promising area for the application of natural language processing (NLP) techniques, among which Legal Judgment Prediction (LJP) (Aletras et al., 2016; Xu et al., 2020; Malik et al., 2021; Feng et al., 2022; Zhao et al., 2022) stands out as a crucial and highly significant task. LJP refers to the task of forecasting court outcomes based on the facts of a legal case, including elements such as the applicable law article, charge, and term of penalty. LJP can not only enhance the efficiency of judicial processes, but also offer valuable legal advice to the general public. Various methods have been proposed and achieved promising results with the application of state-of-the-art Neural Networks (Zhong et al., 2018; Yang et al., 2019; Yue et al., 2021; Zhang et al., 2023) and Large Language Models (LLMs) (He et al., 2024; Wang et al., 2024; Sun et al., 2024; Jiang and Yang, 2024; Chen et al., 2024a).

However, the unique characteristics of legal language, including its high degree of specialization, complexity, and reliance on precise terminology, pose significant challenges for LJP systems. These challenges are further compounded by the growing global population, which has led to an increase in the number of judicial cases, placing immense pressure on the legal systems. Efforts to mitigate these problems have focused on simplifying the simulation of real courtrooms (Jiang and Yang, 2024; Chen et al., 2024a), utilizing Retrieval-Augmented Generation (RAG) techniques (He et al., 2024; Wang et al., 2024), and refining prompt engineering (Sun et al., 2024). However, several challenges persist in the Legal Judgment Prediction (LJP) task:

1) *Reproduction of Real-World Scenarios.* Each legal case may involve one or more charges and statutes. However, most existing legal frameworks struggle to accurately predict multiple charges and relevant laws, making it difficult to forecast the final judicial outcomes. In addition, the existing legal framework is largely centered on supporting text-based input. However, information from other modalities, such as case-related evidence presented in court and trial recordings, also plays a vital role in LJP tasks.

2) *Multivariable Interactions.* Current legal frameworks often rely on straightforward courtroom simulations, where each agent is assigned a

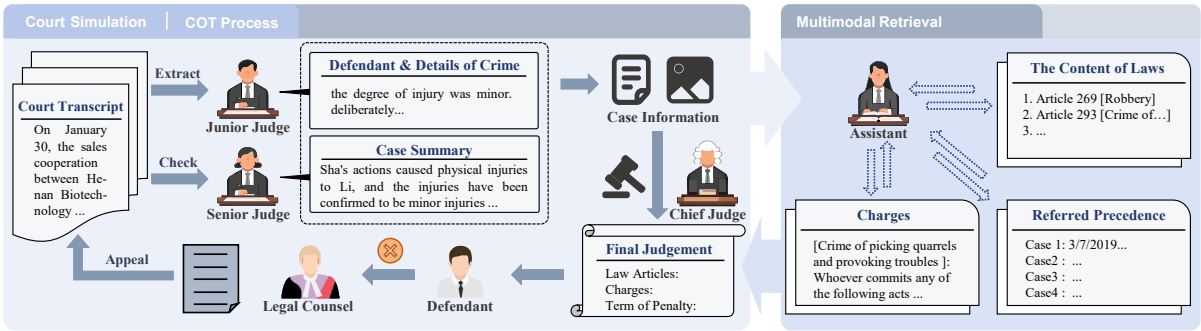


Figure 1: Overview of our framework *JurisAgents*.

role based on real-world scenarios. However, each role in the courtroom typically assumes multiple functions. If a simulated court is built solely based on existing roles while neglecting an analysis of their shared functions and failing to implement a well-structured and procedural design for the trial, most intelligent agents will struggle to grasp the key points of the case. As a result, their conclusions are often one-sided and superficial, making it difficult for them to make sound decisions across various stages. This, in turn, can lead to chaotic and illogical trial outcomes, ultimately undermining the goal of optimizing the entire trial process.

To address the aforementioned challenges, we propose a novel multi-agent framework *JurisAgents* for LJP task. As illustrated in Figure 1, *JurisAgents* leverages the advanced reasoning abilities of LLMs to emulate court proceedings and deliver high-quality outcomes.

Specifically, *JurisAgents* consists of six intelligent agents. Among them, five agents form the courtroom debate module, which adheres to the conventional trial process while detailing the proceedings based on specific tasks. We have established three Judges: Junior Judge, Senior Judge and Chief Judge. The Junior Judge and Senior Judge hold pivotal responsibilities in overseeing the case. Their duties include systematically structuring case information, distilling critical events, and guaranteeing a thorough review of all pertinent details. Ultimately, the Chief Judge is responsible for delivering the final verdict. If the Defendant is dissatisfied with the verdict, they can file an appeal, assisted by the Legal Counsel, for a secondary petition, after which the court conducts a second trial. We also developed a legal knowledge retrieval module to support the Chief Judge in the decision-making process. This module features an intelligent Assistant that provides access to the latest case rulings, updated criminal law documents, charge files, and relevant legal provisions.

Moreover, we introduce a legal dataset *JurisMM*. Given the dynamic and evolving nature of legal systems, existing datasets utilized in previous tasks often include outdated information that no longer aligns with the latest legal standards. Our dataset includes over 50,000 up-to-date real-world trial cases, sourced from official repositories. Furthermore, to explore the multimodal capabilities of the proposed framework, the dataset also contains 83 multimodal samples.

We summarize our contributions as follows:

- *JurisAgents* is a new multi-agent framework capable of conducting comprehensive simulations of court trials based on the provided case information.
- We propose a novel data set comprising single-text modality and multimodal video-text data, consisting of more than 50,000 judicial cases based on the latest criminal law judgments.
- We conducted extensive experiments and ablation studies. The results demonstrate that our framework outperforms existing state-of-the-art methods in all aspects and achieves the best performance across three tasks.

2 The JurisAgents Framework

2.1 Framework Overview

JurisAgents integrates six distinct roles: Junior Judge (JJ), Senior Judge (SJ), Chief Judge (CJ), Assistant, Defendant, and Legal Counsel—-all driven by agents powered by LLM.

Judges play a key role in court trials. In *JurisAgents*, there are three judges at different levels who work collaboratively to perform the following two key tasks: 1) making factual findings: that is, the judge needs to extract, organize, and synthesize the subject matter as well as the specific details of

the case in order to identify the final legal facts; and 2) making a legal value judgement: based on the identification of the legal facts, they rely on the legal provisions, leverage their own legal reserves, and correctly apply the provisions of the law in order to make a final conviction and sentence.

To clearly demonstrate the complete case adjudication process, we assign each judge a specific role categorized into three types: Junior Judge, Senior Judge, and Chief Judge. Through differentiated prompt design schemes, each judge is allocated distinct task perspectives, which are integrated into our *JurisAgents* system as fixed functional components.

1) Junior Judge is primarily responsible for processing the foundational case data submitted by litigants or prosecutorial authorities to the judicial system. This role's core responsibility involves real-time extraction and structured processing of received judicial information to accurately identify the principal elements and factual details of the case. Specifically for the LJP task, this role extracts and integrates information related to the defendant's case, focusing on analyzing behavioral patterns of the defendant to provide critical textual input materials for subsequent legal reasoning.

2) Senior judge undertakes in-depth analysis of the structured data output by the Junior Judge and performs multidimensional comprehensive modeling of the case's legal facts based on existing records in the judicial database. Specifically, this role generates a global description of the case facts by integrating the original case text with the structured case text summarized by the junior judge, and thus achieves the automated extraction of key judicial elements and knowledge refinement.

3) Chief Judge performs multi-level evaluation of the information outputs from both the Junior Judge and Senior Judge, integrates supplementary knowledge bases provided by legal assistants, and applies rule-based legal reasoning engines to invoke relevant legal provisions for generating final judicial decisions. This role synthesizes multi-source input features to produce judicial decision texts containing applicable legal articles, charge classifications, and sentencing predictions.

In addition to the role of the judge, we have also established auxiliary role, Assistant, as well as adversarial roles such as Defendant and Legal Counsel. By analyzing, organizing, and reviewing cases from both auxiliary and adversarial perspectives, we aim to help the judge identify gaps and address any oversights.

Assistant is regarded as the "external brain" of judges, responsible for retrieving external information and knowledge relevant to the final case judgment. In the process of adjudication, the assistant undertakes to extract, sort out and summarize the legal facts finally determined after extracting, sorting out and summarizing the main body and specific details of the case, and collects external information (including legally the content of laws, charges and referred precedence) on the basis of the legal facts, so as to prepare for the final judgment of the chief judge.

Defendant , as the core litigation entity, exhibits behavioral characteristics, including judicial elements such as subjective intent and objective conduct, making basis for case adjudication. Upon termination of the first-instance litigation process, this litigation entity may, in accordance with statutory authority, initiate second-instance appeal proceedings by generating and submitting standardized appeal request documentation. It should be particularly noted that the defendant's procedural choices will significantly impact the iterative evolutionary path of the entire judicial case.

Legal Counsel , functioning as a professional support entity, provides critical decision support to the defendant by leveraging domain knowledge bases and judicial practice experience. Upon system detection of first-instance judgment status updates, this professional support entity initiates the appeal assistance workflow: generating optimized second-instance procedure recommendations, formulating legal document generation strategies, and ultimately producing appeal documentation packages that comply with judicial standards.

2.2 Process Design

JurisAgents comprehensively simulates the entire process through a structured six-stage pipeline, as shown in Figure 2.

Information Extraction, Stage I. This phase corresponds to the initiation of first-instance court proceedings. The initial input consists of a comprehensive description of the criminal case process. Two judicial agents (the Junior Judge and Senior Judge) each receive this identical input and collaboratively process the case materials. The Junior Judge first extracts the defendant's key factual elements from the input text, specifically "defendant identification", "crime details", and "criminal motive". Building upon this extracted information, the

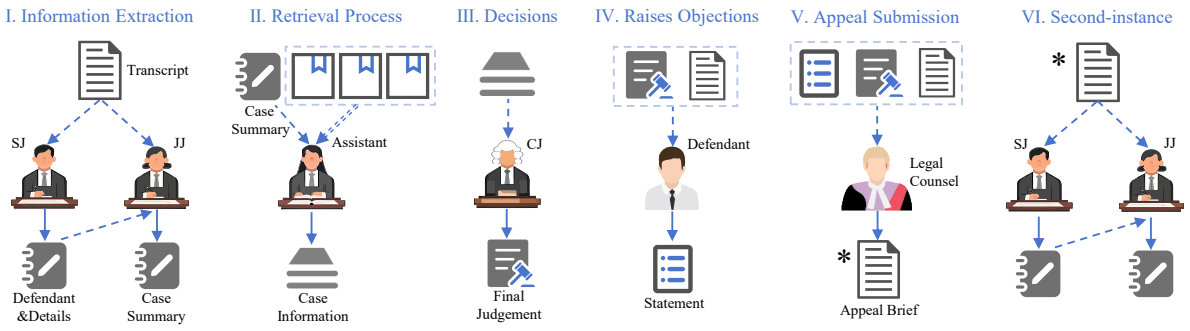


Figure 2: Our *JurisAgents* consists of six phases.

Senior Judge then conducts further analysis of the complete input text to synthesize a concise Case Summary document that captures the essential aspects of the full case.

Retrieval Process, Stage II. The Case Summary generated by the Senior Judge is transferred to the Assistant Agent, which performs searches in three legal knowledge bases (The Content of Laws Database, Charges Database, and Referred Precedence Database). For the Content of Laws Database and Charges Database, a dense vector retrieval approach is employed, with the top 10 ranked results from each serving as references. The Referred Precedence Database utilizes a hybrid retrieval system combining sparse and dense methods, where the top 5 most relevant precedents are selected. The Assistant then consolidates these retrieval results into a unified Case Information Document, incorporating legal provisions, similar cases, and relevant charges.

Decisions, Stage III. The Chief Judge receives three critical documents: 1) the Case Summary from the Junior Judge, 2) the defendant's key factual elements from the Senior Judge, and 3) the Case Information Document from the Assistant. Through comprehensive evaluation of these materials, the Chief Judge conducts an integrated assessment of the case facts. By cross-referencing the retrieved external legal resources with the specific details of the current litigation, the Chief Judge formulates the first-instance judicial decision.The final judgment document is then issued, which formally declares: (i) the relevant articles of the cases, (ii) charges against the defendant, (iii) the term of penalty.

Defendant Raises Objections, Stage IV. Upon receiving the final judgment issued by the Chief Judge, the defendant may either accept the verdict (concluding the case) or exercise the right to appeal if perceiving judicial errors or unfairness. As illustrated in the framework, the defendant formally submits a Statement of Appeal, a legally binding document outlining the grounds for appeal and intention to challenge the verdict. This statement is prepared through systematic analysis of both the final judgment document and complete case proceedings, serving as the legal foundation for initiating appellate proceedings.

Submission Appeal, Stage V. In this stage, legal counsel plays a critical role. Based on the three key documents—the final judgment, the defendant's objection statement, and the original case summary—the legal counsel prepares an appeal brief. This brief synthesizes legal arguments, cited laws and procedural concerns, and references case details extracted in earlier phases, particularly from the transcripts and case summaries handled by JJ and SJ.

Second-instance, Stage VI. Once the appeal brief is submitted, the case enters the second-instance phase. Both the second-instance Senior Judge and the Junior Judge revisit the original transcripts and case summaries, the same documents used in the initial trial. They re-assess the case by cross-referencing the appeal materials with the original evidence and judicial reasoning. This phase is crucial, as it ensures fairness through a comprehensive review. The outcome of this second-instance trial is final, marking the end of the judicial process in a two-instance trial system.

3 The JurisMM Dataset

3.1 Data Collection

Unimodal Text Data. Previous work, such as the CAIL2018 dataset (Xiao et al., 2018), has collected millions of LJP texts. However, due to the real-time nature and frequent updates of criminal law, as well as the rapid development of LLMs, relying on outdated data that references superseded legal statutes

may negatively impact the progress of LJP tasks. To address this issue and ensure the relevance of our experiments, we follow the data collection methods established in prior work to gather the most recent legal documents. This updated dataset supports the experimental validation and further exploration of the framework proposed in this study. We collected criminal judgment documents publicly available on China Judgments Online from 2014 to 2025, amassing approximately 150,000 cases covering all crime types listed in the criminal offense catalog. Preliminary statistics revealed a significant imbalance in the number of cases across different offense categories. To this end, we established an upper limit for the number of cases per offense category: if any category exceeded 2,000 cases, we retained only the most recent 2,000 cases based on judgment date. Concurrently, we implemented a lower limit: if any offense category contained fewer than 100 cases, we removed all cases associated with that offense and excluded the offense from the criminal offense catalog.

Building upon this filtered case repository, we conducted further refinement by eliminating judgments that were inconsistent with the latest provisions of the Criminal Law of the People's Republic of China (2023 Amendment) to ensure legal timeliness of all retained cases. A total of 53,415 data entries were obtained, with 46,365 used as the training set and 7,050 as the test set. For detailed data collection process, see the appendix A.1.

Multimodal Data. Additionally, to address the gap in multimodal data for legal judgment tasks, we collected judicial documents along with their corresponding real legal case videos, ultimately constructing a suitable multimodal legal dataset consisting of both video and text.

We propose a collaborative framework led by legal experts and supported by multimodal large language models. The process begins with extracting key frames from video footage, which are then converted into a more standardized image format. Considering the current challenges in precisely aligning legal judgments with visual evidence, legal experts must semi-automatically select a small set of key frames from relevant videos for each specific case.

As a result, we obtained 83 high-quality multimodal data samples. These multimodal criminal law cases are specifically designed to support the training and evaluation of multimodal judicial decision-making tasks.

3.2 Legal Knowledge Base.

To ensure judicial accuracy and consistency, judges must rely on a comprehensive legal knowledge foundation. Based on the Criminal Law of the People's Republic of China (2023 Amendment), we have constructed a large-scale integrated legal knowledge base comprising three core components: 1) The content of laws with Judicial Interpretations, 2) List of charges, and 3) Repository of referred precedence. All knowledge elements in this legal knowledge base are stored and indexed in a Milvus vector database for efficient retrieval and processing. More details can be found in Appendix B.1.

The Content of Laws with Judicial Interpretations. We obtained legal documents from the authoritative legal information repository. From this database, we filtered all currently effective criminal law provisions, removed repealed clauses, and eliminated articles irrelevant to criminal judgments, ultimately retaining a total of 438 provisions. Additionally, to enhance the interpretability of legal provisions and the precision of legal application, we supplemented the dataset with supporting judicial interpretations issued by the Supreme People's Court.

List of Charges. Based on the Criminal Law of the People's Republic of China, we systematically organized major charges to form a structured inventory. This inventory serves as an independent reference standard that can support downstream tasks including charge prediction, case screening, and legal retrieval.

Repository of Referred Precedence. To facilitate legal reasoning based on case similarity, we incorporated 46,365 cases from the test set into the knowledge base, which serves as authoritative references for judicial decision-making. All selected precedents maintain legal representativeness, thereby ensuring that the referenced cases in judicial proceedings are both authoritative and timely.

4 Experiments

4.1 Experimental Settings

All baselines are categorized into two types: Neural-Networks-based and LLM-based approaches.

1) For the traditional Neural-Network-based methods, training is performed using the training set from our *JurisMM-Text* dataset, while evaluation is conducted on the test set. Additional experimental details can be found in Appendix C.

Model	Law Articles				Charges				Terms of Penalty				
	Acc.	MP	MR	MF	Acc.	MP	MR	MF	Acc.	MP	MR	MF	N-Id
TextCNN	0.703	0.389	0.343	0.334	0.696	0.492	0.457	0.441	0.325	0.321	0.282	0.285	—
TOPJUDGE	0.768	0.455	0.411	0.400	0.751	0.516	0.519	0.491	0.373	0.358	0.364	0.342	—
MPBFN	0.755	0.467	0.391	0.391	0.707	0.427	0.416	0.387	0.324	0.343	0.267	0.263	—
GLM-4V-9B	0.095	0.019	0.027	0.015	0.476	0.169	0.120	0.119	0.199	0.320	0.197	0.163	0.668
mPLUG-7B	0.094	0.027	0.035	0.019	0.319	0.093	0.060	0.058	0.100	0.250	0.168	0.067	0.507
Qwen2.5-VL-3B	0.220	0.089	0.072	0.057	0.258	0.109	0.068	0.068	0.193	0.169	0.122	0.100	0.787
Qwen2.5-VL-7B	0.271	0.133	0.121	0.092	0.134	0.061	0.021	0.026	0.188	0.135	0.119	0.096	0.799
Qwen2.5-7B	0.523	0.133	0.170	0.123	0.500	0.161	0.144	0.134	0.237	0.227	0.145	0.124	0.826
GPT-4o	0.380	0.264	0.207	0.186	0.436	0.253	0.186	0.182	0.372	0.432	0.357	0.319	0.842
JurisAgents	**0.872**	**0.604**	**0.645**	**0.602**	**0.882**	**0.600**	**0.642**	**0.603**	**0.409**	**0.520**	**0.393**	**0.361**	**0.848**
w/o KB	0.653	0.323	0.323	0.281	0.801	0.203	0.190	0.187	0.378	0.454	0.371	0.335	**0.848**
w/o MA	0.767	0.592	0.607	0.569	0.773	0.581	0.599	0.566	0.303	0.475	0.304	0.272	0.803

Table 1: Overall performance on *JurisMM-Text*.

2) For the LLM-based approaches, experiments involving closed-source models are conducted with the temperature parameter set to 0 to ensure deterministic outputs. For open-source models, each experiment is repeated three times, and the average performance is reported. All the multi-agent systems in our framework are powered by GPT-4o.

To compare the performance of the baselines and our methods, we choose four metrics that are widely used for multi-classification tasks, including accuracy (Acc.), macro-precision (MP), macro-recall (MR), and macro-F1 (F1). It is worth noting that for the sentence length prediction task, which involves predicting continuous time durations, traditional classification metrics may not fully capture the model's performance. Therefore, in addition to the above metrics, we also refer to the Normalized Log-Distance (N-Ld) metric used in the LawBench (Fei et al., 2023) benchmark. This metric allows for a more nuanced evaluation by measuring the normalized logarithmic difference between predicted and true values, thus providing a complementary perspective on model accuracy in regression-like settings.

4.2 Baselines

To extensively validate the effectiveness of the proposed framework, we employed the following baselines based on traditional neural networks for comparison. These include TextCNN, TOPJUDGE (Zhong et al., 2018), and MPBFN (Yang et al., 2019).

We then compare several high-performing LLM models. Among the open-source models, we include GLM-4V-9B (GLM et al., 2024), mPLUG-Owl-7B (Ye et al., 2024), Qwen2.5-VL-3B/7B (Bai et al., 2025), and Qwen2.5-7B-Instruct (Qwen et al., 2025). These models are notable for their accessibility, allowing researchers and developers to inspect, modify, and fine-tune them according to specific needs, which fosters innovation and transparency in the field. For closed-source models, we utilize GPT-4o (OpenAI, 2024).

4.3 Overall Performance

Table 1 shows the comprehensive performance of each model on the three tasks of Law Articles, Charges and Terms of Penalty. It can be seen that our proposed framework *JurisAgents* has achieved a significant lead in all tasks and indicators, fully reflecting its advantages in legal text understanding and reasoning.

The performance of *JurisAgents* on multiple evaluation is also very balanced and excellent, showing that the model achieves a good balance between accuracy and recall, avoiding the bias of a single indicator. This balance performance is particularly important for the legal judgment task, because the diversity and complexity of legal texts require that the model neither omits key provisions nor produces too many misjudgments.

In the sentence prediction task, although the overall accuracy was relatively low, *JurisAgents* still outperformed other models, indicating that the task was more difficult, possibly due to the finer grain of inference and quantitative judgment involved in the sentence.

Comparison between NN and LLMs. The results reveal a clear distinction in performance between traditional Neural Network based (NN) models and LLMs on the *JurisMM-Text* dataset. Classical neural network architectures consistently outperform the LLMs across all three legal tasks. This result is consistent with our expectations. The reason is that

traditional neural network models are specifically designed for legal text classification, with both feature extraction and architecture reflecting legal domain rules and expert prior knowledge. In contrast, the large language models evaluated, despite having a vast number of parameters and strong general language understanding, may lack sufficient domain adaptation and detailed legal knowledge, which are essential for accurate LJP.

Performance across Different Legal Tasks. Generally, all models achieve their highest performance on law article prediction, followed by charge prediction, with terms of penalty being the most challenging task. This pattern aligns with the inherent complexity of each task: law articles and charges are often more directly linked to textual cues and legal definitions, whereas penalty terms require nuanced understanding of sentencing guidelines, contextual factors, and legal discretion.

Our framework stands out by achieving the best results across all tasks and metrics, indicating its superior capability in capturing the multifaceted nature of LJP. Notably, its advantage is most pronounced in charge prediction, where it surpasses other models by a significant margin. This suggests that *JurisAgents* effectively integrates legal knowledge and reasoning mechanisms that are crucial for accurately identifying charges, which often involve complex legal interpretations.

In terms of penalty term prediction, although all models show relatively lower scores, *JurisAgents* still leads, demonstrating its ability to handle the most challenging aspect of legal judgment. The relatively lower performance across models in this task highlights the need for further research into modeling sentencing decisions, which may depend on subtle contextual and procedural factors beyond textual information.

Ablation Studies. To thoroughly evaluate the contributions of individual components in *JurisAgents*, we conducted ablation experiments on the test set.

Removing the knowledge base (w/o KB) results in a substantial drop in performance across all tasks, especially in law article and charge prediction. This underscores the critical role of external legal knowledge in enhancing the model's reasoning and decision-making capabilities. Without access to structured legal information, *JurisAgents* struggles to accurately map textual inputs to relevant legal concepts and outcomes.

Similarly, excluding the multi-agent collaboration mechanism (w/o MA) leads to a noticeable

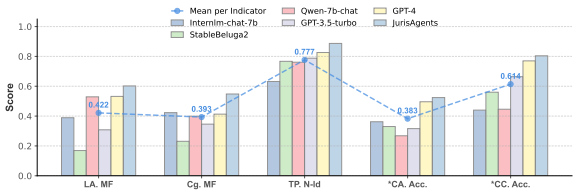


Figure 3: Overall performance on LawBench.

decline, particularly in the terms of penalty task. This indicates that the collaborative interactions among multiple agents within the framework enable the model to incorporate diverse perspectives and legal elements, facilitating more comprehensive and detailed analysis, thereby enhancing its overall robustness. This is crucial for complex decision-making tasks, such as sentencing.

Interestingly, even with the removal of any single component, *JurisAgents* still outperforms many baseline models. This phenomenon not only confirms the reliability of our design mechanisms but also demonstrates the robustness of the *JurisAgents* architecture.

4.4 Expand Experiments

Task Extension. To evaluate the effectiveness of *JurisAgents* across varying data distributions and assess its suitability for legal judgment tasks, we conducted a comparative analysis against several large language models using the LawBench benchmark. LawBench is a comprehensive evaluation suite designed to measure the legal reasoning capabilities of LLMs within the context of the Chinese legal system. It encompasses a wide range of law-related tasks. For our experiments, we selected five representative tasks from the benchmark—three that align directly with the task types *JurisAgents* is designed to handle, and two additional tasks, Case Analysis (CA.) and Crime Amount Calculation (CC.), which are also prevalent in legal practice.

The results, as shown in Table 3, demonstrate that *JurisAgents* achieved the best performance across all five legal knowledge application tasks compared to five other general-purpose large language models. Specifically, in the three fundamental legal judgment tasks, namely, the prediction of law articles, the prediction of charges, and the prediction of prison terms, *JurisAgents* achieved improvements of 7.0%, 12.5%, and 6.1%, respectively. Furthermore, in the two extended tasks, case analysis and crimal damages calculation, our framework yielded improvements of 2.8% and 3.4%, respec-

Model	Law Articles		Charges		Terms of Penalty	
	Acc.	MF	Acc.	MF	N-Id	MF
GPT-4o	0.518	0.165	0.325	0.183	0.724	0.114
w picture	0.518	0.180	0.337	0.177	0.779	0.164
JurisAgents	0.614	0.334	0.518	0.309	0.744	0.105
w picture	**0.639**	**0.336**	**0.554**	**0.353**	**0.816**	**0.184**

Table 2: Overall performance on *JurisMM-Video*.

tively. These results indicate that a multi-agent legal system, refined through task specification and customization, exhibits stronger adaptability to downstream legal application tasks, enabling rapid adjustment of execution goals for various roles to meet diverse legal task requirements.

Modal Extension. We conducted experiments on the collected set of 83 multimodal legal judgment cases, with the results presented in Table 2. Our framework *JurisAgents* demonstrates consistent improvements across all three prediction tasks when incorporating visual information, with accuracy and macro-F1 scores showing notable enhancement. Particularly, the sentence prediction accuracy achieves an 8.5% absolute improvement over text-only conditions. The integration of multimodal data significantly enhances *JurisAgents*' performance in legal judgment prediction. These experimental results indicate that visual information enables the model to capture more comprehensive contextual understanding of cases, thereby improving both the accuracy and robustness of legal provision, charge, and sentence prediction.

5 Related Work

Legal Judgment Prediction. The task of legal judgment prediction (LJP) has undergone several significant methodological transitions. In its early stages, LJP was primarily addressed using mathematical and statistical approaches, relying on existing legal cases to predict the outcomes of specific scenarios (Ruger et al., 2004). These methods often required extensive manual feature engineering. With the rapid advancement of neural networks, researchers have proposed a wide range of models to support legal case analysis (Zhong et al., 2018; Yang et al., 2019).However, due to the rigorousness and fairness required in legal formulation, it remains challenging in practice to distinguish between charges and conviction boundaries described by similar statutes.Some studies leverage contrastive learning to capture subtle distinctions between similar law articles or charges (Gan et al., 2023; Zhang et al., 2023; Liu et al., 2023). With the rapid advancement of LLM-based autonomous agents, their growing capability to make sophisticated decisions in real-world scenarios offers novel perspectives and opportunities for advancing legal judgment prediction. Some recent studies have begun to explore the application of novel LLM-based methods to the LJP task. (He et al., 2024; Wang et al., 2024; Sun et al., 2024; Jiang and Yang, 2024; Chen et al., 2024a).

Multi-agent framework. As the reasoning and planning capabilities of large language models continue to advance, LLM-based agents are also attracting increasing attention. These agents are capable of perceiving environments, making decisions, and taking actions (Guo et al., 2023; Liang et al., 2023; Xi et al., 2023). In complex scenarios, single agents often struggle with intricate task logic, whereas collaborative multi-agent systems can enhance task efficiency. Leveraging the collective expertise of specialized agents has shown promise in tackling challenging problems such as scientific research (Zheng et al., 2023; Chen et al., 2024b), scientific debate (Liang et al., 2024; Park et al., 2024; Du et al., 2024; Xiong et al., 2023), software development (Li et al., 2023; Dong et al., 2024; Qian et al., 2024; Hong et al., 2023; Huang et al., 2023), and so on. In the legal domain, while initial attempts have been made to apply agent-based methods, performance improvements remain limited.

6 Conclusion

We propose *JurisAgents*, a novel multi-agent framework designed to address the inherent complexities of LJP by decomposing trial tasks into structured stages and simulating realistic courtroom procedures. By leveraging the reasoning capabilities of LLMs within a carefully orchestrated multi-agent system, our framework effectively captures the multivariable interactions and procedural nuances that characterize real-world legal cases. To further support this endeavor, we introduced *JurisMM*, a large-scale, up-to-date dataset comprising over 50,000 judicial cases, including both unimodal textual data and multimodal video-text samples.

Overall, our work contributes a robust framework and valuable resources that advance the state of the art in LJP and offer promising directions for future research in legal AI, including more realistic courtroom simulations, multimodal reasoning, and dynamic knowledge integration.

Limitations

First, our exploration of multimodal data in the legal domain remains preliminary. Due to the lack of effective automated methods for constructing large-scale multimodal datasets, the multimodal portion of our dataset—comprising video and text data—was manually curated by legal experts. This manual collection process inherently limits the scale and diversity of multimodal samples, restricting the generalizability and robustness of multimodal learning in legal judgment prediction. Future work should focus on developing scalable and automated approaches for multimodal data acquisition and annotation to better support comprehensive legal AI systems.

Second, our current framework and dataset are primarily centered on criminal law cases derived from Chinese judicial proceedings. Legal systems vary significantly across countries in terms of procedural rules, legal codes, and trial practices. Therefore, extending our framework to accommodate multiple legal systems and constructing more universal datasets will be essential to enhance the adaptability and global relevance of our methods.

References

Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. 2016. Predicting judicial decisions of the european court of human rights: A natural language processing perspective. *PeerJ Computer Science*, 2:99–110.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. 2025. Qwen2.5-vl technical report.

Guhong Chen, Liyang Fan, Zihan Gong, Nan Xie, Zixuan Li, Ziqiang Liu, Chengming Li, Qiang Qu, Shiwen Ni, and Min Yang. 2024a. Agentcourt: Simulating court with adversarial evolvable lawyer agents.

Pei Chen, Shuai Zhang, and Boran Han. 2024b. CoMM: Collaborative multi-agent, multi-reasoning-path prompting for complex problem solving. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 1720–1738, Mexico City, Mexico. Association for Computational Linguistics.

Yihong Dong, Xue Jiang, Zhi Jin, and Ge Li. 2024. Self-collaboration code generation via chatgpt. *ACM Transactions on Software Engineering and Methodology*, 33(7):1–38.

Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. 2024. Improving factuality and reasoning in language models through multiagent debate. In *Forty-first International Conference on Machine Learning*.

Zhiwei Fei, Xiaoyu Shen, Dawei Zhu, Fengzhe Zhou, Zhuo Han, Songyang Zhang, Kai Chen, Zongwen Shen, and Jidong Ge. 2023. Lawbench: Benchmarking legal knowledge of large language models. *arXiv preprint arXiv:2309.16289*.

Yi Feng, Chuanyi Li, and Vincent Ng. 2022. Legal judgment prediction via event extraction with constraints. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 648–664, Dublin, Ireland. Association for Computational Linguistics.

Leilei Gan, Baokui Li, Kun Kuang, Yating Zhang, Lei Wang, Anh Luu, Yi Yang, and Fei Wu. 2023. Exploiting contrastive learning and numerical evidence for confusing legal judgment prediction. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 12174–12185, Singapore. Association for Computational Linguistics.

Team GLM, :, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu Feng, Hanlin Zhao, Hanyu Lai, Hao Yu, Hongning Wang, Jiadai Sun, Jiajie Zhang, Jiale Cheng, Jiayi Gui, Jie Tang, Jing Zhang, Jingyu Sun, Juanzi Li, Lei Zhao, Lindong Wu, Lucen Zhong, Mingdao Liu, Minlie Huang, Peng Zhang, Qinkai Zheng, Rui Lu, Shuaiqi Duan, Shudan Zhang, Shulin Cao, Shuxun Yang, Weng Lam Tam, Wenyi Zhao, Xiao Liu, Xiao Xia, Xiaohan Zhang, Xiaotao Gu, Xin Lv, Xinghan Liu, Xinyi Liu, Xinyue Yang, Xixuan Song, Xunkai Zhang, Yifan An, Yifan Xu, Yilin Niu, Yuantao Yang, Yueyan Li, Yushi Bai, Yuxiao Dong, Zehan Qi, Zhaoyu Wang, Zhen Yang, Zhengxiao Du, Zhenyu Hou, and Zihan Wang. 2024. Chatglm: A family of large language models from glm-130b to glm-4 all tools.

Taicheng Guo, Kehan Guo, Bozhao Nan, Zhenwen Liang, Zhichun Guo, Nitesh V. Chawla, Olaf Wiest, and Xiangliang Zhang. 2023. What can large language models do in chemistry? a comprehensive benchmark on eight tasks.

Zhitao He, Pengfei Cao, Chenhao Wang, Zhuoran Jin, Yubo Chen, Jiexin Xu, Huaijun Li, Kang Liu, and Jun Zhao. 2024. AgentsCourt: Building judicial decision-making agents with court debate simulation and legal knowledge augmentation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 9399–9416, Miami, Florida, USA. Association for Computational Linguistics.

Sirui Hong, Xiawu Zheng, Jonathan Chen, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven

Ka Shing Yau, Zijuan Lin, Liyang Zhou, et al. 2023. Metagpt: Meta programming for multi-agent collaborative framework. *arXiv preprint arXiv:2308.00352*, 3(4):6.

Dong Huang, Jie M Zhang, Michael Luck, Qingwen Bu, Yuhao Qing, and Heming Cui. 2023. Agentcoder: Multi-agent-based code generation with iterative testing and optimisation. *arXiv preprint arXiv:2312.13010*.

Cong Jiang and Xiaolei Yang. 2024. Agents on the bench: Large language model based multi agent framework for trustworthy digital justice. *ArXiv*, abs/2412.18697.

Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. 2023. Camel: Communicative agents for" mind" exploration of large scale language model society.

Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. 2024. Encouraging divergent thinking in large language models through multi-agent debate. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17889–17904, Miami, Florida, USA. Association for Computational Linguistics.

Zhenwen Liang, Wenhao Yu, Tanmay Rajpurohit, Peter Clark, Xiangliang Zhang, and Ashwin Kalyan. 2023. Let GPT be a math tutor: Teaching math word problem solvers with customized exercise generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 14384–14396, Singapore. Association for Computational Linguistics.

Yifei Liu, Yiquan Wu, Yating Zhang, Changlong Sun, Weiming Lu, Fei Wu, and Kun Kuang. 2023. Mlljp: Multi-law aware legal judgment prediction. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '23, page 1023–1034, New York, NY, USA. Association for Computing Machinery.

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 *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4046–4062, Online. Association for Computational Linguistics.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

OpenAI. 2024. Gpt-4o system card.

Someen Park, Jaehoon Kim, Seungwan Jin, Sohyun Park, and Kyungsik Han. 2024. PREDICT: Multi-agent-based debate simulation for generalized hate speech detection. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 20963–20987, Miami, Florida, USA. Association for Computational Linguistics.

Chen Qian, Wei Liu, Hongzhang Liu, Nuo Chen, Yufan Dang, Jiahao Li, Cheng Yang, Weize Chen, Yusheng Su, Xin Cong, Juyuan Xu, Dahai Li, Zhiyuan Liu, and Maosong Sun. 2024. ChatDev: Communicative agents for software development. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15174–15186, Bangkok, Thailand. Association for Computational Linguistics.

Qwen, :, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. 2025. Qwen2.5 technical report.

Theodore W Ruger, Pauline T Kim, Andrew D Martin, and Kevin M Quinn. 2004. The supreme court forecasting project: legal and political science approaches to predicting supreme court decisionmaking. *Colum. L. Rev.*, 104:1150.

Jingyun Sun, Shaobin Huang, and Chi Wei. 2024. Chinese legal judgment prediction via knowledgeable prompt learning. *Expert Systems with Applications*, 238:122177.

Maosong Sun, Xinxiong Chen, Kaixu Zhang, Zhipeng Guo, and Zhiyuan Liu. 2016. Thulac: An efficient lexical analyzer for chinese. Technical report, Technical Report.

Xuran Wang, Xinguang Zhang, Vanessa Hoo, Zhouhang Shao, and Xuguang Zhang. 2024. Legalreasoner: A multi-stage framework for legal judgment prediction via large language models and knowledge integration. *IEEE Access*, 12:166843–166854.

Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Yuhao Zhou, Weiran Wang, Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zhangyue Yin, Shihan Dou, Rongxiang Weng, Wensen Cheng, Qi Zhang, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjing Huang, and Tao Gui. 2023. The rise and potential of large language model based agents: A survey.

Chaojun Xiao, Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Yansong Feng, Xianpei

Han, Zhen Hu, Heng Wang, and Jianfeng Xu. 2018. CAIL2018: A large-scale legal dataset for judgment prediction. *CoRR*, abs/1807.02478.

Kai Xiong, Xiao Ding, Yixin Cao, Ting Liu, and Bing Qin. 2023. Examining inter-consistency of large language models collaboration: An in-depth analysis via debate. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 7572–7590, Singapore. Association for Computational Linguistics.

Nuo Xu, Pinghui Wang, Long Chen, Li Pan, Xiaoyan Wang, and Junzhou Zhao. 2020. Distinguish confusing law articles for legal judgment prediction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3086–3095, Online. Association for Computational Linguistics.

Wenmian Yang, Weijia Jia, Xiaojie Zhou, and Yutao Luo. 2019. Legal judgment prediction via multi-perspective bi-feedback network. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 4085–4091. International Joint Conferences on Artificial Intelligence Organization.

Qinghao Ye, Haiyang Xu, Guohai Xu, Jiabo Ye, Ming Yan, Yiyang Zhou, Junyang Wang, Anwen Hu, Pengcheng Shi, Yaya Shi, Chenliang Li, Yuanhong Xu, Hehong Chen, Junfeng Tian, Qi Qian, Ji Zhang, Fei Huang, and Jingren Zhou. 2024. mplug-owl: Modularization empowers large language models with multimodality.

Linan Yue, Qi Liu, Binbin Jin, Han Wu, Kai Zhang, Yanqing An, Mingyue Cheng, Biao Yin, and Dayong Wu. 2021. Neurjudge: A circumstance-aware neural framework for legal judgment prediction. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 973–982. ACM.

Han Zhang, Zhicheng Dou, Yutao Zhu, and Ji-Rong Wen. 2023. Contrastive learning for legal judgment prediction. *ACM Trans. Inf. Syst.*, 41(4).

Jie Zhao, Ziyu Guan, Cai Xu, Wei Zhao, and Enze Chen. 2022. Charge prediction by constitutive elements matching of crimes. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pages 4517–4523.

Zhiling Zheng, Oufan Zhang, Ha L Nguyen, Nakul Rampal, Ali H Alawadhi, Zichao Rong, Teresa Head-Gordon, Christian Borgs, Jennifer T Chayes, and Omar M Yaghi. 2023. Chatgpt research group for optimizing the crystallinity of mofs and cofs. *ACS Central Science*, 9(11):2161–2170.

Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018. Legal judgment prediction via topological learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3540–3549, Brussels, Belgium. Association for Computational Linguistics.

A JurisMM Details

A.1 Date Collection Process

Building upon this filtered case repository, we conducted further refinement by eliminating judgments that were inconsistent with the latest provisions of the Criminal Law of the People's Republic of China (2023 Amendment) to ensure legal timeliness of all retained cases.

We implemented a phased data screening methodology with the following procedure: First, we performed a systematic comparison between the 2023 version (effective from March 1, 2024) and four previous editions of the Criminal Law implemented during 2014-2025. This comparative analysis identified all modified legal articles between the 2023 version and each preceding edition. Subsequently, we examined cases published during each version's effective period and removed those involving any modified legal provisions.For concrete illustration, taking the 2020 edition as an example: our comparison with the 2023 version revealed seven modified articles. Since the 2020 edition was effective from March 1, 2021 to March 1, 2024, we accordingly removed all cases involving these seven modified articles that were published during this period.

A.2 Statistics

We conducted a comprehensive statistical analysis of the charge categories, including the overall quantitative distribution and frequency statistics in the test set, with the detailed distribution characteristics presented in Table 3 and Figure 4.

Category	Total Kinds	Average
The Content of Laws	132	921.70
Charges	115	907.18
Terms of Penalty	-	23.05

Table 3: Summary of Relevant Articles, charges, and Terms of Penalty.

B JurisAgents Details

B.1 The Details of Legal Knowledge Base

For the retrieval process, we implemented a hybrid approach that combines:

- **Dense retrieval**: Captures semantic information through vector embeddings using inner product similarity metrics
- **Sparse retrieval**: Employs keyword matching based on BM25 scoring

The final results primarily utilize dense retrieval outputs, with sparse retrieval results serving as supplementary information, followed by reranking through reciprocal rank fusion.

C Experimental Setup

All baselines are categorized into two types: Neural-Networks-based and LLM-based approaches.

1) For the Neural-Network-based methods, since our *JurisMM* dataset consists of continuous Chinese text without explicit word boundaries, we first apply THULAC (Sun et al., 2016) to perform word segmentation on the factual descriptions of the cases. After segmentation, we pretrain word embeddings using word2vec (Mikolov et al., 2013) with an embedding dimension of 200.

We implement the NN-based methods using PyTorch. In the CNN-based baseline models, the maximum document length is configured to 512 tokens. For the LSTM-based baselines, we define the maximum sentence length as 128 tokens, with each document limited to 32 sentences.

The hidden state dimension d_s is set to 256. For the CNN encoder, the number of filters d_c is set to 256, and the window sizes h are set to 2,3,4,5 with 64 filters for each window size. During the training part, we use the Adam optimizer with a learning rate of 10^{-3} and a dropout rate of 0.5. All models are trained for 16 epochs on an RTX 4090 GPU with a batch size of 512.Then we evaluate the final model on the testing set.

2) For the LLM-based approaches, experiments involving closed-source models are conducted with the temperature parameter set to 0 to ensure deterministic outputs. For open-source models, each experiment is repeated three times, and the average performance is reported.

Table 5 presents the detailed experimental results corresponding to Figure 3.

D Prompts

As shown in the following Tables, we present the specific prompts corresponding to each individual role.

Charge	Count	Charge	Count	Charge	Count
Abandonment	12	Abuse	1	Abuse of Power	142
Accepting Bribes	65	Accident Involving Dangerous Substances	1	Arson	179
Assaulting Police	9	Assisting Criminals to Evade Punishment	25	Bid Rigging	10
Bigamy	15	Bill Fraud	6	Bribery	149
Bribery by Leveraging Influence	7	Bribery of Corporations	2	Bribery of Non-State Personnel	62
Bribery to Non-State Personnel	2	Child Abduction	15	Child Molestation	78
Contract Fraud	137	Copyright Infringement	20	Corporate Acceptance of Bribes	4
Corporate Bribery	21	Corruption	117	Counterfeiting Currency	1
Counterfeiting Registered Trademark	78	Credit Card Fraud	106	Damaging Computer Information Systems	10
Damaging Flammable or Explosive Equipment	53	Damaging Power Equipment	45	Damaging Traffic Facilities	5
Damaging Transportation Vehicles	1	Dangerous Driving	35	Defamation	3
Dereliction of Duty	115	Disrupting Production and Operation	23	Disrupting Regulatory Order	6
Distributing Obscene Materials	1	Embezzlement of Funds	189	Embezzlement of Public Funds	158
Endangering Public Safety with Dangerous Means	85	Environmental Pollution	11	Escape	1
Excessive Logging	37	Explosion	4	Extortion	109
Facilitating Bribery	19	False Accusation	15	False Reporting of Registered Capital	3
Forced Labor	1	Forced Transactions	25	Forcing Others to Use Drugs	2
Fraud	130	Fundraising Fraud	52	Gambling	20
Gathering to Disturb Social Order	9	Gathering to Fight	121	Gathering to Loot	3
Harboring Drug Users	23	Harming Rare or Endangered Wildlife	9	Illegal Absorption of Public Deposits	145
Illegal Business Operation	161	Illegal Cultivation of Narcotic Plants	104	Illegal Detention	104
Illegal Drug Possession	474	Illegal Entry into Residence	311	Illegal Fishing of Aquatic Products	35
Illegal Loan Issuance	16	Illegal Logging	64	Illegal Medical Practice	30
Illegal Mining	63	Illegal Occupation of Agricultural Land	43	Illegal Organization of Blood Selling	1
Illegal Sale of Invoices	8	Impersonating Military Personnel for Fraud	36	Impersonation Fraud	54
Insult	9	Insurance Fraud	31	Intentional Destruction of Property	145
Intentional Homicide	360	Intentional Injury	94	Interfering with Credit Card Management	72
Issuing False Invoices	1	Kidnapping	40	Loan Fraud	31
Major Liability Accident	65	Misappropriation	11	Misappropriation of Designated Funds	1
Money Laundering	3	Negligent Homicide	537	Negligent Serious Injury	247
Obstructing Public Duties	80	Obstructing Testimony	12	Occupational Embezzlement	172
Operating a Gambling House	37	Organizing Prostitution	1	Perjury	2
Perverting the Course of Justice	4	Picking Quarrels and Provoking Trouble	33	Possession of Forged Invoices	24
Private Division of State Assets	11	Rape	14	Refusal to Pay Labor Remuneration	77
Releasing Dangerous Substances	35	Retaliating Against Witnesses	2	Robbery	41
Selling Counterfeit Trademark Goods	410	Smuggling Waste	3	Snatching	225
Spreading Sexually Transmitted Diseases	2	Tax Evasion	9	Teaching Criminal Methods	4
Theft	69	Throwing Objects from Heights	4	Traffic Accident	45
Usurious Relending	2				

Table 4: Distribution of Charges in test date of *JurisMM-Text* .

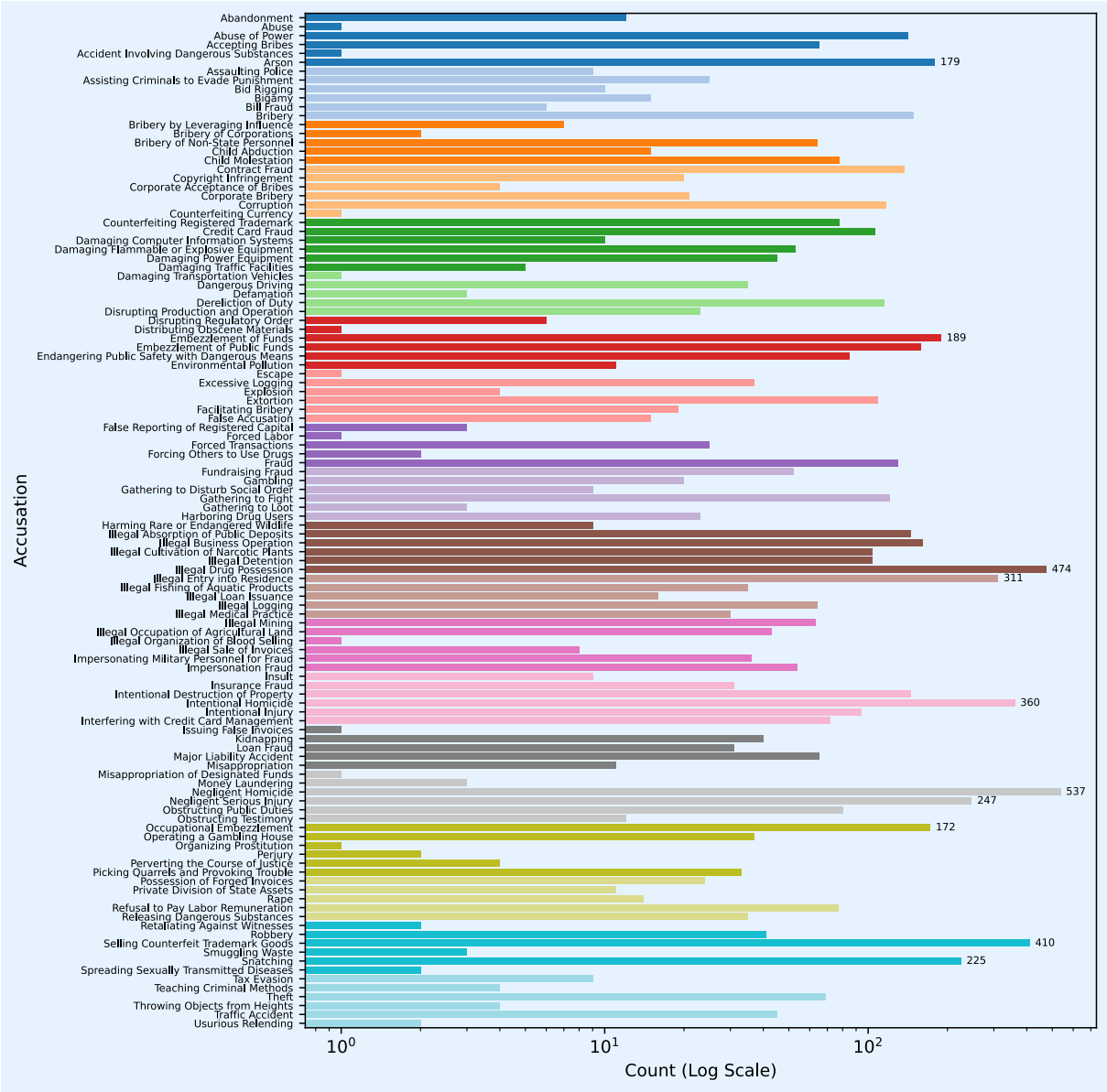


Figure 4: Distribution of Charges of *JurisMM-Text*.

Model	LA. MF	Cg. MF	TP. N-ld	*CA. Acc.	*CC. Acc.
Internlm-chat-7b	0.389	0.423	0.632	0.362	0.440
StableBeluga2	0.169	0.231	0.767	0.330	0.560
Qwen-7b-chat	0.529	0.399	0.761	0.268	0.446
GPT-3.5-turbo	0.308	0.346	0.788	0.316	0.664
GPT-4	0.532	0.413	0.826	0.496	0.770
JurisAgents	**0.602**	**0.548**	**0.887**	**0.524**	**0.804**

Table 5: Overall performance on LawBench.

E Cases

Chen Case of Illegal Hunting

Case Information

Cause of Action: Illegal Hunting
Case Type: Criminal Case
Defendant: Chen
Prosecution: Procuratorate

(1) Defendant Background:

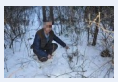
Defendant Chen Mouwen illegally hunted a protected roe deer in Jilin using snares on December 20, 2024. Seized items included hunting tools and the deer carcass (valued at RMB 3,000). Charged under Criminal Law Article 341(2), facing possible fines or imprisonment..

(2) Established Facts:

On December 20, 2024, defendant Chen Mouwen illegally placed over 10 snares in Sanchazi Forest Farm, Jilin, capturing one protected roe deer (valued at RMB 3,000). Police seized hunting tools (23 rabbit snares, 10 deer snares, 50 steel pellets, 1 slingshot) and the deer carcass. The accomplice Chen received non-prosecution.

(3)Prosecution Claims

The prosecution charges that on December 20, 2024, defendant Chen Mouwen unlawfully entered...



(4) Defense Statement

The defendant acted out of necessity for subsistence, without knowledge of the protected status of either the area or species. As a first-time offender demonstrating full cooperation...

(5) Defendant Statement

The defendant maintains this was for subsistence purposes without knowledge of the area's protected status or the roe deer's classification as a protected species...

Court of First Instance

(6) Judicial Decision – First Instance

Junior Judge

- Defendant & Details of Crime:
The defendant Chen illegally hunted a protected roe deer using snares in Jilin's Sanchazi Forest Farm (Dec 20, 2024). Seized items included hunting tools and the carcass (valued at ¥3,000). Charged under Criminal Law Art.341(2).

Senior Judge

- Case summary:
The defendant illegally hunted protected roe deer in Jilin forest (Dec 2024). ¥3,000 wildlife damage. Charged under Criminal Law Art.341(2).

Chief Judge

- Final Judgement:
Accusations:llegal hunting of a protected roe deer using prohibited snares in a protected forest area.
Relevant Articles: Criminal Law Art.341(2).
Term of Imprisonment: Under 3 years.

appeal

(7) Grounds for Appeal

The appeal challenges the 10-year sentence as unlawful since hunting one "Three-Protected" deer (§341(2)) caps at 3 years. Errors: (1) no proof snares were locally banned; (2) ignored appellant's confession/remorse; (3) overstated harm. Request reduction per statutory limits.

Second Instance

(8) Findings of the Appellate Court

The appellate court finds the original 10-year sentence excessive as the offense involves hunting a common protected species under Article 341(2) rather than endangered wildlife. Considering the appellant's clean record, prompt confession, and minimal ecological impact, the sentence is reduced to [X] years in accordance with statutory limits for such offenses.

Chief Judge

(9) Final Judgment – Second Instance

The court reduces the sentence to [X] years, holding the offense constitutes illegal hunting under Article 341(2) given the appellant's clean record and minimal ecological impact. This judgment is final.

Sha Case of Intentional Injury

Case Information

Cause of Action: Intentional Injury
Case Type: Criminal Case
Defendant: Sha
Prosecution: Procuratorate

(1) Defendant Background:

On July 29, 2015, the defendant Sha, after consuming alcohol, was present at Lujin Barbecue Restaurant located at the intersection of Quannei Street in Quannei Township. At approximately 11:00 PM, he punched the victim Li in the face, causing a Level-2 minor injury, as confirmed by forensic evaluation.

(2) Established Facts:

On the night of July 29, 2015, after consuming alcohol, the defendant Sha was at Lujin Barbecue Restaurant located at the intersection of Quannei Street in Quannei Township. At approximately 11:00 PM, he punched the victim Li in the face without provocation. The assault resulted in physical injuries, which were later identified through forensic examination as constituting a Level-2 minor injury.

(3)Prosecution Claims

Defendant Sha, under the influence of alcohol, punched victim Li in the face at Lujin Barbecue Restaurant...

(4) Defense Statement

The defense acknowledges that Sha struck Li, causing injury. However, it contends that the act was impulsive and influenced by alcohol consumption, lacking pre-meditation...

(5) Defendant Statement

Defendant Sha expresses deep remorse for his actions on July 29, 2015. He admits to punching Li while intoxicated and acknowledges the harm caused. Sha has apologized...

Court of First Instance

(6) Judicial Decision – First Instance

Junior Judge

- Defendant & Details of Crime:
The defendant Sha, assaulted Li Moumou by punching him in the face during a late-night altercation at a barbecue restaurant in Quannei Township on July 29, 2015. The attack, committed while Sha was intoxicated, resulted in a Level-2 minor injury as confirmed by medical assessment.

Senior Judge

- Case summary:
The defendant Sha was found guilty of intentional injury after causing facial harm to Li Moumou.

Chief Judge

- Final Judgement:
Accusations:guilty of intentional injury
Relevant Articles: Article 234 of the Criminal Law.
Term of Imprisonment: 12 months of imprisonment

Prompt for Junior Judge
Case Description and Facts { Fact } **Task** Based on the Criminal Law of the People's Republic of China (amended on December 29, 2023), act as a senior judge to analyze the case description and facts provided in the criminal legal document and complete the following tasks. **Task Requirements** 1. The case may involve multiple defendants; analyze only the primary defendant explicitly specified in the document and clearly state their name. 2. Summarize the defendant's criminal acts, listing the specific actions, resulting consequences, and subjective intent (negligence or intent). 3. Do not make a final conviction or sentencing decision; focus solely on analyzing the criminal acts and their nature. 4. Strictly adhere to the output format, listing criminal acts as separate items. If there is only one act, list only that item. 5. Base the analysis solely on the Criminal Law and the facts provided, without considering other laws or unmentioned circumstances (e.g., voluntary surrender, recidivism).
Output Format
Defendant: [Name] Defendant's Criminal Acts: 1. [Action and consequence] [Negligence/Intent] 2. [Action and consequence] [Negligence/Intent] *One act listed if applicable, no blanks.*

Table 6: Prompt for Junior Judge in *JurisAgents*.

Prompt for Senior Judge
Case Description and Facts { Fact } **Defendant Information** { Response from Junior Judge } **Task** Based on the Criminal Law of the People's Republic of China (amended on December 29, 2023), act as a senior judge to analyze the case description, facts, and defendant information provided in the criminal legal document. Extract and summarize key case information to serve as the basis for subsequent predictions of charges, applicable legal provisions, and sentencing. **Task Requirements** 1. The case may involve multiple defendants; analyze only the primary defendant explicitly specified in the document, clearly stating their name. 2. Summarize key case information, including the defendant's specific actions, consequences, subjective intent (negligence or intent), time, location, and relevant evidence or statements. 3. Do not predict charges, cite legal provisions, or predict sentencing; focus solely on extracting and summarizing facts and information. 4. Output as a single paragraph, concise yet detailed, emphasizing key information and avoiding verbatim reproduction of the original text. 5. Base the analysis solely on the Criminal Law and the provided facts and defendant information, without considering other laws or unmentioned circumstances (e.g., voluntary surrender, recidivism). 6. Do not include the thought process; strictly adhere to the output format.
Output Format
[A single paragraph summarizing the case facts and information in detail.]

Table 7: Prompt for Senior Judge in *JurisAgents*.

Prompt for Chief Judge
Defendant Information { Response from Junior Judge } **Case Summary** { Response from Senior Judge } **Candidate Charges** { top10_rel_accu } **Candidate Legal Provisions** { top10_rel_rules } **Similar Cases** { top5_rel} **Task** As a senior presiding judge, based on the Criminal Law of the People's Republic of China (amended on December 29, 2023), determine the defendant's charges, relevant legal provisions, and penalty according to the defendant information, case summary, candidate charges, and legal provisions. Prioritize selecting from the candidate charges and legal provisions; if the case requires, other applicable charges or legal provisions may be chosen. **Task Requirements** 1. Output a single line in JSON format for the judgment result, strictly adhering to the specified format. 2. Do not output the thought process or judgment analysis. 3. Charges (‘charge‘) should not include the word “crime,” e.g., “intentional homicide crime” should be output as “intentional homicide.” 4. Legal provisions (‘relevant_articles‘) are represented by integer numbers, sourced from the Criminal Law of 2023, amended on December 29. 5. The term of imprisonment (‘imprisonment‘) is expressed in months as an integer, e.g., 1 year is 12, 3.5 years is 42; for death penalty or life imprisonment, ‘imprisonment‘ is 0. 6. In ‘term_of_imprisonment‘, ‘death_penalty‘ and ‘life_imprisonment‘ are boolean values; a suspended death sentence is treated as a death penalty (‘death_penalty: true‘). 7. Do not use code blocks (e.g., “‘json) for output; directly output the JSON string.
Output Format
{"relevant_articles": [Number1, Number2, ...], "charge": ["Charge1", "Charge2", ...], "term_of_imprisonment": {"death_penalty": true/false, "imprisonment": Integer, "life_imprisonment": true/false}}

Table 8: Prompt for Chief Judge in *JurisAgents*.

Prompt for Defendant
Case Description and Facts { Fact } **First-Instance Verdict** { Response from Junior Judge } **Task** You are a defendant in a criminal case. Based on the provided case description, facts, and first instance judgment (charges and penalty), decide whether to accept the first instance judgment. If you accept, it indicates no objections to both the charges and penalty; if you do not accept, you must clearly state your objection to the charges or penalty and provide reasons. **Task Requirements** 1. Strictly adhere to Output Format 1 (accept) or Output Format 2 (do not accept) for your response. 2. When not accepting, clearly specify whether the objection is to the charges or penalty, and provide reasonable reasons based on the case facts. 3. When accepting, output only "accept" without any additional text, spaces, or newlines. 4. Do not output the thought process.
Output Format 1 (Accepting the First Instance Judgment)
accept
Output Format 2 (Not Accepting the First Instance Judgment)
not accept, because [clearly state the objection to the charges or penalty, and provide reasons based on the case facts]

Table 9: Prompt for Defendant in *JurisAgents*.

Prompt for Legal Counsel
Case Description and Facts { Fact } **First-Instance Verdict** { Result From Chief Judge } **Defendant's Grounds for Appeal** { Response from Defendant } **Task** As a professional criminal defense lawyer, based on the Criminal Law of the People's Republic of China (amended on December 29, 2023), draft an appeal petition on behalf of the defendant using the provided case description, first-instance verdict, and defendant's grounds for appeal. Present the appeal request, combining facts, evidence, and legal arguments to challenge the impropriety of the first-instance verdict and advocate for a revised judgment or reduced penalty. **Task Requirements** 1. The case may involve multiple defendants; draft the appeal petition only for the primary defendant explicitly specified, clearly stating their name. 2. The appeal petition must include the following elements: appellant information, appeal request, facts and reasons (integrating case details, defendant's grounds, evidence, and legal analysis), legal basis, and concluding request. 3. If the defendant's grounds are insufficient or absent, supplement with reasonable legal arguments based on the case and verdict (e.g., errors in fact-finding, insufficient evidence, or excessive sentencing). 4. Output only the complete appeal petition text, excluding any other text or reasoning process. 5. Legal arguments should primarily rely on the Criminal Law of the People's Republic of China (amended on December 29, 2023). 6. Do not consider unmentioned circumstances (e.g., voluntary surrender, recidivism); base the petition solely on the provided facts, verdict, and grounds.
Output Format
[Complete appeal petition text, including appellant information, appeal request, facts and reasons, legal basis, and concluding request, with no additional text.]

Table 10: Prompt for Junior Judge in *JurisAgents*.

Prompt for Junior Judge (Second Instance)
Case Description and Facts { Fact } **Appeal Petition** * A legal sentencing task is now underway, based on the Criminal Law of the People's Republic of China (amended on December 29, 2023). A first-instance verdict has been issued, but the defendant has chosen to appeal. The appeal petition is as follows: * { Appeal } **Task** Based on the Criminal Law of the People's Republic of China (amended on December 29, 2023), act as a senior judge to analyze the case description and facts provided in the criminal legal document and complete the following tasks. **Task Requirements** 1. The case may involve multiple defendants; analyze only the primary defendant explicitly specified in the document and clearly state their name. 2. Summarize the defendant's criminal acts, listing the specific actions, resulting consequences, and subjective intent (negligence or intent). 3. Do not make a final conviction or sentencing decision; focus solely on analyzing the criminal acts and their nature. 4. Strictly adhere to the output format, listing criminal acts as separate items. If there is only one act, list only that item. 5. Base the analysis solely on the Criminal Law and the facts provided, without considering other laws or unmentioned circumstances (e.g., voluntary surrender, recidivism). 6. You may assess the reasonableness of the appeal based on the case description. If the appeal is deemed unreasonable, you can disregard the appeal petition and uphold your original judgment.
Output Format
Defendant: [Name] Defendant's Criminal Acts: 1. [Action and consequence] [Negligence/Intent] 2. [Action and consequence] [Negligence/Intent] *One act listed if applicable, no blanks.*

Table 11: Prompt for Junior Judge (Second Instance) in *JurisAgents*.

Prompt for Senior Judge (Second Instance)
Case Description and Facts { Fact } **Defendant Information** { Response from Junior Judge } **Appeal Petition** * A legal sentencing task is now underway, based on the Criminal Law of the People's Republic of China (amended on December 29, 2023). A first-instance verdict has been issued, but the defendant has chosen to appeal. The appeal petition is as follows: * { Appeal } **Task** Based on the Criminal Law of the People's Republic of China (amended on December 29, 2023), act as a senior judge to analyze the case description, facts, and defendant information provided in the criminal legal document. Extract and summarize key case information to serve as the basis for subsequent predictions of charges, applicable legal provisions, and sentencing. **Task Requirements** 1. The case may involve multiple defendants; analyze only the primary defendant explicitly specified in the document, clearly stating their name. 2. Summarize key case information, including the defendant's specific actions, consequences, subjective intent (negligence or intent), time, location, and relevant evidence or statements. 3. Do not predict charges, cite legal provisions, or predict sentencing; focus solely on extracting and summarizing facts and information. 4. Output as a single paragraph, concise yet detailed, emphasizing key information and avoiding verbatim reproduction of the original text. 5. Base the analysis solely on the Criminal Law and the provided facts and defendant information, without considering other laws or unmentioned circumstances (e.g., voluntary surrender, recidivism). 6. Do not include the thought process; strictly adhere to the output format. 7. You may assess the reasonableness of the appeal based on the case description. If the appeal is deemed unreasonable, you can disregard the appeal petition and uphold your original judgment.
Output Format
[A single paragraph summarizing the case facts and information in detail.]

Table 12: Prompt for Senior Judge (Second Instance) in JurisAgents.

Prompt for Chief Judge (Second Instance)
Defendant Information { Response from Junior Judge } **Case Summary** { Response from Senior Judge } **Appeal Petition** { Appeal } **Candidate Charges** { top10_rel_accu } **Candidate Legal Provisions** { top10_rel_rules } **Similar Cases** { top5_rel} **Task** As a senior presiding judge, based on the Criminal Law of the People's Republic of China (amended on December 29, 2023), determine the defendant's charges, relevant legal provisions, and penalty according to the defendant information, case summary, candidate charges, and legal provisions. Prioritize selecting from the candidate charges and legal provisions; if the case requires, other applicable charges or legal provisions may be chosen. **Task Requirements** 1. Output a single line in JSON format for the judgment result, strictly adhering to the specified format. 2. Do not output the thought process or judgment analysis. 3. Charges (‘charge‘) should not include the word “crime,” e.g., “intentional homicide crime” should be output as “intentional homicide.” 4. Legal provisions (‘relevant_articles‘) are represented by integer numbers, sourced from the Criminal Law of 2023, amended on December 29. 5. The term of imprisonment (‘imprisonment‘) is expressed in months as an integer, e.g., 1 year is 12, 3.5 years is 42; for death penalty or life imprisonment, ‘imprisonment‘ is 0. 6. In ‘term_of_imprisonment‘, ‘death_penalty‘ and ‘life_imprisonment‘ are boolean values; a suspended death sentence is treated as a death penalty (‘death_penalty: true‘). 7. Do not use code blocks (e.g., “‘json) for output; directly output the JSON string.
Output Format
{"relevant_articles": [Number1, Number2, ...], "charge": ["Charge1", "Charge2", ...], "term_of_imprisonment": {"death_penalty": true/false, "imprisonment": Integer, "life_imprisonment": true/false}}

Table 13: Prompt for Chief Judge (Second Instance) in *JurisAgents*.