IALR: Instance-Adaptive Legal Rubric Supervision for Legal Judgment Prediction

Anonymous ACL submission

Abstract

Legal judgment prediction (LJP) is important for supporting judicial decision-making and improving judicial efficiency. Existing LLM-based approaches often rely on manually specified reasoning pipelines or external legal structures, while feedback-based methods typically provide generic signals of correctness or coherence rather than case-specific legal supervision. We propose **IALR**: Instance-Adaptive Legal Rubric Supervision for Legal Judgment Prediction. For each training instance, IALR contrasts model responses of varying quality and distills their legally consequential failures into a weighted Instance-Adaptive Legal Rubric. These rubrics convert judgment-relevant conditions into discriminative preference signals for Direct Preference Optimization, enabling the model to learn charge prediction without following a predefined reasoning paradigm. At inference time, IALR requires no additional evaluator or correction module. Experiments on four Chinese criminal charge-prediction benchmarks show that IALR consistently improves Qwen2.5-7B, yielding average absolute accuracy gains of 5.68% in-domain and 8.73% out-of-domain. These results demonstrate the effectiveness and cross-domain transferability of Instance-Adaptive Legal Rubric supervision. The code and data are available at https://anonymous.4open.science/r/2BEB-B0521F7209.

1 Introduction

Legal judgment prediction (LJP) infers judicial outcomes from case facts and is a central task in legal NLP with important implications for legal analysis and decision support. Earlier studies primarily formulated it as supervised fact-to-outcome prediction (Luo et al., 2017; Xiao et al., 2018; Zhong et al., 2018; Xu et al., 2020; Lyu et al., 2023; Huang et al., 2024). With the emergence of large language models (LLMs), LJP research has increasingly considered the reliability and interpretability of the underlying legal reasoning (Deng et al., 2024; Shi et al., 2025; Dai et al., 2025).

Unfortunately, LLM-based methods remain challenged by complex fact patterns, confusable charges, multiple defendants, and conditions that exclude criminal liability (Deng et al., 2024; Yuan et al., 2024; Yu et al., 2025; Zhang et al., 2024). Accurate prediction in these settings requires comprehensive reasoning over the case facts and legal conditions that determine charge boundaries. Existing studies mainly pursue two directions. One direction introduces explicit reasoning structures, legal knowledge, or verification mechanisms to guide legal judgment prediction (Deng et al., 2024; Liu et al., 2025; Shi et al., 2025; Yang et al., 2025). However, these methods often rely on predefined procedures or expert-provided structures, which limits their generalization across diverse cases and tasks. The other direction optimizes legal reasoning through feedback or reinforcement learning (Dai et al., 2025; Nigam et al., 2025), converting quality differences among generated analyses into training signals. However, existing feedback typically relies on task-level signals, such as final-answer correctness, overall response quality, or reasoning-induced confidence gains, without assessing whether the reasoning uses sufficient case facts and legal knowledge for the specific instance. These limitations motivate fine-grained, case-specific supervision without manually specified reasoning structures.

To address this need, we derive for each case a tailored set of legal reasoning criteria and convert these criteria into fine-grained preference supervision that targets compositional reasoning over decisive facts, constitutive legal elements, and charge-boundary distinctions. Building on this design, we propose **IALR** (**I**nstance-**A**daptive **L**egal **R**ubric supervision). To reduce reliance on manually specified fine-grained reasoning structures, IALR per-

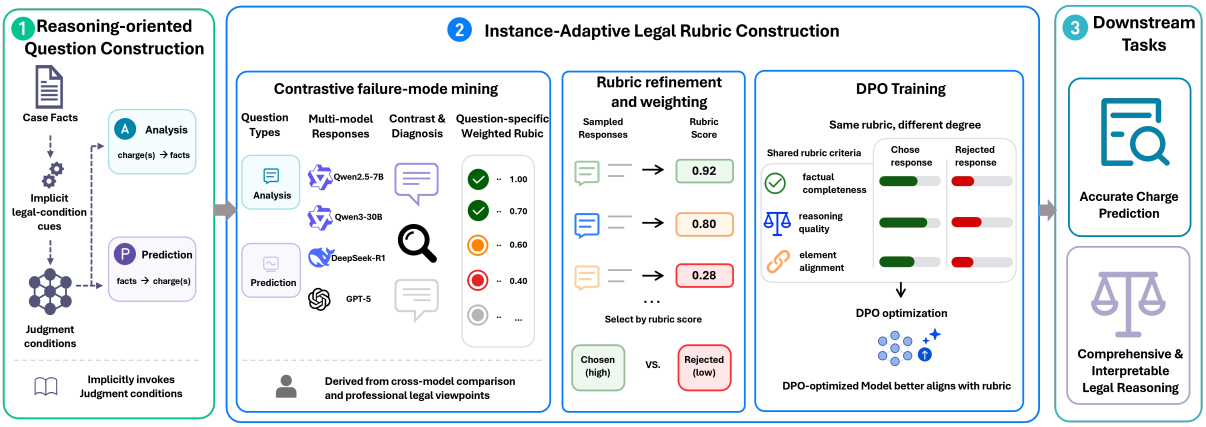

Figure 1: Overview of IALR. Instance-Adaptive Legal Rubrics are induced from multi-model response comparison and then used to construct legally constrained preference pairs for DPO training.

forms contrastive failure-mode mining over diverse model responses to the same case, extracting response deficiencies that reveal which aspects of legal reasoning require examination for that instance. To provide fine-grained feedback centered on the judgment-consequential factors of each case, IALR distills these failure modes into a weighted Instance-Adaptive Legal Rubric. The preference data constructed using this case-adaptive rubric are used for Direct Preference Optimization (DPO; Rafailov et al., 2023). At inference time, the optimized model directly performs charge prediction without relying on rubrics or any additional modules.

We evaluate IALR on four Chinese criminal charge-prediction benchmarks: CAIL2018, MultiLJP, CMDL, and LawBench. Experimental results show that IALR consistently improves charge-prediction accuracy, with average absolute gains of 5.68% in-domain and 8.73% out-of-domain. These results demonstrate that instance-adaptive legal rubric supervision improves prediction performance within the preference-construction distribution and transfers effectively to unseen legal datasets.

Our contributions are threefold:

- We propose IALR, an Instance-Adaptive Legal Rubric supervision framework that targets judgment-relevant conditions in charge prediction, moving beyond fixed reasoning procedures and holistic response-level feedback.
- We develop an automated preference construction pipeline that mines contrastive legal failures from multi-model responses, distills them into weighted Instance-Adaptive Legal Rubrics, and selects legally discriminative preference pairs for DPO while preserving direct inference at test time.
- We conduct experiments on four Chinese criminal charge-prediction benchmarks, demonstrating consistent improvements over the base model in both in-domain and out-of-domain settings and providing empirical evidence for the transferability of Instance-Adaptive Legal Rubric supervision.

2 Methodology

We present IALR, an Instance-Adaptive Legal Rubric supervision framework for legal judgment prediction. Given case facts x, a target defendant d, and a reasoning-oriented question q, IALR trains the model to identify decisive facts, align them with constitutive legal elements, reason over their composition, and distinguish adjacent charges when necessary. Instead of relying on case-level expert annotations, fixed reasoning pipelines, or manually crafted legal templates, IALR induces Instance-Adaptive Legal Rubrics from model responses and uses them to construct preference pairs for DPO training.

Figure 1 provides an overview of IALR. We first convert each legal judgment sample into charge-prediction and judgment-analysis questions. For each case–question pair, multi-model responses are compared to expose legally meaningful failure modes, which are refined into a weighted Instance-Adaptive Legal Rubric. The rubric scores candidate responses and selects chosen–rejected pairs whose quality gap reflects decisive fact coverage, fact–element alignment, charge-boundary discrimi-

nation, or conclusion support. The rubric and evaluator are used only for offline preference construction; inference uses only the optimized model.

2.1 Reasoning-oriented Question Construction

To capture judgment-relevant conditions from complementary outcome- and justification-oriented perspectives, we transform each sample into two question types, $\{q^{\text{pred}}, q^{\text{ana}}\}$. The charge-prediction question asks the model to infer the defendant's charge set from case facts, while the judgment-analysis question asks the model to analyze why the facts support a specified legal judgment. The former targets outcome-oriented legal prediction, whereas the latter targets justification-oriented legal analysis. Both require a fact–element–judgment chain, allowing the subsequent rubric construction to target shared judgment conditions from different reasoning perspectives. Complete prompts are provided in Appendix A.

2.2 Instance-Adaptive Legal Rubric Construction

To obtain fine-grained legal criteria tailored to each case without manually specifying reasoning structures, IALR constructs an Instance-Adaptive Legal Rubric for each case–question pair. Checklist-based alignment suggests that instance-specific criteria can provide more interpretable supervision than fixed reward criteria for general instruction following (Viswanathan et al., 2025). In contrast to generic instruction checklists, the Instance-Adaptive Legal Rubric is legally constrained: it is induced from responses to the same criminal case, and each criterion targets case-specific judgment conditions such as decisive facts, fact–element alignment, charge-boundary distinctions, or conclusion support.

Multi-model reference responses. For each case–question pair (x, q), we collect K reference responses from models with different capabilities and reasoning styles. Prior work on scalable AI feedback shows that diverse model responses can support richer feedback construction (Cui et al., 2023). These responses are not treated as ensemble predictions or generic preference labels. Instead, cross-response variation provides contrastive evidence for case-specific legal adequacy: responses may share a final charge but differ in decisive fact coverage, element mapping, charge-boundary handling, or conclusion support.

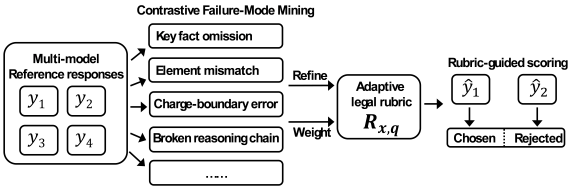

Figure 2: Illustration of Instance-Adaptive Legal Rubric construction. Multi-model responses expose contrastive legal failure modes, which are refined into the weighted rubric $\mathcal{R}_{x,q}$ for scoring candidates and selecting preference pairs.

Contrastive failure-mode mining. A strong evaluator compares the reference responses and induces a checklist of verifiable legal criteria. It focuses on failures that affect the legal conclusion or its support, including decisive fact omission, element mismatch, peripheral-fact overemphasis, missing compositional reasoning, charge-boundary confusion, and unsupported conclusions. Since these criteria are mined from within-instance response differences, they remain discriminative for separating stronger and weaker legal reasoning on that case.

Rubric refinement and weighting. We remove redundant, stylistic, peripheral, or weakly discriminative criteria and weight the remaining ones by their role in the judgment chain. The resulting Instance-Adaptive Legal Rubric is $\mathcal{R}_{x,q} = \{(c_i, w_i)\}_{i=1}^{m}$, where c_i is a legal criterion and $w_i > 0$ is its importance weight. Given a response y, the evaluator assigns binary criterion indicators $z_i(x, q, y) \in \{0, 1\}$, and the rubric score is

$$S_{\mathcal{R}}(x, q, y) = \frac{\sum_{i=1}^{m} w_i z_i(x, q, y)}{\sum_{i=1}^{m} w_i}.$$

This score measures satisfaction of weighted, case-specific legal conditions rather than holistic response quality. Figure 2 illustrates the process.

2.3 Preference Pair Construction with Instance-Adaptive Legal Rubrics

For each (x, q), we sample N candidate responses from the model to be optimized and score each candidate with $\mathcal{R}_{x,q}$. A higher-scoring response is selected as y^+, a lower-scoring response as y^-, and their preference strength is $\Delta(y^+, y^-) = S_{\mathcal{R}}(x, q, y^+) - S_{\mathcal{R}}(x, q, y^-)$. To reduce noisy preferences, we retain only pairs whose margin exceeds

a threshold:

$$\mathcal{D}_{\text{pref}} = \left\{(x, q, y^+, y^-) \mid \Delta(y^+, y^-) > \tau\right\}.$$

Because both responses are compared under the same Instance-Adaptive Legal Rubric, the retained pairs are enriched for legally discriminative contrasts. Compared with the rejected response, the chosen response tends to satisfy more case-specific judgment conditions, such as decisive fact coverage, element alignment, charge-boundary handling, and conclusion support.

2.4 DPO Training and Inference

We optimize the target model with DPO (Rafailov et al., 2023) over $\mathcal{D}_{\text{pref}}$, where each prompt consists of case facts and a reasoning-oriented question, and each chosen–rejected pair is selected by the corresponding Instance-Adaptive Legal Rubric. At inference time, the rubric, checklist, and evaluator are no longer used; the optimized model directly generates charge predictions or judgment analyses without additional test-time modules.

3 Experimental Setup

3.1 Datasets and Evaluation Metrics

We evaluate on four Chinese criminal judgment prediction benchmarks. **CAIL2018** (Xiao et al., 2018) contains standard charge prediction instances from criminal facts; we use its official test set of 32,508 cases. **MultiLJP** (Lyu et al., 2023) targets multi-defendant and multi-charge prediction and contains 2,370 test instances. **CMDL** (Huang et al., 2024) is a large-scale Chinese multi-defendant LJP benchmark designed to evaluate charge prediction under complex defendant–charge associations, with 7,879 test instances. **LawBench 3-3** (Fei et al., 2024) is the charge prediction subset of LawBench, a broader benchmark for evaluating legal knowledge and reasoning in large language models; it contains 500 test instances. Together, these datasets cover settings ranging from standard single-charge prediction to complex multi-defendant and multi-charge judgment prediction. Preference data are constructed from the training/validation splits of CAIL2018 and MultiLJP.

We report **case-level exact-match accuracy**: a prediction is counted as correct only when the predicted charge set exactly matches the gold set. For CAIL2018, following ADAPT (Deng et al., 2024), generated charge expressions outside the predefined label inventory are mapped to the nearest valid label using BGE representations. To ensure comparability, all systems are evaluated with the same output parsing and label-normalization protocol on each benchmark.

Source	Prediction	Analysis	Total
CAIL2018	2,540	1,446	3,986
MultiLJP	2,773	2,337	5,110
Total	5,313	3,783	9,096

Table 1: Statistics of IALR preference pairs.

3.2 Baselines

We use Qwen2.5-7B (Qwen Team, 2025) as the foundation model for the principal experiments and compare IALR with the following representative baselines.

ADAPT (Zero-shot). ADAPT applies the Ask–Discriminate–Predict reasoning procedure at inference time without parameter updates (Deng et al., 2024).

ADAPT (SFT). ADAPT fine-tunes the same Qwen2.5-7B backbone on synthetic reasoning trajectories constructed under the ADAPT framework (Deng et al., 2024). Although the original paper reports training trajectories constructed from both CAIL2018 and MultiLJP, only the CAIL2018 trajectories were publicly available at the time of our experiments. We therefore reproduce ADAPT (SFT) using the released CAIL2018 training resources.

LegalReasoner. LegalReasoner employs LLaMA-3.1-70B together with a verification–correction pipeline for legal judgment prediction (Shi et al., 2025). Since it uses a substantially larger backbone and a different inference pipeline, we include its reported CMDL result only as an external reference rather than as a controlled baseline.

LegalΔ. LegalΔ improves legal reasoning through chain-of-thought-guided information gain and combines supervised fine-tuning with GRPO-based policy optimization (Dai et al., 2025).

3.3 Preference Construction and Training

For each reasoning-oriented question, `gpt-5-chat-latest` (OpenAI, 2025) constructs an Instance-Adaptive Legal Rubric from four multi-model reference responses and scores three candidate responses sampled from the target model. The highest- and lowest-scoring candidates form a preference pair when their normalized

Method	In-domain		Out-of-domain	
	CAIL2018	MultiLJP	CMDL	LawBench
Qwen2.5-7B	77.05	76.47	72.53	55.26
ADAPT (Zero-shot)	78.29	62.42	67.60	56.08
ADAPT (SFT)	81.54	47.19	43.97	56.00
LegalReasoner†	–	–	76.95	–
LegalΔ	80.93	74.52	78.62	63.92
IALR (Ours)	**85.10**	**79.77**	**80.29**	**64.95**

Table 2: Main results on legal judgment prediction. Accuracy (%) is reported. † indicates an external reported result rather than a controlled baseline.

score margin is at least 0.25. This produces 9,096 preference pairs from CAIL2018 and MultiLJP (Table 1); detailed construction statistics are provided in Appendix B.

For IALR, we use Qwen2.5-7B as the foundation model and perform DPO training with the TRL framework (von Werra et al., 2020) on preference pairs constructed from CAIL2018 and MultiLJP. We adopt LoRA for parameter-efficient fine-tuning.

Variant	CAIL2018	MultiLJP
Qwen2.5-7B	77.05	76.47
IALR (w/o Rubric Guidance)	79.84	77.02
IALR (w/o Analysis Supervision)	78.34	71.24
IALR (w/o Prediction Supervision)	78.52	71.30
IALR (Full)	**85.10**	**79.77**

Table 3: Ablation results on rubric guidance and reasoning-oriented supervision sources. Accuracy (%) is reported.

4 Results

4.1 Main Results

As shown in Table 2, IALR achieves the best accuracy on all four benchmarks. Compared with the Qwen2.5-7B foundation model, IALR yields average absolute improvements of 5.68% on the in-domain benchmarks and 8.73% on the out-of-domain benchmarks. IALR also outperforms ADAPT (Zero-shot) across all evaluation benchmarks, with particularly large gains of 17.35% on MultiLJP and 12.69% on CMDL. Under the reproducibility condition described in Section 3.2, IALR further surpasses ADAPT (SFT) by 3.56% on CAIL2018 and by 8.95% on the unseen LawBench benchmark. These results indicate that IALR provides more effective and transferable supervision than both inference-time structured reasoning prompts and supervised reasoning-trajectory tuning.

Compared with LegalΔ, IALR consistently achieves stronger performance across both in-domain and out-of-domain benchmarks, with the most pronounced gain observed on MultiLJP ($+5.25\%$). This advantage suggests that, beyond rewarding reasoning for its contribution to the final prediction, explicitly supervising whether a response captures decisive case facts, aligns them with constitutive legal elements, and distinguishes relevant charge boundaries provides a more legally discriminative learning signal for charge prediction.

On CMDL, IALR with a 7B backbone achieves 80.29%, surpassing the reported 76.95% of LegalReasoner based on LLaMA-3.1-70B. Although this is not a strictly controlled comparison, it further demonstrates the effectiveness of IALR.

4.2 Ablation Study

We conduct ablation studies to examine the contributions of rubric-guided preference construction and the two reasoning-oriented supervision sources. IALR (W/O RUBRIC GUIDANCE) removes Instance-Adaptive Legal Rubric construction and rubric-based scoring, and instead directly asks `gpt-5-chat-latest` (OpenAI, 2025) to select preferred and dispreferred responses from the same candidate pool. Its preference data are matched to IALR in source composition and size. IALR (W/O ANALYSIS SUPERVISION) removes preference pairs derived from judgment-analysis questions, retaining only those constructed from charge-prediction questions. Conversely, IALR (W/O PREDICTION SUPERVISION) retains only preference pairs derived from judgment-analysis questions. All variants use the same foundation model, DPO configuration, and evaluation protocol.

As shown in Table 3, removing rubric guidance reduces accuracy by 5.26% on CAIL2018 and

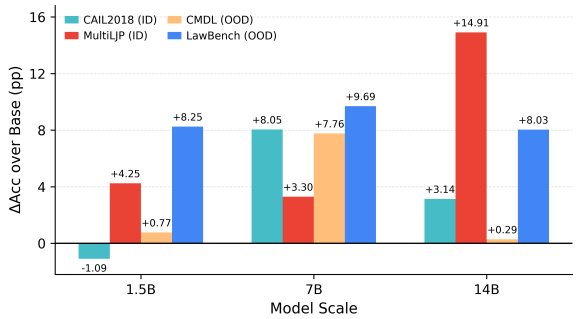

Figure 3: Generalization of IALR across Qwen2.5 model scales. Bars report absolute accuracy changes over the corresponding foundation model; negative values indicate performance degradation.

2.75% on MultiLJP. Compared with direct model selection, rubric-based scoring identifies responses with more complete legal reasoning, while margin-based filtering retains preference pairs with clearer quality separation.

Beyond rubric guidance, the two supervision sources are also complementary. On MultiLJP, removing judgment-analysis supervision or charge-prediction supervision results in comparable drops of 8.53% and 8.47%, respectively. This suggests that complex multi-defendant and multi-charge prediction benefits from jointly learning to discriminate charge outcomes and to identify the judgment-relevant conditions that support them.

4.3 Generalization Across Model Scales

As shown in Figure 3, IALR improves accuracy in 11 of the 12 model–benchmark combinations, demonstrating that its gains extend across multiple model scales. The most consistent improvements are observed on the out-of-domain LawBench benchmark, where all three model scales achieve substantial gains, indicating that rubric-guided preference supervision transfers effectively to unseen charge-prediction distributions. In addition, the particularly large improvement of the 14B model on MultiLJP suggests that models with greater capacity may better leverage instance-adaptive legal supervision in complex multi-defendant and multi-charge settings. Overall, these results support the applicability and cross-domain transferability of IALR across different model scales.

4.4 Efficiency Analysis

Table 4 compares the policy-optimization cost of IALR with LegalΔ under the same Qwen2.5-7B backbone. The DPO stage in IALR requires fewer than 10 H100 GPU-hours, whereas GRPO-based LegalΔ requires more than 672 H100 GPU-hours, yielding a conservative reduction of more than 67$\times$. This comparison concerns policy optimization only and excludes offline Instance-Adaptive Legal Rubric construction and scoring costs. IALR incurs no additional test-time evaluator or rubric calls.

Method	Opt.	Training Cost	Inference
LegalΔ	GRPO	4$\times$H100, $>$ 7 days ($>$ 672 h)	–
IALR	DPO	2$\times$H100, $<$ 5 h ($<$ 10 h)	None

Table 4: Policy-optimization cost comparison.

4.5 Qualitative Analysis

We further examine whether IALR can better identify and apply judgment-relevant conditions in cases involving confusable charges. Table 5 presents a representative case requiring the distinction between *Gathering a Crowd to Fight* and *Picking Quarrels and Provoking Trouble*. The case facts are abridged to retain legally decisive information, while the two model outputs are translated with their substantive reasoning preserved.

Both methods recognize that Defendant C mobilized others and triggered the subsequent assault. However, LegalΔ primarily characterizes the case through the organization of collective violence, leading to the incorrect prediction of *Gathering a Crowd to Fight*. In contrast, IALR additionally identifies and applies the preceding disturbance of public order, the refusal to cooperate with the police investigation, the face-saving retaliatory motive, and the targeted violence against venue staff, thereby reaching the correct charge of *Picking Quarrels and Provoking Trouble*. This case illustrates that IALR grounds charge determination in instance-specific contextual and motivational conditions that define the legal boundary between confusable charges. It further suggests that reasoning supervision without instance-specific legal criteria may favor superficially plausible analyses while overlooking legally decisive distinctions between closely related charges.

5 Related Work

5.1 Legal Judgment Prediction and Structured Legal Reasoning

Legal judgment prediction (LJP) predicts judicial outcomes, such as charges, statutes, or penalties,

Case facts	While watching a performance at an entertainment plaza, Defendant C and Defendant A **disrupted the performance and argued with staff**. After the staff called the police, Defendant C **refused to cooperate and left the scene**. Feeling humiliated, Defendant C instructed Defendant D to gather others and return to **retaliate against the staff**. Under Defendant C's identification, several defendants assaulted staff members, causing one victim a minor injury.	
Gold charge	***Picking Quarrels and Provoking Trouble***	
Method	**LegalΔ**	**IALR**
Model output	Defendant C's conduct constitutes the organizing and directing behavior in *Gathering a Crowd to Fight*. After having a dispute with staff members of the entertainment plaza, he called Defendant D and others to the scene for support, and later identified the victims, directly leading to the fight. Although he did not personally participate in the beating, his conduct satisfies the requirements for a principal offender in *Gathering a Crowd to Fight*. The other defendants were the direct executors in the joint crime, whereas Defendant C, as the ringleader, should bear full responsibility for *Gathering a Crowd to Fight*.	According to the provided information, Defendant C was involved in a group assault. Specifically, while watching a performance, Defendant C **disrupted the performance by violating public order** and argued with staff members of the entertainment plaza. Afterwards, he **refused to cooperate with the police investigation** and left the scene. Later, in order to **save face**, Defendant C instructed Defendant D to assemble others and return to the scene for **retaliation**. Ultimately, Defendant C identified the staff members and participated in the assault against them, causing one staff member a minor injury.
	[Answer] *Gathering a Crowd to Fight*	**[Answer]** ***Picking Quarrels and Provoking Trouble***
Reasoning focus	Focuses on Defendant C's role in gathering and directing multiple assailants, and therefore characterizes the incident as a group fight.	Additionally uses the preceding disturbance of public order, refusal to cooperate, face-saving retaliatory motive, and targeted assault on venue staff to determine the appropriate charge boundary.

Table 5: A representative charge-boundary case. Case facts are abridged to retain judgment-relevant information, and model outputs are translated with substantive reasoning preserved. Bold text highlights the conditions explicitly used by IALR for charge determination.

from case facts. Early neural studies mainly treated LJP as supervised fact-to-outcome prediction (Luo et al., 2017; Xiao et al., 2018; Chalkidis et al., 2019). Large-scale benchmarks such as CAIL2018 further standardized Chinese criminal judgment prediction under this setting (Xiao et al., 2018). Later work moved beyond independent label prediction by modeling dependencies among legal subtasks and legal concepts, including task topology, charge-aware penalty prediction, confusing law article discrimination, and long-document legal representation (Zhong et al., 2018; Chen et al., 2019; Xu et al., 2020; Xiao et al., 2021).

With the rise of large language models, recent LJP research has increasingly explored generative and reasoning-oriented formulations. Representative methods decompose cases, compare candidate charges, verify dispute points, incorporate legal knowledge, or augment reasoning with retrieval (Deng et al., 2024; Shi et al., 2025; Yang et al., 2025; Liu et al., 2025). These studies share the view that legal prediction benefits from structured reasoning over facts, elements, and charge boundaries, rather than from direct outcome generation alone. Different from methods that prescribe explicit reasoning procedures or rely on externally provided legal structures, our method constructs instance-adaptive evaluation criteria from contrastive model responses and uses them for offline preference construction, without introducing additional reasoning modules at inference time.

5.2 Legal Feedback Learning and Preference Alignment

Feedback learning offers another line of work for improving model behavior. General alignment methods such as RLHF, DPO, and LLM-as-a-judge use human preferences, chosen–rejected response pairs, or strong-model evaluations to rank and optimize model outputs (Ouyang et al., 2022; Rafailov et al., 2023; Zheng et al., 2023). These methods typically operate at the response level and provide coarse but scalable supervision for open-ended generation.

Recent legal NLP work has also explored feedback signals for legal reasoning. LegalΔ optimizes reasoning with chain-of-thought-guided information gain, MALR uses multi-agent collaboration and knowledge feedback for confusing charge prediction, and LegalReasoner verifies and corrects intermediate reasoning steps (Dai et al., 2025; Yuan et al., 2024; Shi et al., 2025). Together, these

studies show a growing interest in moving from final-answer supervision toward reasoning-aware feedback. IALR follows this direction by using Instance-Adaptive Legal Rubrics to construct preference pairs for DPO training. Rather than relying on holistic or task-level feedback, our method uses instance-adaptive legal rubrics to evaluate whether model responses identify and correctly apply the judgment-relevant conditions of each case, yielding fine-grained preference supervision for DPO.

6 Conclusion

In this paper, we presented IALR, an Instance-Adaptive Legal Rubric supervision framework for legal judgment prediction. Instead of relying on fixed reasoning procedures or holistic preference signals, IALR constructs case-specific legal rubrics from contrastive failures in multi-model responses and uses them to select legally discriminative preference pairs for DPO. Through complementary charge-prediction and judgment-analysis supervision, the framework encourages models to better capture decisive facts, fact–element correspondence, compositional judgment conditions, and charge-boundary distinctions. Experiments on four Chinese criminal charge-prediction benchmarks show that IALR consistently improves the Qwen2.5-7B base model in both in-domain and out-of-domain settings, with particularly strong gains on previously unseen evaluation distributions. Moreover, because rubrics and external evaluators are used only during offline preference construction, IALR retains direct and efficient inference at test time. These results suggest that instance-adaptive, legally grounded preference supervision provides an effective and transferable approach to improving LLM-based legal judgment prediction.

Limitations

This work has three limitations. First, our experiments focus on Chinese criminal charge prediction. Although we evaluate IALR on both in-domain and out-of-domain benchmarks, the out-of-domain settings still represent distribution shifts within the same broad task and jurisdiction. Therefore, the applicability of IALR to other legal tasks, legal systems, languages, or forms of judicial reasoning remains to be examined.

Second, IALR relies on an external large language model during offline preference construction to induce Instance-Adaptive Legal Rubrics and score candidate responses. Although this dependency introduces no additional inference-time cost, it incurs offline computational and financial overhead and may affect reproducibility when the external model or service changes. More importantly, automatically generated rubrics and preference judgments may inherit biases or legal inaccuracies from the evaluator model. Future work should investigate more reproducible rubric construction with open-source evaluators and systematic auditing of evaluator-induced errors.

Third, we do not conduct direct evaluation by legal experts of either the induced rubrics or the legal reasoning expressed in model outputs. The downstream accuracy improvements and qualitative case study provide evidence that IALR yields useful supervision for charge prediction, but they do not establish that the generated rubrics are legally complete or that the resulting reasoning is consistently aligned with professional legal judgment. Future work should incorporate expert assessment of rubric validity, reasoning sufficiency, and charge-boundary analysis.

Ethical Considerations

IALR is developed for research on legal judgment prediction and is not intended to replace professional legal judgment or support automated judicial decision-making without human oversight. Legal predictions produced by language models may contain factual or doctrinal errors and should not be treated as legal advice or dispositive evidence in real-world proceedings.

References

Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. 2019. Neural legal judgment prediction in english. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 4317–4323, Florence, Italy. Association for Computational Linguistics.

Huajie Chen, Deng Cai, Wei Dai, Zehui Dai, and Yadong Ding. 2019. Charge-based prison term prediction with deep gating network. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 6363–6367, Hong Kong, China. Association for Computational Linguistics.

Ganqu Cui, Lifan Yuan, Ning Ding, Guanming Yao, Bingxiang He, Wei Zhu, Yuan Ni, Guotong Xie,

Ruobing Xie, Yankai Lin, and 1 others. 2023. Ultrafeedback: Boosting language models with scaled ai feedback. *arXiv preprint arXiv:2310.01377*.

Xin Dai, Buqiang Xu, Zhenghao Liu, Yukun Yan, Huiyuan Xie, Xiaoyuan Yi, Shuo Wang, and Ge Yu. 2025. LegalΔ: Enhancing legal reasoning in llms via reinforcement learning with chain-of-thought guided information gain. *arXiv preprint arXiv:2508.12281*.

Chenlong Deng, Kelong Mao, Yuyao Zhang, and Zhicheng Dou. 2024. Enabling discriminative reasoning in llms for legal judgment prediction. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 784–796, Miami, Florida, USA. Association for Computational Linguistics.

Zhiwei Fei, Xiaoyu Shen, Dawei Zhu, Fengzhe Zhou, Zhuo Han, Alan Huang, Songyang Zhang, Kai Chen, Zhixin Yin, Zongwen Shen, and 1 others. 2024. Lawbench: Benchmarking legal knowledge of large language models. In *Proceedings of the 2024 conference on empirical methods in natural language processing*, pages 7933–7962, Miami, Florida, USA. Association for Computational Linguistics.

Wanhong Huang, Yi Feng, Chuanyi Li, Honghan Wu, Jidong Ge, and Vincent Ng. 2024. Cmdl: A large-scale chinese multi-defendant legal judgment prediction dataset. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 5895–5906, Bangkok, Thailand. Association for Computational Linguistics.

Huanghai Liu, Quzhe Huang, Qingjing Chen, Yiran Hu, Jiayu Ma, Yun Liu, Weixing Shen, and Yansong Feng. 2025. Jurex-4e: Juridical expert-annotated four-element knowledge base for legal reasoning. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 3794–3814, Suzhou, China. Association for Computational Linguistics.

Bingfeng Luo, Yansong Feng, Jianbo Xu, Xiang Zhang, and Dongyan Zhao. 2017. Learning to predict charges for criminal cases with legal basis. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2727–2736, Copenhagen, Denmark. Association for Computational Linguistics.

Yougang Lyu, Jitai Hao, Zihan Wang, Kai Zhao, Shen Gao, Pengjie Ren, Zhumin Chen, Fang Wang, and Zhaochun Ren. 2023. Multi-defendant legal judgment prediction via hierarchical reasoning. In *Findings of the association for computational linguistics: EMNLP 2023*, pages 2198–2209, Singapore. Association for Computational Linguistics.

Shubham Kumar Nigam, Tanuj Tyagi, Siddharth Shukla, Aditya Kumar Guru, Balaramamahanthi Deepak Patnaik, Danush Khanna, Noel Shallum, Kripabandhu Ghosh, and Arnab Bhattacharya. 2025. Regal: A first look at ppo-based legal ai for judgment prediction and summarization in india.

OpenAI. 2025. GPT-5 Chat Latest. https://platform.openai.com/docs/models/gpt-5-chat-latest. Accessed: 2026-05-24.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744.

Qwen Team. 2025. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Advances in neural information processing systems*, volume 36.

Weijie Shi, Han Zhu, Jiaming Ji, Mengze Li, Jipeng Zhang, Ruiyuan Zhang, Jia Zhu, Jiajie Xu, Sirui Han, and Yike Guo. 2025. Legalreasoner: Step-wised verification-correction for legal judgment reasoning. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7297–7313, Vienna, Austria. Association for Computational Linguistics.

Vijay Viswanathan, Yanchao Sun, Shuang Ma, Xiang Kong, Meng Cao, Graham Neubig, and Tongshuang Wu. 2025. Checklists are better than reward models for aligning language models. *Preprint*, arXiv:2507.18624.

Leandro von Werra, Younes Belkada, Lewis Tunstall, Edward Beeching, Tristan Thrush, Nathan Lambert, Shengyi Huang, Kashif Rasul, and Quentin Gallouédec. 2020. TRL: Transformer reinforcement learning. https://github.com/huggingface/trl.

Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, and Maosong Sun. 2021. Lawformer: A pre-trained language model for chinese legal long documents. *AI Open*, 2:79–84.

Chaojun Xiao, Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Yansong Feng, Xianpei Han, Zhen Hu, Heng Wang, and 1 others. 2018. Cail2018: A large-scale legal dataset for judgment prediction. *arXiv preprint arXiv:1807.02478*.

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.

Xinyu Yang, Chenlong Deng, and Zhicheng Dou. 2025. Glare: Agentic reasoning for legal judgment prediction. *arXiv preprint arXiv:2508.16383*.

Yaoyao Yu, Leilei Gan, Yinghao Hu, Bin Wei, Kun Kuang, and Fei Wu. 2025. Evaluating test-time scaling llms for legal reasoning: Openai o1, deepseek-r1, and beyond. *arXiv preprint arXiv:2503.16040*.

Weikang Yuan, Junjie Cao, Zhuoren Jiang, Yangyang Kang, Jun Lin, Kaisong Song, Tianqianjin Lin, Pengwei Yan, Changlong Sun, and Xiaozhong Liu. 2024. Can large language models grasp legal theories? enhance legal reasoning with insights from multi-agent collaboration. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 7577–7597, Miami, Florida, USA. Association for Computational Linguistics.

Kepu Zhang, Haoyue Yang, Xu Tang, Weijie Yu, and Jun Xu. 2024. Beyond guilt: Legal judgment prediction with trichotomous reasoning. *arXiv preprint arXiv:2412.14588*.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, and 1 others. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. In *Advances in Neural Information Processing Systems*, volume 36.

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 Prompts

This appendix reports the prompts used in IALR.

A.1 Judgment-Analysis Question Construction

```
CREATE_ANALYSIS_QUESTION = """
You need to read the {facts} first.
Now, you need to ask a question so that the respondent can briefly analyze, based on the {facts}, what is the reason why {criminal} was convicted of
       the following crime: {accusation}?

Question requirements:
1. The purpose of the question is to analyze the basis of the judgment.
2. The analysis in the question is based on the conditional analysis in the {facts}.
3. The question must be natural.
Now, please only return one of your questions, do not return other content.
"""
```

Figure 4: Prompt for judgment-analysis question construction.

A.2 Charge-Prediction Question Construction

```
CREATE_PREDICT_QUESTION = """
You need to read the {facts} first.
Now, you need to ask a question so that the respondent can briefly analyze, based on the {facts}, what crime would {criminal} be convicted of?

Question requirements:
1. The purpose of the question is to predict the crime of the judgment; there may be more than one crime.
2. The analysis in the question is based on the conditional analysis in the {facts}.
3. The question must be natural.
Now, please only return one of your questions, do not return other content.
"""
```

Figure 5: Prompt for charge-prediction question construction.

A.3 Failure-Mode Mining for Judgment-Analysis Responses

```
CONTRAST_ANALYSIS_RESPONSE = """
You are required to generate a series of questions, serving as a checklist of failure modes, to assess the quality of legal responses. These questions
        are designed to identify instances where the response suffers from a significant failure in its "analysis of the conditions for a valid
       judgment" due to a specific omission.

The general principles for formulating these questions are as follows:
1. Each checklist item must be a question that allows for a clear, binary determination of whether the condition is "met" or "not met."
2. Each checklist item must be specific and concrete, avoiding vague suggestions or platitudes.
3. Each checklist item must center specifically on the "conditions upon which the validity of the judgment depends."
4. Priority should be given to checking for "high-value" defects: the omission of critical facts, a failure to properly map facts to legal elements, a
        lack of mutual support among multiple legal elements, or errors in distinguishing between different criminal charges.
5. Do not generate items that merely restate the same requirement using different phrasing.
6. Do not generate items that pertain solely to writing style, general completeness, or professional polish, but offer little substantive value to the
        analysis of the judgment's validity.
7. If two items address the same level of analytical issue, retain only the one that is more specific and possesses greater diagnostic power.

The current prompt is:
{question}

My primary focus is on whether the response successfully accomplishes the following:
1. Extracts the critical conditions, those that genuinely determine the validity of the judgment or conviction, from the [factual narrative of the
       crime], rather than merely restating the case facts.
2. Explicitly maps these critical conditions to specific legal elements (constituents of the offense), including but not limited to: {components}.
3. Articulates which factual conditions support the establishment of which specific legal elements, rather than simply listing facts and legal
       elements in separate, disconnected lists.
4. Explains how multiple legal elements combine, constrain one another, or collectively support the final conclusion.
5. Explains how the absence, insufficiency, or failure to establish a specific critical condition would impact the conviction or the final judgment.
6. Distinguishes between core conditions for validity and supplementary factors, avoiding the mischaracterization of peripheral issues as fundamental
       defects.
7. Prioritizes the examination of critical analytical steps that serve to differentiate between distinct criminal charges, alternative paths to
       conviction, or varying judgment outcomes.

Now, please compare the following responses, which vary in quality, and compile a comprehensive checklist of "all high-value failure modes":
{multi_quality_response}

Output Requirements:
1. The output must consist solely of a numbered list (e.g., 1., 2., 3., etc.).
2. Each item should be written in the format "Reply or not...";
3. Each item should correspond to a unique, verifiable, and non-repeating failure point.
4. Prioritize retaining items that truly differentiate response quality.
5. Keep the number of items to 8-12 if possible.
6. Do not return other content.
"""
```

Figure 6: Prompt for failure-mode mining of judgment-analysis responses.

A.4 Failure-Mode Mining for Charge-Prediction Responses

```
CONTRAST_PREDICTION_RESPONSE = """
You are required to generate a series of questions, serving as a checklist of potential failure modes, to evaluate the quality of a legal response.
      These questions are designed to identify critical omissions: if a specific item is missing, it indicates a significant failure in the response's
       "analysis of the conditions for a valid judgment."

The general principles for formulating these questions are as follows:
1. Each checklist item must be a question that allows for a clear, binary determination of whether the condition has been "met" or "not met."
2. Each checklist item must be specific and concrete, avoiding vague suggestions or platitudes.
3. Each checklist item must center on the "conditions upon which the validity of the judgment depends."
4. Priority should be given to items covering the following areas: the extraction of key factual conditions, the mapping of facts to specific legal
      elements, the interconnections between multiple legal elements, and the evidentiary basis for the final criminal charge or predictive conclusion
      .
5. Do not generate duplicate entries that merely rephrase the same requirement.
6. Do not generate entries that contribute solely to the completeness of expression but offer little value in validating the soundness of the legal
      conclusion.
7. If a potential criminal charge represents only a marginal possibility, do not generate an excessive number of checklist items focused on it.
8. Prioritize items where an omission would directly impact the determination of guilt, the differentiation of charges, the analysis regarding
      probation, or the validity of the predictive conclusion.

The current query is:
{question}

My primary focus is on ensuring the response successfully addresses the following points:
1. Identifies, from the [factual circumstances of the crime], the specific key conditions that truly determine the judgment, conviction, probation
      analysis, or predictive conclusion.
2. Explicitly maps these key conditions to specific legal elements, including, but not limited to: {components}.
3. Does not merely list the legal elements, but explains how the factual circumstances of the crime satisfy those elements.
4. Does not merely list the facts, but explains how multiple factual conditions collectively support multiple legal elements.
5. Explains the relationships between the various legal elements, and how these relationships collectively support the final determination of the
      charge for {crime}, the probation analysis, or other predictive conclusions.
6. Distinguishes between the conditions that constitute the core basis for the judgment and those that serve merely as auxiliary factors in the
      analysis.
7. Ensures that the legal analysis regarding {crime} is fundamentally accurate, while avoiding the generation of excessive checklist items focused on
      marginal or "noise" elements that may appear within the list of potential charges.

Now compare the following multiple responses of different qualities, and then write a list of "all high-value failure modes":
{multi_quality_response}

Output Requirements:
1. Only output a list like 1. 2. 3. ...
2. Each item should be written in the form of "Reply/Not..."
3. Each item must correspond to an independent, identifiable, and non-repeating failure point.
4. Prioritize items whose absence would directly affect the final legal conclusion.
5. Keep the number of items to 8-12 if possible.
6. Do not return any other content.
"""
```

Figure 7: Prompt for failure-mode mining of charge-prediction responses.

A.5 Checklist Refinement

```
EXAMINE_RAW_CHECKLIST = """
You need to review this checklist and condense it into a version characterized by "high value, low redundancy, and strong distinctiveness."

Original Checklist:
<checklist>
{raw_checklist}
</checklist>

I want the final checklist to focus more on:
1. Whether truly critical factual conditions have been extracted, rather than merely restating the case narrative.
2. Whether key factual conditions have been mapped to specific legal elements, including, but not limited to: {components}.
3. Whether a logical chain has been established linking "Factual Condition -> Corresponding Legal Element -> Judgment Conclusion."
4. Whether the interconnections between multiple legal elements have been analyzed, rather than merely listing them in isolation.
5. Whether core conditions for establishing liability have been distinguished from supplementary factors.
6. Whether it explains how the absence, insufficiency, or failure to establish a specific key condition would impact the final conclusion.

Please filter the items according to the following criteria:
1. Remove duplicate items: If two items check for the exact same failure point, retain only the one that is more specific and provides greater
      diagnostic value.
2. Remove hierarchical redundancy: If a broader item is already fully covered by a more specific sub-item, delete the broader one.
3. Remove low-value items: If the absence of a specific item would have negligible impact on the validity of the final judgment conclusion, prioritize
       its removal.
4. Remove "noise" items: If an item focuses excessively on peripheral charges, speculative candidate charges, or unstable/trivial details, prioritize
      its removal.
5. Prioritize retaining the items that best differentiate the quality of a response:
- Extraction of critical factual conditions.
- Mapping of facts to specific core legal elements.
- Analysis of interconnections between multiple legal elements.
- Differentiation between charges, handling of concurrent offenses, or assessment of the threshold for criminal liability.
- The evidentiary basis for the final judgment conclusion.
6. Do not retain more than two "broad summary items." If there are already multiple specific items covering the details, delete the generalized
      summary items.
7. The final list should contain no more than 10 items; ideally, keep it within the 8-10 range.

Output Requirements:
1. Output only the items that have been retained.
2. Format the list using numbered points: 1. 2. 3. ...
3. Each item should remain phrased in the form of "Does the response..."
4. Do not include any other content in the output.
"""
```

Figure 8: Prompt for checklist refinement.

A.6 Rubric Weight Assignment

```
SET_WEIGHTS = """
You need to assign an "importance" weight (0-100) to each item in the checklist. These checklist items are used to evaluate whether a legal response
       truly fulfills the task of analyzing the "conditions for establishing a judgment."

Checklist to be weighted:
<checklist>
{raw_checklist}
</checklist>

The current issue is:
{cause_event}

I am primarily focused on whether the response achieves the following points:
1. Extracts the key conditions that truly determine the validity of the judgment from the [facts of the crime], rather than merely restating the case
       details.
2. Maps these key conditions to specific constituent elements of the crime, including but not limited to: {components}.
3. Establishes a clear logical chain linking "factual conditions -> constituent elements -> final judgment/conviction conclusion."
4. Analyzes the interconnections between multiple constituent elements and how they collectively support {effect_event}.
5. Distinguishes between core conditions for establishment, criteria for differentiating between charges, conditions regarding concurrence of crimes
       or thresholds for criminal liability, and supplementary reasoning factors.
6. Ensures that the judgment, prediction, or legal analysis regarding {effect_event} is fundamentally accurate.

Please assign weights according to the following rules:
1. If the omission of an item would directly render it impossible to determine "whether the specific charge/judgment conclusion is valid," assign a
       weight of 90 or 100.
2. If the omission of an item would significantly impact the differentiation of charges, the handling of concurrent crimes, the threshold for criminal
        liability, or the determination of core constituent elements, assign a weight of 70 or 80.
3. If an item primarily affects supplementary reasoning, comparative analysis, tendencies in judicial practice, or overall completeness, assign a
       weight of 40, 50, or 60.
4. If an item is merely a summary statement, and is already covered by more specific items, its weight should be lower than the specific items it
       summarizes.
5. Do not assign a weight of 90/100 to generic items, catch-all items, or items related solely to phrasing/expression.
6. Items such as "extraction of key factual conditions," "clear mapping of facts to core constituent elements," and "multiple elements collectively
       supporting the conclusion" should generally be weighted higher than general summary items.
7. Assign the highest weights (90/100) only to items that truly lie at the core of the judgment chain and whose omission would severely undermine the
       final conclusion.
8. You must ensure a clear differentiation in scoring levels; do not allow the majority of entries to cluster within the 80, 90, or 100 range.

Here, the value of n is restricted to:
40, 50, 60, 70, 80, 90, 100

Please append the score to the end of each entry using the following format:
XXXXXXXX(n/100)

Output Requirements:
1. Output only the weighted checklist.
2. Maintain the numbered list format (1., 2., 3., etc.).
3. Append the score solely to the end of the original entry; do not include any other content in the output.
"""
```

Figure 9: Prompt for rubric weight assignment.

B Implementation Details

B.1 Rubric Construction and Candidate Sampling

For each case–question pair, we collect four reference responses, one from each of Qwen2.5-7B, Qwen3-30B, DeepSeek-R1, and GPT-5. These multi-model responses are provided to `gpt-5-chat-latest` (OpenAI, 2025) for contrastive failure-mode mining, rubric refinement, and weight assignment. In this way, the Instance-Adaptive Legal Rubric is constructed from response differences that expose case-specific deficiencies in decisive fact coverage, fact–element alignment, compositional reasoning, and charge-boundary discrimination.

After rubric construction, we sample three candidate responses from the target model for each case–question pair. In the principal experiments, the target model is Qwen2.5-7B; in the scale-generalization experiments, it is the corresponding Qwen2.5 model at the 1.5B, 7B, or 14B scale. For response sampling, we use a temperature of 0.8, `top_p` of 0.9, and `top_k` of 30.

The evaluator scores the three candidate responses using the corresponding Instance-Adaptive Legal Rubric. We select the highest-scoring candidate as the chosen response and the lowest-scoring candidate as the rejected response. A preference pair is retained only when their normalized rubric-score margin is at least 0.25, i.e., $\Delta(y^+, y^-) \geq 0.25$. This filtering strategy discards weakly distinguishable candidates and retains preference signals with clearer legally relevant quality differences.

B.2 Preference Pair Statistics

Table 6 reports the number of case–question pairs processed and the number of preference pairs retained after rubric-guided margin filtering. The prediction and analysis columns in Table 1 correspond to the charge-prediction and judgment-analysis question types, respectively.

Source	Question Type	Processed	Retained	Rate
CAIL2018	Judgment analysis	8,325	1,446	17.4%
CAIL2018	Charge prediction	8,360	2,540	30.4%
MultiLJP	Judgment analysis	5,247	2,337	44.5%
MultiLJP	Charge prediction	5,247	2,773	52.9%
Total	–	27,179	9,096	33.5%

Table 6: Detailed statistics of rubric-guided preference pair construction.

Scale	Model	CAIL2018	MultiLJP	CMDL	LawBench
1.5B	Qwen2.5-1.5B	63.87	61.24	59.69	38.14
1.5B	IALR-1.5B	62.78	65.49	60.46	46.39
7B	Qwen2.5-7B	77.05	76.47	72.53	55.26
7B	IALR-7B	85.10	79.77	80.29	64.95
14B	Qwen2.5-14B	78.49	65.29	77.42	57.53
14B	IALR-14B	81.63	80.20	77.71	65.56

Table 7: Absolute accuracy (%) across Qwen2.5 model scales.

C Detailed Results Across Model Scales

Table 7 reports the absolute accuracy scores of the foundation and IALR-optimized models used to compute the gains shown in Figure 3. The results complement the gain-based visualization by providing the underlying performance values for each model scale and benchmark.

Artifact licenses. We use publicly available datasets, models, and software in accordance with their respective licenses and terms of use. The released code and constructed preference data are provided through the anonymous repository accompanying this submission.