Explanation of Revisions

This revised version incorporates several substantial additions and improvements to enhance the clarity, reproducibility, and empirical depth of the paper. The following changes have been made since the previous submission:

1. **Inter-Annotator Agreement (IAA) Scores and Analysis:** To ensure the reliability of expert-based evaluations, we conducted an IAA analysis using five standard metrics: Fleiss' Kappa, Cohen's Kappa, Intraclass Correlation Coefficient (ICC), Krippendorff's Alpha, and Pearson Correlation Coefficient. Detailed IAA tables for both *Factual Accuracy* and *Completeness & Comprehensiveness* have been added to the main paper along with corresponding discussions.

2. **Insights from Legal Experts:** We include a qualitative summary of observations from domain experts, reflecting their perspectives on the strengths and limitations of AI-generated legal drafts. These insights help validate the practical relevance and usability of our system.

3. **Expanded Experimental Setup and Reproducibility:** To improve transparency and reproducibility, we now describe our experimental setup in greater detail, including model configurations, training hyperparameters, and system settings. This makes it easier for future researchers to replicate our results.

4. **Analysis of SFT Degradation:** The paper now includes a more comprehensive discussion on why Supervised Fine-Tuning (SFT) underperformed in our experiments. We compare the instruction tuning strategy used in SFT with the wrapper-based prompting method, and explore potential overfitting and format misalignment issues.

5. **Document Categorization and Visual Representation:** A figure illustrating the distribution of legal document categories in the dataset has been added to the Appendix. This representation highlights the diversity and structure of the dataset, supporting our claims regarding its comprehensiveness.

6. **Updated Related Work:** The Related Work section has been revised to include citations for all major legal NLP applications mentioned (judgment prediction, summarization, segmentation, NER), incorporating several recent papers recommended by the community.

7. **System Flow Diagram Update:** The flow diagram illustrating the two-phase wrapper architecture has been updated for better clarity and alignment with the written methodology.

8. **Ablation Study on Retrieval and Wrapper Components:** To isolate the impact of our retrieval-based contextualization and wrapper structure, we added ablation studies and comparative baselines. This helps disentangle the contribution of

different components in the proposed framework.

9. **Wrapper Overhead and Efficiency Considerations:** The revised paper now includes a discussion of the computational overhead introduced by the wrapper mechanism. We report inference latencies and address implications for deployment in legal workflows.