Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ28917

Chinese Legal Language Modeling Based on Judicial Corpora: A Case Study on Criminal Reasoning with a Large Language Model

Andrew Dong[1*], Yuhang Yan[2], Jinxuan Shao[3]

[1]Sino - Canada International School, Suzhou, China
[2]The High School Affiliated to Renmin University of China, Beijing, China
[3]Shanghai Guanghua Cambridge International School, Shanghai, China
**Corresponding Author. Email: treeinhouse@qq.com*

Abstract. In today's world, many fields are supported with AI, but there are few research about the applications of large language models on sentencing crime. This paper formulates a legal-large language model specialized in Chinese criminal adjudication. The objective of the model is to give charge recommendations and sentencing range that is in accordance to China's laws. It is based on a transformer-based model (Qwen-7B) and fine-tuned on more than 500,000 anonymized judicial documents. The data we used are from CAIL2018 and CAIL2020, two large legal datasets that store huge amounts of legal cases. Our model addresses charge prediction and sentencing recommendation tasks through the incorporation of prompt-based instruction tuning to simulate judicial logic. The model is able to receive long case inputs and give meaningful outputs. It shows high precision in determining charges and suggesting sentencing ranges. The model also shows high capability in aligning outputs with laws and statutes, giving the charges and sentencing ranges according to legal statutes. Interpretive pattern analysis of the model indicates capabilities in statutory fit. The model shows weakness in addressing complicated case details, where the model will show decrease in precision of output. The results of our model provide useful conclusions for the development of AI-supported judicial systems, where Large Language Models can help and support judges' decision.

Keywords: Legal Language Model, Criminal Reasoning, Chinese Judicial Corpora, Charge Prediction, Sentencing Recommendation, Transformer Architecture

1. Introduction

There are many recent developments in the field of LegalAI. Examples include using AI to give legal advices and boosting efficiency of lawyers by extracting and summarizing data from long documents. In the court, many judges use AI statistical tools to help determine charges. Those developments in LegalAI has helped to decrease the probability of wrong charges, and saved a lot of resources. In China, the progress of AI-assisted court is fast, and approved by the government. Many courts of China uses AI tools to help determine charges and to improve accuracy and efficiency of

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.TJ28917

sentencing, and the most famous one is the “206” system that is used in Shanghai [1]. However, how judges and AI work together to make sentencing of crime fairer is understudied. Our team wants to make contribution to the applications of LegalAI on the determining sentencing ranges and charges. We present a model that is based on Lawformer [2], able to analyze long legal cases and give charges and sentencing ranges that are in accordance with the Chinese criminal law. By incorporating prompt-based instruction tuning, the model is designed to mimic judicial reasoning, providing insights into the effectiveness and limitations of legal AI on real and synthetic criminal cases. This research aims to promote AI-powered judicial systems by bridging deep learning with Chinese legal theory.

1.1. Literature review

The rapid progress in natural language processing (NLP) has revolutionized legal informatics, making automated solutions for statutory interpretation, case prediction, and criminal reasoning possible. Early NLP models face several problems in legal applications. The performance of the models was constrained by the complexity of legal texts, scarcity of data (few datasets), and domain-specific challenges such as logical reasoning and cross-jurisdictional adaptability. Large language models (LLMs), with their strong performance in general domains, encounter difficulties in specialized legal domains due to the complexity and subtlety of the judicial language. Classic rule-based systems are incapable of dealing with the open-textured nature of legal norms, whereas early statistical NLP methods had no deep semantic insight. Recent advances in domain-specific fine-tuning, for example, LegalBERT and Lawformer, have enhanced the applicability of the model for reading and analyzing long legal texts, which is crucial in the legal field, as many legal texts are lengthy. However, the application of LLM to Chinese criminal adjudication is still understudied. This paper presents a transformer-based legal language model trained on more than 500,000 anonymized Chinese judicial documents, with charge prediction and sentencing recommendation as target tasks. Early legal expert systems focused on symbolic logic, taxonomies, and hand-crafted rule trees [3]. Although methodologically sound, these systems struggled with ambiguity and exceptions, common features in criminal law. The rise of statistical NLP brought new approaches to document classification and retrieval, but lacked deep reasoning. More recently, domain-specific adaptations of large language models have emerged, which have avoided some of the flaws of the models mentioned above. LegalBERT and CaseLawBERT demonstrated that fine-tuning on legal corpora boosts performance across tasks like multi-label classification and statute mapping. However, those models used BERT as the base model, which cannot process texts that are longer than 512 tokens, as many legal texts exceed. In China, Lawformer [4], a model based on Longformer, introduced whole word masking suited for Chinese legal texts, offering improved semantic representation. Lawformer is a model that is particularly good at processing and reading long inputs, so the model has improved performance on long legal texts compared to the models mentioned above. Data sets such as CAIL2018, CAIL2020 [5], and China Judgments Online have supported training efforts by providing annotated charges, legal facts, and sentencing decisions. Instruction tuning, prompt-based learning, and RLHF (reinforcement learning from human feedback) now contribute to teaching nuanced behaviors of LLM, especially in legal judgment tasks. Also, Our work builds upon these foundations, applying instruction tuning on criminal prompts with a model inspired by Lawformer but modified for generative legal reasoning.

DOI: 10.54254/2755-2721/2026.TJ28917

1.2. Model description

We fine-tuned a legal language model built upon Lawformer, a transformer-based architecture optimized for Chinese legal text via whole-word masking and uses Domain-Adaptive Pretraining to improve performance. The model is designed to interpret the descriptions of criminal cases and generate predictions on applicable charges, sentencing suggestions, and relevant statutory references. Functionally, the model receives raw case descriptions in Chinese, tokenizes them using a legal-specific encoder, and performs semantic mapping to criminal charges and statutory articles. The output contains the sentencing range and the charge. A decoding head trained for causal generation produces natural language interpretations, often including references to China's criminal law. The decoding parameters include top-p sampling (0.9), temperature (0.25), and repetition penalty (1.0) to balance coherence and diversity. The model is optimized for deterministic output suitable for legal recommendation tasks.

2. Training data and procedure

Table 1. The performance level of the model

Task Type	Sample Size	Quantity	Model Performance
Charge predictions	5000 samples	Accuracy	84.7 percent
Statute Matching	5000 samples	Precision	78.9 percent
Sentencing Range	3000 samples	Range Match	81.5 percent
Generation Fluency	500 samples	Human rating	4.2/ 5

Table 1 shows the performance level of the model. We trained on a corpus composed of CAIL2018 and CAIL2020, two large-scale datasets containing huge amounts of real criminal cases and the verdicts of the judges. Large-scale Chinese criminal AI and law challenge datasets offer criminal case descriptions and verdicts with annotations, which can largely improve performance for models. Retrieved from [https://github.com/china-ai-law-challenge/CAIL2018], 2018-2020. China Judgments Online Corpus: Over 500,000 anonymized criminal rulings segmented into facts, charges, rationale, and sentences. The NLP annotated legal samples: Supporting statutory relevance and judgment summaries. The preprocessing of the data included entity masking, sentence resegmentation, token normalization, and alignment of judicial elements. The data was structured into prompt-response pairs simulating real-world legal questions, such as "Please identify applicable charges given the following fact pattern." The model training was carried out over 5 epochs using mixed precision FP16 with device map='auto' and offload configuration. In the process of optimization, we used AdamW with learning rate scheduling and warm-up, and a batch size of 8 samples. Instruction tuning guided model behavior toward reasoning steps rather than classification. Each analysis takes approximately 2 minutes on the local machine and 10-20 seconds on Google Colab, depending on the complexity of the task. Some evaluation metrics included: Charge prediction accuracy: 82.5 percent, Sentence range alignment:

76.4 percent, Statutory mapping precision: 79.1 percent. The model was deployed using Hugging Face Transformers with command line.

DOI: 10.54254/2755-2721/2026.TJ28917

3. Experiment setup and case analysis

3.1. Prompt format of input

Defining the structure of input prompts is crucial for guiding the model's reasoning process. Each input prompt includes:

- Case description (fact pattern)
- Instruction (e.g., "Identify applicable charges and recommend sentencing range")

An example of the input format is provided below:

Case Description: Defendant Zhao got into an argument with someone at a bus stop and hit the other person's face with his fist, causing minor injuries. Instruction: Identify applicable charges and recommend sentencing range

Expected output: Charge: Intentional Injury Sentencing: Six to twelve months imprisonment under Article 234, Criminal Law.

3.2. Case results

Evaluating the model across diverse scenarios is essential to assess its robustness. Various examples were tested:

- Case 1: A defendant involved in a minor theft incident.
- Case 2: A defendant charged with multiple offenses in a complex dispute.
- Case 3:A synthetic case with mitigating circumstances.

Preliminary results indicate the model accurately identifies charges and proposes sentencing ranges, with further analysis to be detailed in subsequent sections.

The defendant stole ¥1,200 worth of electronics from a retail store.	Charge: Theft; Statute: Article 264; Sentence: 3–6 months
The defendant attacked a man in a park, causing minor facial injuries.	Charge: Intentional Injury; Statute: Article 234; Sentence: 6–12 months
After drunk driving, the defendant ran over a pedestrian and fled.	Charge: Traffic Offense; Statute: Article 133; Sentence: 2–4 years
The defendant distributed 30 grams of illegal drugs in a nightclub.	Charge: Drug Trafficking; Statute: Article 347; Sentence: 7–10 years
The defendant forged a university diploma to secure a job.	Charge: Document Forgery; Statute: Article 280; Sentence: 1–3 years

Figure 1. Some typical inputs and outputs

In Figure 1, we can see some examples of input and the output of the model. Our model is able to analyze case descriptions and propose suitable charges and sentencing range, and gives the charges and range according to the legal statutes. The model interprets the input into output that shows statutory fit.

3.3. Discussion

In this paper, our team trained a model that specializes on predicting charge and sentencing range. Our model showcases promising capabilities in replicating judicial reasoning. The model is able to receive case inputs and give accurate charge suggestions and sentencing ranges. While not perfect, the outputs exhibit statutory understanding and sentencing calibration. Errors occur largely with

DOI: 10.54254/2755-2721/2026.TJ28917

complex multi-charge cases or cases involving mitigating circumstances not explicitly described. The use of whole-word masking and instruction tuning appears crucial in achieving legal-specific fluency. Furthermore, the interpret ability of the model, achieved by structured prompt design and consistent output formatting, improves user trust and facilitates manual review. However, our model is far from perfect, and some of the problems our model face include: Lack of full statutory coverage (certain niche crimes excluded), insufficient understanding of courtroom nuances such as plea bargains or mental health considerations, and absence of multilingual legal translation capacity.

4. Conclusion

This study presents a large language model fine-tuned on Chinese criminal corpora. The model is specialized on adjudication, and can evaluate many different types of crime. The results shows that the model is capable of identifying charges, suggesting sentencing ranges, and aligning outputs with legal statutes accurately. The model proves useful as a legal assistant and as a proof-of-concept for generative law modeling, and can help dealing with crime sentencing. Some of the Future work includes: Expanding to civil, administrative law domains, integrating dialogue-based legal consultation modes, and enhancing explainability and citation verification Building web-based or mobile interfaces for legal practice. The convergence of NLP and law is no longer hypothetical—it is actionable, impactful, and transformative. We hope this work contributes meaningfully to the field of legal AI in China and beyond.

References

[1] Wanqiang Wu, Xifen Lin, 2025, Access to technology, access to justice: China's artificial intelligence application in criminal proceedings.
[2] Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, Maosong Sun, 2021, Lawformer: A Pre-trained Language Model for Chinese Legal Long Documents
[3] Ashley, 2017, Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age
[4] Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, Ion An- droutsopoulos, 2020, LEGAL-BERT: The Muppets straight out of Law School
[5] Chaojun Xiao, Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Yansong Feng, Xianpei Han, Zhen Hu, Heng Wang, Jianfeng Xu CAIL2018: A Large-Scale Legal Dataset for Judgment Prediction