Simulating Judicial Decisions with LLMs: How Public Opinion on Social Media Shapes Labor Law Outcomes

Anonymous ACL submission

Abstract

This paper explores how social media discussions influence judicial decision-making in labor disputes. Using 309,642 comments on labor market conditions from Chinese social media platform Douyin and 10,000 representative labor case outcomes, we analyzed key labor issues and their sentiment patterns, revealing growing dissatisfaction with labor practices. Through a simulation experiment with Large Language Models (LLMs), we examined the impact of public opinion on judicial decisions. Our findings show that social sentiments significantly influence judicial outcomes, with a stronger effect on cases involving lower-skilled occupations. Additionally, different LLMs exhibit varying sensitivities to public opinion, with legal-specific models displaying the highest sensitivity, contrary to expectations. Notably, introducing public sentiment substantially alters the judicial decisions of certain LLMs, particularly in cases related to labor rights and lower-skilled workers. This study highlights the potential of social media discourse to shape judicial fairness, especially in labor disputes.

1 Introduction

To achieve a fair institutional design, judges are typically required to separate legislative and judicial powers. However, the limitations of the law necessitate that judges interpret unclear legal provisions, exercising what is termed 'discretionary power' (Dworkin, 1986). While this power can be essential, there is also the risk of its abuse, especially when cases attract significant public attention, leading to potential pressure on judges to influence their decisions. Public opinion can enhance transparency, prompting judges to exercise their discretion more judiciously and provide adequate legal justifications. However, excessive attention may also undermine impartiality and lead to bias in judicial decisions (Epstein and Knight, 1997).

The social and legal challenges faced by workers in labor disputes in China have garnered increasing attention. Workers, often described as vulnerable and marginalized, deal with widespread issues such as unpaid wages, unemployment, and poor labor protections (Shen, 2008). Legal recourse, including labor disputes, remains one of the few available solutions for these individuals. Social media has become a significant platform for voicing such concerns, with the "anti-996 movement" gaining traction. The "996" system, where employees work from 9 AM to 9 PM, six days a week, has sparked public debate primarily on Chinese social media, highlighting the absence of open discourse in traditional public forums (Yang and Zhang, 2023).

Judges' discretionary powers in handling labor dispute cases are influenced by multiple factors, including these societal realities. Although labor law does not directly address every aspect of workers' daily struggles, it is important to understand how public opinion, shaped through media discussions, might influence judicial outcomes. Preliminary observations suggest that non-legal factors, such as a worker's profession (Neitz, 2013) or judges' ideological beliefs (Garoupa et al., 2022), may affect judicial decisions. Thus, we aim to explore how social media comments, reflecting people's perceptions of labor market conditions, can shed light on public attitudes towards labor issues.

In light of ethical constraints and the unavailability of real-world decision-making experiments in labor dispute cases, we adopt a novel approach using Large Language Models (LLMs) to simulate judicial decision-making within complex social environments. This methodology allows us to explore labor dispute scenarios and examine the interaction between public opinion, legal frameworks, and judicial discretion. In addition to simulating the influence of social media on judicial decisions, we also investigate the biases inherent in LLMs, which share similarities with human judges. Recent studies show that LLMs exhibit biases, such as political and economic preferences, that can influence decisions in labor disputes (Bang et al., 2024; Rozado, 2024; Barkhordar et al., 2024). By exploring these biases, our research not only analyzes the impact of public opinion on judicial decisions but also addresses the risks of bias in LLMs, contributing to the goal of ensuring fairness in future Chinese judicial processes.

Based on the analysis of the data of 309,642 comments in Douyin and 10000 representative cases from China Judgments Online (CJOL) , an online database focusing on the judicial documents of the courts at different levels in China, we study the following 4 research questions (RQs):

- *RQ1*: What are the main labor-related issues and their corresponding public sentiment on Chinese social media?
- *RQ2*: How do simulation outcomes differ in LLM-based judicial simulations with and without social media comment input?
- *RQ3*: How does sensitivity to public opinion vary across labor dispute cases in different skill-level occupations?
- *RQ4*: How do inherent value preferences and biases in different LLMs influence their judicial decisions?

The results show that incorporating public sentiment has a considerable impact on the judicial decisions of some LLMs, especially in cases involving lower-skilled occupations and labor rights. Notably, even legal-oriented LLMs like Farui-plus, which is designed to prioritize legal principles, demonstrated a strong responsiveness to social media sentiment, prompting concerns about the potential impartiality of AI in legal decision-making.

2 Related Work

2.1 Social Media and Public Oponion

While there has been a large amount of literature examining how social media affects labor market outcomes by structuring social networks for employee job search and referrals(Sharone, 2017), performing as "a medium for labor activism"(Wu, 2024), or other mechanisms, they have rarely considered how social media, as public opinion, shapes perceptions of labor relations and affects overall labor relations. According to Dong & Lian (2021), social media-based public opinion analysis is a newly emerging trend in various disciplines in recent years(Dong and Lian, 2021), and this approach is an important foundation for our study. In addition, we also provide insights into how public opinion in the context of the new media era affects social outcomes in a broader sense, such as equality in labor relations or legal fairness.

2.2 Public opinion interferes with the judiciary

Public opinion significantly influences the judicial system in modern democracies. Most researchers now use an interdisciplinary approach, combining theories from communication, psychology, sociology, and jurisprudence, to study the impact of social media and online public opinion on judicial decisions. Public sentiments can greatly influence judges' decisions in high-profile cases, especially on socially contentious issues, where judges may feel pressure from the public. (Black et al., 2016) For instance, a study found that judges take shifts in social opinion into account when making decisions to align with evolving societal values. (Giles et al., 2008) However, empirical research indicates that this alignment with public opinion can sometimes lead judges to render biased decisions. (Rachlinski and Wistrich, 2017) Additionally, other studies have highlighted how social media shapes public opinion, subsequently affecting judicial discretion, and have analyzed the impact of social media on public sentiment in legal cases. (Gruce, 2024)

2.3 Value Preferences of Large Language Moedels

Recent studies have explored the value preferences and biases embedded in Large Language Models (LLMs). Bang et al. found that LLMs exhibit varied political biases across topics, with liberal stances on reproductive rights and conservative views on immigration. (Bang et al., 2024) They also show a US-centric focus, and larger models are not necessarily more neutral. Rozado revealed that conversational LLMs tend to show left-of-center political preferences, primarily due to supervised fine-tuning and reinforcement learning stages. (Rozado, 2024) However, Barkhordar et al. demonstrated that biases extend beyond Western political spectra, as Persian language models also exhibit political and economic biases, including authoritarian tendencies. (Barkhordar et al., 2024) Scherrer et al. showed that LLMs align with commonsense morality in low-ambiguity scenarios but exhibit high uncertainty in high-ambiguity ones, with some models displaying clear preferences likely due to fine-tuning. (Scherrer et al., 2023) Ashery et al. found that social conventions can spontaneously emerge in LLM populations through local interactions, with collective biases developing even when individual agents appear unbiased. (Ashery et al., 2024) Our research explores the influence of public opinion on Chinese labor dispute rulings, drawing parallels between the observed biases in LLMs and the complex factors influencing human judges. This suggests LLMs could be valuable for analyzing public sentiment and its impact on judicial decisions.

3 Dataset Construction

3.1 Douyin Comments

The dataset is based on comments from China's leading short video platform*Douyin*. We developed a web crawler using the open-source tool MediaCrawler[1]. We initially collected 386 short videos and 319,448 comments. Specific relevant keywords are shown in Table 7 in Appendix. 309,642 comments remained after removing comments containing only punctuation, emojis, or @usernames.

3.2 Judgements

we selected case data from 2019 to 2021 collected from China Judgments Online (CJOL) [2] as the research subject.

3.2.1 Selection Critiria

In judicial practice concerning labor disputes, judges frequently exercise discretionary power based on specific case circumstances, particularly when determining labor remuneration amounts. To ensure that our dataset captures cases where such judicial discretion is exercised quantitatively, we specifically selected cases related to Article 38, Paragraph 1 of the Labor Contract Law of the People's Republic of China (2013). This provision establishes three quantitative criteria for determining "failure to pay full labor remuneration": (1) whether base wages meet contractual standards, (2) compliance with statutory minimum wage requirements, and (3) proper payment of supplementary compensation such as overtime pay. Since these criteria provide a structured framework for assessing judicial discretion, selecting cases related to this provision allows our dataset to focus on labor disputes where clear economic and legal benchmarks are applied.

Furthermore, the cases were filtered according to the provisions of Part VI, "Labor Disputes and Personnel Disputes", of the "Regulations on Causes of Action in Civil Cases (2020 Revision)[3]". The specific causes of action selected include: Labor contract disputes, Disputes over confirmation of labor relations, Collective contract disputes, Labor dispatch contract disputes, Part-time employment disputes, Disputes over recovery of labor remuneration, Disputes over economic compensation, Disputes over non-compete agreements, Social insurance disputes, Disputes over pension benefits, Disputes over work-related injury insurance benefits, Disputes over medical insurance benefits, Disputes over maternity insurance benefits, Disputes over unemployment insurance benefits, Disputes over welfare benefits, Employment contract disputes, Appointment contract disputes, Resignation disputes, Dismissal disputes.

3.2.2 Data Preprossing

To ensure the representativeness and quality of our sample, we employed a stratified random sampling method, randomly selecting 3,333, 3,334, and 3,333 cases from the 2019 and 2020 datasets, respectively, resulting in a dataset of 10,000 cases. For preprocessing, we meticulously removed judges' legal reasoning, final court judgments, and extraneous information such as address identifiers and personal details (e.g., names, regions, and genders) to eliminate irrelevant factors. This allowed the large language model (LLM) to focus on simulating the role of a judge in decision-making, while preserving core elements such as plaintiff and defendant claims, defendant's occupation, and factual descriptions, ensuring a streamlined and analytically robust dataset.

3.2.3 Occupation Extraction and Classification

We extracted 3,623 job titles from the complete judicial documents using the ChatGLM-4 model. Three researchers with social socience background then manually annotated the job group based on

[1] https://github.com/NanmiCoder/MediaCrawler
[2] https://wenshu.court.gov.cn/

[3] https://tzqfy.bjcourt.gov.cn/article/detail/2010/07/id/4018438.shtml

the International Standard Classification of Occupations (ISCO-08) [4], excluding Armed Forces Occupations. To assess the reliability of the annotations, we calculated the interannotator agreement using Fleiss's Kappa, which yielded a value of 0.829, indicating a high level of agreement among the annotators. Occupations and related information are shown in the Table 1.

Table 1: Occupations, Their Counts, and Percentages

Occupations	Counts	Percentage (%)
Clerical Support Workers	376	3.79
Craft and Related Trades Workers	2555	25.75
Elementary Occupations	740	7.46
Managers	1406	14.17
Plant and Machine Operators, and Assemblers	674	6.79
Professionals	558	5.63
Service and Sales Workers	2515	25.36
Skilled Agricultural, Forestry and Fishery Workers	627	6.32
Technicians and Associate Professionals	473	4.77

4 Methods

4.1 Text Analysis

4.1.1 Sentimant Analysis

For fine-tuning purposes, we manually annotated 10,000 examples from which we extracted 6,000 comments, ensuring a balanced distribution across categories. The dataset was divided into training and testing sets following a 3:1 ratio. We utilized the Erlangshen-MacBERT-110M-BinaryClassification-Chinese model (Zhang et al., 2022), fine-tuning it over 4 epochs with a batch size of 8 and a learning rate set at 3e-5.

The model demonstrated robust performance, achieving an accuracy of 92.24% and an F1-score of 91.81%. These results indicate that the model effectively distinguishes between negative and non-negative sentiments.

4.1.2 Topic Clustering

For this task, we employed BERTopic (Grootendorst, 2022) along with TopicTuner[5] to optimize the parameters for "min cluster size" and "sample size". To capture subtle differences in word usage between negative and positive sentiment comments, we conducted separate topic modeling for each sentiment category following the sentiment analysis. This approach allowed us to better distinguish thematic nuances within each sentiment.

Subsequently, we removed irrelevant topics to refine the models. Following this filtration process, we manually clustered the remaining topics to enhance thematic coherence and interpretability. The first-level classification primarily includes five categories: worker identity, worker income, evaluation of employers, labor legal relationships, job seeking difficulties. The detailed description of topics are shown in Table 8 in Appendix.

4.2 LLM Judicial Decision Simulations

In the context of labor dispute cases, the relationship between the plaintiff's victory status and the amount to be paid by the defendant plays a crucial role. Specifically, if the plaintiff wins the case, the amount requested from the defendant is typically fully satisfied. In contrast, if the plaintiff loses, the defendant is not required to pay any amount. In cases where the plaintiff partially wins, the defendant is generally required to pay a portion of the amount requested by the plaintiff. This partial win scenario is the most common and is central to our analysis of judicial decisions.

4.2.1 Quantitative Metrics

The following key elements must be extracted for the LLM decision-making process:

1. **Plaintiff's Victory Status**: This indicates whether the plaintiff wins, loses, or partially wins the case.

   - "Yes" means the plaintiff's full request is supported by the judge.
   - "No" means the plaintiff's request is entirely rejected.
   - "Partial Win" means that the judge supports only a portion of the plaintiff's requesint.

   The replacement rules are shown in Table 2. By converting the status of winning cases into numerical values, it facilitates the subsequent calculation of changes in the win rate.

2. **Defendant's Payment to the Plaintiff**: This refers to the various amounts the defendant is required to pay to the plaintiff. While the total amount often includes interest, which can be difficult to accurately determine, we instruct the model to exclude interest calculations. Considering the LLM's limitations in mathematical computations, we do not require the model to compute the total amount directly; instead, we perform the final calculation of the total payment amount ourselves.

[4] https://isco-ilo.netlify.app/en/isco-08/
[5] https://github.com/drob-xx/TopicTuner

These two key elements provide the basis for comparing LLM judicial decisions in the baseline phase (without public opinion influence, relevant prompt is shown in Table 10) and in the experimental phase (with public opinion influence, relevant prompt is shown in Table 11).

Table 2: Replacement Rules Table

Original Value	Replaced Value
Yes	2
No	-2
Partially win	1
Cannot determine	0
Yes/Partially win	1.5
No/Partially win	-0.5

4.2.2 Baseline Judicial Decision Simulations

In the baseline phase, we simulated a judge tasked with making decisions strictly according to labor law, independent of external factors such as public sentiment. This phase utilized a structured system prompt to instruct the LLM to evaluate the case facts, calculate the defendant's payment obligations, and determine case outcomes.

The prompt for this step is as follows:

4.2.3 Judicial Decision Simulations with Public Opinion Influence

To systematically incorporate public opinion into the judicial decision-making simulation while minimizing inconsistencies caused by directly inputting raw social media comments, we structured public opinion as topic-based indicators rather than verbatim comments. This approach ensures that variations in input length and linguistic style do not interfere with the LLM's reasoning process. Specifically, we utilized the engagement matrix (refer to Table 3) to extract topic-based social sentiment indicators.

For each case, we selected five distinct public opinion topics, each accompanied by its respective engagement metrics. These included 1) total number of comments under the topic (*comment_count*), 2) proportion of negative comments (*neg_prop*), 3) user engagement level, reflecting the intensity of discussion (*engagement*), and 4) average number of replies per comment, indicating the depth of discussion (*sub_comment_count*).

These structured inputs were sequentially provided to the model, replacing the original judicial context, to assess how different public opinion climates influenced the model's judicial decisions. The modified prompt ensured that public sentiment was explicitly included as an external factor in judicial reasoning while maintaining computational rigor in determining compensation.

By structuring public opinion as topic-based numerical indicators, this method isolates the effect of social sentiment from confounding linguistic biases, allowing for a more controlled evaluation of how public opinion influences judicial decisions.

$$\text{WRC}^{6} = \begin{cases} \text{Blank Value} & \text{if } |\text{Worker Should Win 1}| = 0, \\ \frac{\text{Worker Should Win 2} - \text{Worker Should Win 1}}{|\text{Worker Should Win 1}|} & \text{otherwise.} \end{cases} \quad (1)$$

$$\text{CCR}^{7} = \begin{cases} \text{Blank Value} & \text{if } |\text{Defendant Payment 1}| = 0, \\ \frac{\text{Defendant Payment 2} - \text{Defendant Payment 1}}{|\text{Defendant Payment 1}|} & \text{otherwise.} \end{cases} \quad (2)$$

5 Results and Analysis

5.1 Topic Clustering

This analysis explores user comments across various labor-related topics to answer our RQ1, categorized into primary and secondary classifications. The primary dimensions focus on broad themes such as employer evaluations, legal labor relationships, worker identities, income, and job-seeking challenges. Each primary dimension contains several secondary subcategories like wages, working hours, work environment, and overtime pay, offering a deeper look into user discussions. Details of public attitudes towards secondary topics are shown in Table 9.

5.2 Sentimant Analysis

Sentiment analysis provides a valuable perspective on the reactions users have towards these topics, making it possible to answer the other part of RQ1. The legal labor relationships category, for instance, shows a high level of negative sentiment, with 23,826 of the 27,670 total comments being negative. Similarly, worker income generates 30,029 negative comments out of 37,440 total comments, showing widespread dissatisfaction with income-related issues. Details are shown in Table 6 and Figure 1.

At the subcategory level, issues such as working hours and overtime pay reveal high levels of negative sentiment. Working hours, with 84,620 sub-comments and 27,099 total comments, sees 21,799 negative responses, reflecting concerns over time management and workload. Likewise, overtime pay has 2,405 negative comments out of 2,405

[5]WRC is the Win Rate Change.
[6]CCR is the Compensation Change Rate.

Table 3: Comment Statistics and User Engagement for Primary Topic Dimensions

Primary Topic Dimension	sub comment count	comment count	Neg. sentiment count	Sentiment Extremity	User Engagement
Evaluation of Employers	207511	60510	44798	0.5193	51301
Legal Labor Relationships	81500	27670	23826	0.2778	23808
Worker Identity	12219	5438	4129	0.4814	5263
Worker Income	97095	37440	30029	0.3959	32454
Job-seeking Challenges	87619	15479	11794	0.4761	13950

total comments, signaling overwhelming dissatisfaction with compensation for extra work.

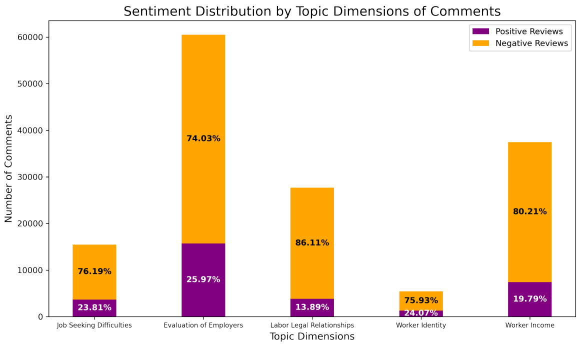

Figure 1: Sentiment Distribution by Topic Dimensions of Comments

5.3 Impact Across Various Public Opinion Dimensions

In this study, we analyzed the influence of public opinion on judicial decision-making in labor disputes in response to RQ2, specifically focusing on Win Rate Change (WRC) and Compensation Change Rate (CCR) for different labor categories. Public opinion was simulated using engagement metrics, such as comment count, negative sentiment proportion, user engagement, and the number of sub-comments. These metrics were derived from large-scale social media analysis, and each model was exposed to five distinct topics, each with its own associated metrics.

As shown in Table 5 and Figure 2, across all models, the most significant shifts in both WRC and CCR were observed in categories related to Legal Labor Relationships and Worker Identity. These categories are closely tied to issues such as employment rights, discrimination, and the legal protection of workers, which are prominent topics in social discourse. The strong public sentiment surrounding worker rights and labor conditions led to noticeable shifts in judicial outcomes, particularly in lower-skilled worker categories.

Worker Income and Job-Seeking Challenges were also notably impacted by public opinion, with public sentiment significantly altering compensation decisions. However, the degree of influence in these categories was not as extreme as in the aforementioned labor rights and identity issues. Finally, Employer Evaluation showed moderate changes in decision-making, suggesting that while employer-related discussions on social media do have an impact, they do not exert as much influence as worker-focused issues.

5.4 Impact Across Various Skill-level Occupations

To address RQ3, we examined how different skill-level occupations respond to public opinion in judicial decisions. Lower-skilled occupations showed greater sensitivity to social sentiment. As shown in Table 4, Table 13 and Figure 3, occupations like Clerical Support Workers (CSW), Craft and Related Trades Workers (CRTW), Elementary Occupations (EO), and Plant and Machine Operators (PMOA), all categorized under Skill Level 2, demonstrated the highest sensitivity to public opinion. These roles, representing lower-skilled workers, are more likely to be affected by societal concerns regarding fair wages, job security, and labor conditions. For example, Farui-plus showed robust sensitivity to public opinion, with high WRC and CCR values in these categories, indicating that the model adjusted its judgments significantly based on public sentiment.

In contrast, Technicians and Associate Professionals (TAP), categorized under Skill Level 3, displayed moderate shifts in judicial outcomes. While Farui-plus and internlm2.5-7b-chat models still showed notable responsiveness to public opinion, the shifts were less drastic compared to lower-skilled occupations. This suggests that mid-tier skilled workers experience less volatility in their judicial outcomes due to public sentiment, likely due to perceived stability in employment and wages.

Managers (MGRS) and Professionals (PROS), representing Skill Levels 3.5 and 4, showed the lowest sensitivity to public opinion. These high-

Table 4: Win Rate Change and Compensation Change Rate by Models and Occupations

Model	Metric	CSW	CRTW	EO	MGRS	PMOA	PROS	SSW	SAFW	TAP
farui-plus	**WRC**	0.707	0.812	0.773	0.72	0.736	0.74	0.761	0.804	0.714
	CCR	2.06	0.871	9.067	1.719	11.381	3.94	1.831	0.784	1.913
ChatGlm-4-9b-chat	**WRC**	0.044	0.016	0.01	0.06	0.04	0.047	0.039	-0.004	0.033
	CCR	8.45	0.6	0.267	6.873	0.334	6.028	0.692	0.192	6.928
DeepSeek V2.5	**WRC**	0.001	0.021	0.035	-0.02	-0.037	-0.016	-0.001	0.046	-0.022
	CCR	0.457	0.263	1.103	0.417	0.567	1.265	0.862	0.063	0.179
gemma-2-9b-it	**WRC**	0.069	0.061	0.089	0.101	0.145	0.105	0.095	0.061	0.151
	CCR	1.474	1.314	2.788	1.065	15.944	1.452	0.894	0.185	5.719
internlm2.5-7b-chat	**WRC**	0.202	0.095	0.135	0.179	0.155	0.235	0.136	0.117	0.134
	CCR	1.114	0.543	1.02	1.025	19.863	0.649	7.509	0.122	0.706
Qwen2.5-7B-Instruct	**WRC**	0.031	-0.012	-0.003	0.014	0.067	0.025	0.014	0.011	0.023
	CCR	0.512	0.147	1.05	1.075	1.983	0.68	3.709	0.197	0.828

Table 5: Win Rate Change (WRC) and Compensation Change Rate (CCR) by Models and Topics. Full terms: WRC - Win Rate Change, CCR - Compensation Change Rate, Evaluation of Employers, Legal Labor Relationships, Worker Identity, Worker Income, Job-seeking Challenges.

Model	Metric	Employers	Labor Relations	Worker Identity	Worker Income	Job-seeking
farui-plus	**WRC**	0.697	0.761	0.789	0.837	0.74
	CCR	2.089	3.768	4.437	2.283	1.432
ChatGlm-4-9b-chat	**WRC**	0.043	0.02	0.028	0.038	0.021
	CCR	1.801	2.223	3.806	1.858	0.616
DeepSeek V2.5	**WRC**	0.034	0.015	-0.02	-0.016	0.013
	CCR	0.45	0.578	0.678	0.562	0.481
gemma-2-9b-it	**WRC**	0.093	0.101	0.094	0.112	0.05
	CCR	2.219	2.481	2.848	2.898	1.37
internlm2.5-7b-chat	**WRC**	0.152	0.095	0.135	0.179	0.155
	CCR	3.718	3.669	3.55	3.595	3.561
Qwen2.5-7B-Instruct	**WRC**	0.031	-0.012	-0.003	0.014	0.067
	CCR	0.834	1.679	1.555	1.239	1.9

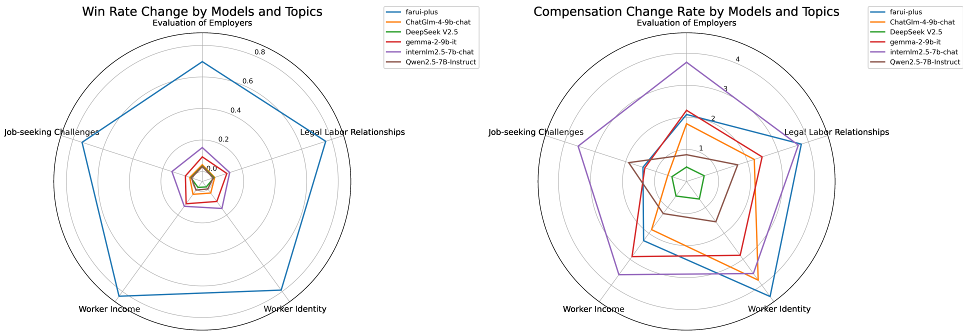

Figure 2: Win Rate Change and Compensation Change Rate by Models and Topics

Topic Dimension	Total	Neg. (Count/Prop.)	Pos. (Count/Prop.)
Job Seeking Difficulties	15479	11794 / 76.19	3685 / 23.81
Evaluation of Employers	60510	44798 / 74.03	15712 / 25.97
Labor Legal Relationships	27670	23826 / 86.11	3844 / 13.89
Worker Identity	5438	4129 / 75.93	1309 / 24.07
Worker Income	37440	30029 / 80.21	7411 / 19.79

Table 6: Sentiment Distribution by Topics of Comments

skilled occupations were the least impacted by social discourse, with minimal shifts in both WRC and CCR. This suggests that managerial and professional roles, due to their higher legal protections and relatively stable economic standing, are less influenced by societal pressures and are more governed by the established legal frameworks in their decision-making processes. Occupations and corresponding acronyms and skill levels are shown in Table12 in Appendix.

5.5 Impact Across Various Models

To answer RQ4, public opinion influenced different LLMs differently, with legal models being more sensitive to social sentiment than general-purpose ones due to their inherent value preferences and biases.

Farui-Plus, the legal-specific LLM, exhibited the highest sensitivity to public opinion across all categories. Despite its legal focus, Farui-plus displayed a high degree of responsiveness to labor-related discussions, shifting both its decision on plaintiff victory and compensation amounts when exposed to social sentiment. This high degree of responsiveness raises an important question: while legal

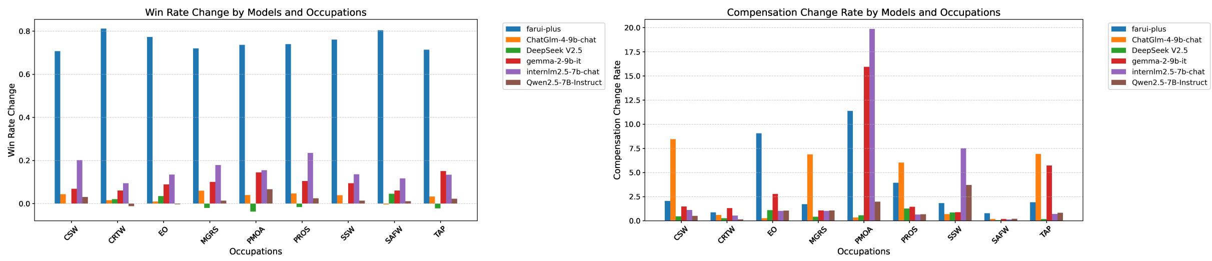

Figure 3: Win Rate Change and Compensation Change Rate by Models and Occupations

LLMs are expected to strictly adhere to established legal principles, Farui-plus appears to incorporate broader labor rights concerns, which aligns with the public discourse on labor issues.

Internlm2.5-7b-chat, while still highly sensitive to public opinion, showed a somewhat lower level of influence compared to Farui-plus. The shifts in its WRC and CCR values were still significant, but not as extreme, indicating a moderate response to public sentiment. This suggests that Internlm2.5-7b-chat is influenced by social discourse but to a lesser extent than specialized legal models.

ChatGLM-4-9b-chat and Gemma-2-9b-it exhibited more moderate responses to public opinion. These models showed some sensitivity to labor issues, but the shifts in WRC and CCR values were less pronounced than in Farui-plus and Internlm2.5-7b-chat. This implies that these models provide a more balanced approach, reacting to public opinion but not allowing it to dramatically alter their decision-making process.

DeepSeek V2.5 and Qwen2.5-7B-Instruct displayed the least sensitivity to public opinion. Their WRC and CCR values remained relatively stable, even when exposed to social sentiment, suggesting that these models are the most legally rigid. Their decision-making is less influenced by public sentiment, indicating that they prioritize intrinsic reasoning over external societal factors when determining judicial outcomes.

6 Key Findings

The primary objective of this study was to investigate how public opinion, as reflected through social media discussions, influences judicial decision-making in labor disputes, particularly with respect to the Win Rate Change (WRC) and Compensation Change Rate (CCR). The results indicate that the introduction of public sentiment significantly alters the judicial decisions of certain LLMs, particularly in lower-skilled occupations and labor rights issues. Interestingly, even legal-specific LLMs, such as Farui-plus, which is designed to focus on legal principles, showed high sensitivity to social media sentiment, raising questions about the impartiality of AI in judicial contexts.

7 Conclusion

This study provides valuable insights into how public opinion influences AI-driven judicial decisions in labor disputes. While legal LLMs like Farui-plus are designed to follow established legal frameworks, they still exhibit notable sensitivity to public discourse. This reflects the growing importance of considering social sentiment in legal decision-making processes. The study highlights the need for balanced models that can make fair, impartial decisions while being aware of societal concerns without being excessively swayed by them. Further research is needed to explore how we can reduce the influence of bias and public sentiment in legal AI systems, ensuring that these tools serve to enhance, rather than compromise, judicial fairness.

8 Limitations

This study has several limitations. First, the dataset relies solely on comments from the Douyin platform, which may not fully represent broader public opinion, as user demographics and discourse styles can vary across platforms. Second, the study's reliance on simulations using LLMs introduces biases inherent in these models, which could affect judicial outcomes, highlighting the need for models that minimize these biases. Third, the research isolates the impact of public sentiment without accounting for other external factors, such as economic conditions or legal precedents, that may influence judicial decisions. Lastly, while public sentiment is classified as positive or negative, this binary classification overlooks more nuanced emotions, which could impact the judicial process in complex real-world scenarios.

References

Ariel Flint Ashery, Luca Maria Aiello, and Andrea Baronchelli. 2024. The dynamics of social conventions in llm populations: Spontaneous emergence, collective biases and tipping points. *Preprint*, arXiv:2410.08948.

Yejin Bang, Delong Chen, Nayeon Lee, and Pascale Fung. 2024. Measuring political bias in large language models: What is said and how it is said. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11142–11159, Bangkok, Thailand. Association for Computational Linguistics.

Ehsan Barkhordar, Surendrabikram Thapa, Ashwarya Maratha, and Usman Naseem. 2024. Why the unexpected? dissecting the political and economic bias in Persian small and large language models. In *Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages @ LREC-COLING 2024*, pages 410–420, Torino, Italia. ELRA and ICCL.

Ryan C Black, Ryan J Owens, Justin Wedeking, and Patrick C Wohlfarth. 2016. The influence of public sentiment on supreme court opinion clarity. *Law & Society Review*, 50(3):703–732.

Xuefan Dong and Ying Lian. 2021. A review of social media-based public opinion analyses: Challenges and recommendations. *Technology in Society*, 67:101724.

Ronald Dworkin. 1986. *Law's empire*. Harvard University Press.

Lee Epstein and Jack Knight. 1997. *The choices justices make*. Sage.

Nuno Garoupa, Fernando Gómez-Pomar, and Adrián Segura. 2022. Ideology andcareer judges: reviewing labor law in the spanish supreme court. *Journal of Institutional and Theoretical Economics*, 178(2):170–190.

Micheal W Giles, Bethany Blackstone, and Richard L Vining Jr. 2008. The supreme court in american democracy: Unraveling the linkages between public opinion and judicial decision making. *The Journal of Politics*, 70(2):293–306.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Jordan Gruce. 2024. *Social Media and the Court: Exploring Impacts, Challenges, and Legal Considerations in the Digital Age*. Ph.D. thesis.

Michele Benedetto Neitz. 2013. Socioeconomic bias in the judiciary. *Clev. St. L. Rev.*, 61:137.

Jeffrey J Rachlinski and Andrew J Wistrich. 2017. Judging the judiciary by the numbers: Empirical research on judges. *Annual Review of Law and Social Science*, 13(1):203–229.

David Rozado. 2024. The political preferences of llms. *PLOS ONE*, 19(7):1–15.

Nino Scherrer, Claudia Shi, Amir Feder, and David Blei. 2023. Evaluating the moral beliefs encoded in llms. In *Advances in Neural Information Processing Systems*, volume 36, pages 51778–51809. Curran Associates, Inc.

Ofer Sharone. 2017. Linkedin or linkedout? how social networking sites are reshaping the labor market. In *Emerging Conceptions of Work, Management and the Labor Market*, volume 30 of *Research in the Sociology of Work*, pages 1–31. Emerald Publishing Limited, Leeds.

Jie Shen. 2008. The characteristics and historical development of labour disputes in china. *Journal of Management History*, 14:161–173.

Tommy Wu. 2024. The potential and limits of spontaneous labor activism on social media: The search for justice among chinese restaurant workers. *Labor Studies Journal*.

Duanyi Yang and Tingting Zhang. 2023. Voice without representation: Worker voice in china's networked public sphere. Available at SSRN: `https://ssrn.com/abstract=4417492`.

Jiaxing Zhang, Ruyi Gan, Junjie Wang, Yuxiang Zhang, Lin Zhang, Ping Yang, Xinyu Gao, Ziwei Wu, Xiaoqun Dong, Junqing He, Jianheng Zhuo, Qi Yang, Yongfeng Huang, Xiayu Li, Yanghan Wu, Junyu Lu, Xinyu Zhu, Weifeng Chen, Ting Han, Kunhao Pan, Rui Wang, Hao Wang, Xiaojun Wu, Zhongshen Zeng, and Chongpei Chen. 2022. Fengshenbang 1.0: Being the foundation of chinese cognitive intelligence. *CoRR*, abs/2209.02970.

A APPENDIX

Table 7: Crawling Keywords

关键词	**Keywords**	关键词	**Keywords**
996	996 work system	进厂黑幕	dark side of entering factories
八小时工作制	eight-hour work system	进厂日常	daily life in factories
被裁	laid off	进厂生活	life in factories
毕业生打工	graduate part-time job	进厂实习	factory internship
毕业生工作	graduate employment	劳动法	labor law
毕业生进厂	graduate entering factories	劳动节	Labor Day
毕业生就业现状	current graduate employment situation	老年人就业	employment for the elderly
毕业生实习	graduate internship	离职	resignation
毕业生实习黑幕	dark side of graduate internships	流水线	assembly line
毕业生现状	current situation of graduates	赔偿	compensation
编制	establishment	骗子中介	fraudulent broker
补偿金	compensation	企业裁员	corporate layoffs
裁员	layoffs	实习黑幕	dark side of internships
辞退	dismissal	实习期	internship period
辞职	resignation	实习生	intern
打工	part-time job	退休	retirement
打工人	part-time worker	延迟退休	delayed retirement
大学生打工	college student part-time job	研究生就业现状	current employment situation of graduate students
大学生打工黑幕	dark side of college student part-time jobs	正式员工	full-time employee
大专实习黑幕	dark side of junior college internships	职场被裁	laid off in the workplace
电子厂	electronics factory	职场被开除	fired in the workplace
工厂打工	factory work	职场辞职	resignation in the workplace
工厂日常	factory daily life	职场氛围	workplace atmosphere
工厂生活	factory life	职场环境	workplace environment
工厂实习	factory internship	职场离职	leaving the workplace
工作制	work system	职场失业	unemployment in the workplace
黑中介	illegal broker	职场现状	current situation in the workplace
技校进厂黑幕	dark side of vocational school factory work	职高进厂黑幕	dark side of vocational high school students entering factories
技校实习黑幕	dark side of vocational school internships	职高实习黑幕	dark side of vocational high school internships
加班	overtime	职校进厂黑幕	dark side of vocational school students entering factories
加班文化	overtime culture	职校实习黑幕	dark side of vocational school internships
进厂打工	work in factories	中专进厂黑幕	dark side of secondary vocational school students entering factories
进厂打工黑幕	dark side of working in factories	中专实习黑幕	dark side of secondary vocational school internships

Topic	Description
Worker Identity	Differentiated legal protections are provided for specific groups in China, such as women, minors, persons with disabilities, and elderly workers. The existence of a valid labor contract determines the applicability of tiered legal protections.
Worker Income	This section categorizes economic rights and interests of workers according to China's labor laws, including five types of insurance and one fund: social security benefits such as pensions and medical care. Detailed calculations of wages, overtime pay, and economic compensation explore the level of economic security and protection for workers.
Evaluation of Employers	This section delves into workers' work experience and satisfaction through an analysis of working hours, the working environment, and credit evaluation. The credit evaluation mainly focuses on the employer's wage payment record and the economic stability of the worker's job.
Labor Legal Relationships	There are four categories in this section: (a) Termination of Relationship: Legal termination conditions and breach of contract handling. (b) Contract Duration: Fixed-term, indefinite-term, and probationary period regulations. (c) Dispute Resolution: Arbitration, mediation, and litigation pathways. (d) Contract Validity: Standardization of contract signing and legal binding force.
Job Seeking Difficulties	Reflects workers' perceptions of macro-level pressures, such as workplace competition, providing social insights for the judicial environment.

Table 8: Description of Key Labor Topics

Subcategory	Total	Neg. (Count/Prop.)	Pos. (Count/Prop.)
Subsidies and Grants	501	501 / 100.00	0 / 0.00
Wages	15187	10973 / 72.25	4214 / 27.75
Work Environment	22220	13697 / 61.64	8523 / 38.36
Working Hours	27099	21799 / 80.44	5300 / 19.56
Overtime Pay	2405	2405 / 100.00	0 / 0.00
Economic Compensation	5352	5148 / 96.19	204 / 3.81
Duration of Labor Relationship	4521	3275 / 72.44	1246 / 27.56
Termination of Labor Relationship	12474	11057 / 88.64	1417 / 11.36
Labor Contract	3842	3041 / 79.15	801 / 20.85
Judicial Labor Disputes	6833	6453 / 94.44	380 / 5.56
Formal Employee Status	3878	2569 / 66.25	1309 / 33.75
Supervisor Status	438	438 / 100.00	0 / 0.00
Special Groups	193	193 / 100.00	0 / 0.00
Five Insurances and One Fund	13995	11002 / 78.61	2993 / 21.39
Credit Evaluation	11191	9302 / 83.12	1889 / 16.88

Table 9: Sentiment Distribution Across Subcategories

Table 10: Prompt for Baseline Judicial Decision-Making

Chinese Prompt	English Prompt
'content': ("'你是一名法官，任务是根据提供的劳动法相关案情进行裁决。请列出被告应支付金额的各款项及其计算式。请简要给出以下格式的输出，勿回复额外内容："')	'content': ("'You are a judge, tasked with adjudicating labor law-related cases based on the provided circumstances. List each item of the amount the defendant should pay and its calculation formula. Briefly provide the output in the following format, without additional content:"')
{"劳动者应否胜诉": "是/部分胜诉/否", "被告应支付金额":"元"}	{"Should the worker win": "Yes/Partially/No", "Amount to be paid by defendant":"$"}

Table 11: Prompt for Judicial Decision-Making with Public Opinion Influence

中文**Prompt**	English Prompt
system_prompt = ('你是一名法官，任务是根据公众对劳动相关议题的舆论裁决案件。请简要给出以下格式的输出，勿回复额外内容：')	system_prompt = ('You are a judge, tasked with adjudicating cases based on public opinion regarding labor-related issues. Briefly provide the output in the following format, without additional content:')
{"以上舆论能否监督裁判":"能/否", "劳动者应否胜诉": "是/部分胜诉/否", "被告应支付金额":"元"}	{"Can public opinion supervise the judgment":"Yes/No", "Should the worker win": "Yes/Partially/No", "Amount to be paid by defendant":"$"}
user_prompt = f"'案情: case_details 舆论话题: topic_with_description 话题评论数: comment_count 负面评价比例: neg_prop 用户参与度: $engagement$ 评论的子评论数: $sub\_comment\_count$"'	user_prompt = f"'Case details: case_details Topic description: topic_with_description Comment count: comment_count Negative evaluation ratio: neg_prop User engagement: $engagement$ Sub-comment count: $sub\_comment\_count$"'

Table 12: Occupations, Acronyms, and Skill Levels

Occupation	Acronyms	Skill Level
Clerical Support Workers	CSW	2
Craft and Related Trades Workers	CRTW	2
Elementary Occupations	EO	2
Managers	MGRS	3.5
Plant and Machine Operators, and Assemblers	PMOA	2
Professionals	PROS	4
Service and Sales Workers	SSW	2
Skilled Agricultural, Forestry and Fishery Workers	SAFW	2
Technicians and Associate Professionals	TAP	3

Table 13: Mean of WRC and CCR by Occupation Skill Level.

Occupation Skill Level	Mean of WRC	Mean of CCR
2	1.045	16.702
3	1.033	16.273
3.5	1.054	12.174
4	1.136	14.014

UTF8gbsn