WORLD LAUREATES FORUM
世界顶尖科学家论坛
2026
国际工程智能大会
2026 International Conference on AI for Engineering
World Laureates Foundation
上海世界顶尖科学家发展基金会
工程智能研究院
INSTITUTE OF AI FOR ENGINEERING

Traffic Accident Liability Determination Using Large Language Models

Junhua Wang[1,2], Wenhao Zhang[1,2], Ting Fu[1,2] , Qiangqiang Shangguan[1,2]

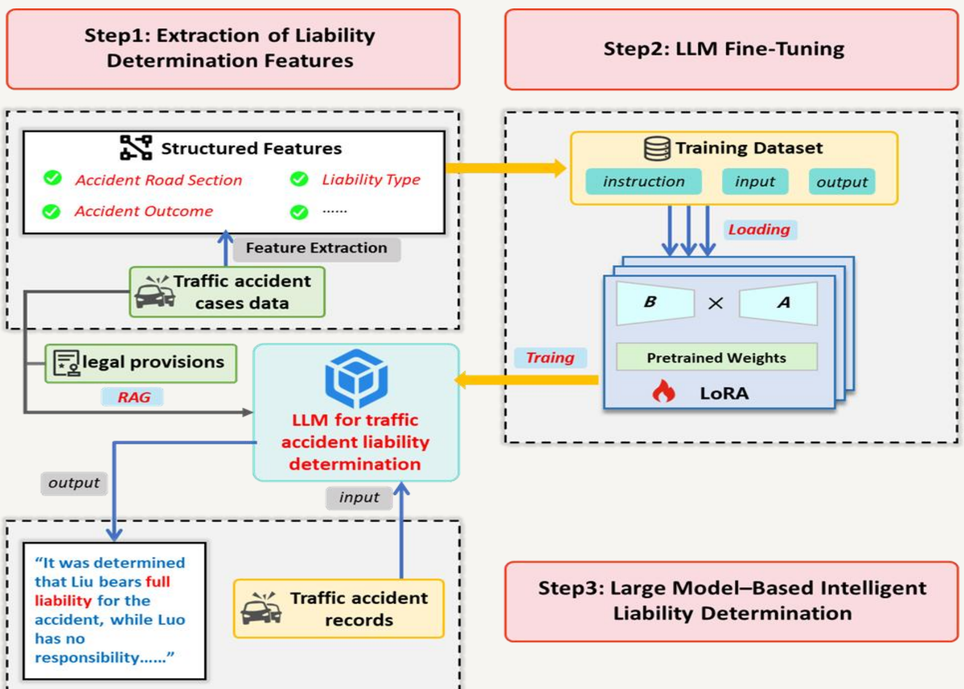
[1] The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University
[2] College of Transportation, Tongji University

01 ABSTRACT

Traffic accident liability determination plays a vital role in traffic safety management. However, existing approaches largely rely on subjective judgment, leading to limited efficiency, scalability, and consistency. This paper proposes an automated and interpretable framework for traffic accident liability determination based on large language models (LLMs). The model is lightweight fine-tuned using Low-Rank Adaptation (LoRA) on 30,379 real-world accident cases and enhanced with Retrieval-Augmented Generation (RAG) and prompt engineering. By dynamically incorporating legal provisions and similar cases, the framework performs factor extraction, behavior analysis, liability reasoning, and final responsibility allocation directly from textual accident descriptions. Experiments are conducted on several mainstream LLMs with different parameter scales (7B, 14B, 20B, and 32B), including DeepSeek-R1, Qwen3, and GPT-oss. The fine-tuned DeepSeek-R1-32B achieves the best performance, with accuracies of 94.8% for full vs. zero liability, 93.5% for primary vs. secondary liability, and 91.5% for equal liability cases, outperforming baseline models. Performance analysis shows that the proposed model exhibits strong logical consistency, domain understanding, and interpretability in both liability reasoning and result generation. In addition, this study proposes a data filtering criterion for accident cases to construct a higher-quality fine-tuning dataset. Empirical results demonstrate that training with the filtered dataset significantly improves determination accuracy and output consistency across different liability types. Overall, the proposed model shows clear advantages in process automation, accuracy, interpretability, and practical applicability, providing an efficient and feasible method for intelligent traffic accident liability determination.

02 METHOD

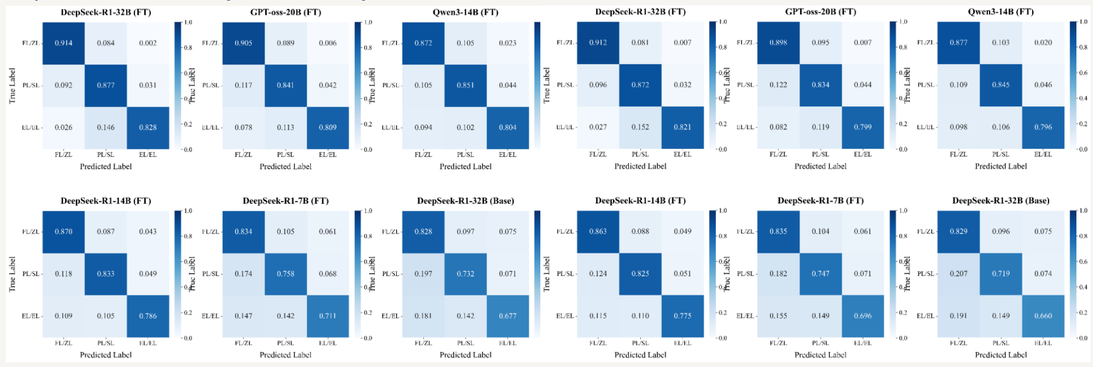
This section aims to describe the fine-tuning methodology and training process of the LLM. First, a locally deployed LLM is used to load the training dataset and is fine-tuned using the LoRA. The training parameters are repeatedly adjusted to achieve optimal model performance. Subsequently, RAG is introduced to integrate the traffic accident liability determination knowledge base, further enhancing the model's capabilities. As a result, an intelligent and accurate LLM for traffic accident liability determination is obtained



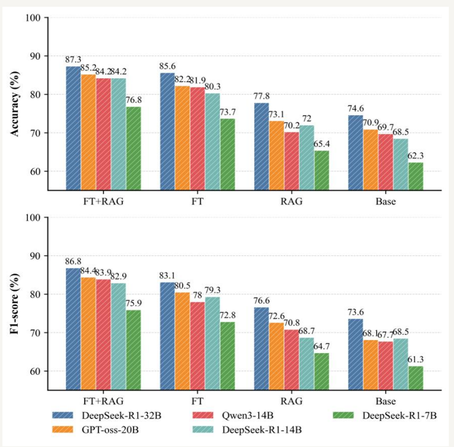
To systematically evaluate fine-tuning performance and to examine the adaptability and scale effects of different types of LLMs in accident liability determination tasks, several mainstream LLMs with varying parameter sizes were selected for fine-tuning, followed by comparative experiments. The model types and parameter scales are listed in Table 4. Specifically, three variants of the DeepSeek-R1 series (7B, 14B, and 32B) developed by Hangzhou DeepSeek Artificial Intelligence Co., Ltd. (DeepSeek-AI et al., 2025), the GPT-oss-20B model developed by OpenAI (Bi et al., 2025), and the Qwen3-14B model developed by Alibaba Cloud Computing Co. Ltd (Yang et al., 2025) were fine-tuned for accident liability determination.

03 RESULTS

Overall, all fine-tuned models significantly outperform the non-fine-tuned DeepSeek-R1-32B (Base). Among them, DeepSeek-R1-32B (FT) achieves the best overall performance with the strongest stability across all evaluation metrics. In addition, all fine-tuned models show better performance on FL/ZL cases, indicating that the models can effectively determine cases with relatively clear liability boundaries. In contrast, the performance on PL/SL and EL/EL cases is generally lower for all models. This is mainly because these cases involve more complex traffic violations, while the semantic similarity of the textual records is higher, leading to greater difficulty in discrimination. Among the compared models, GPT-oss-20B (FT) and Qwen3-14B (FT) also demonstrate strong performance. By comparison, DeepSeek-R1-7B (FT) and DeepSeek-R1-32B (Base) perform relatively poorly overall. The recall and F1-score decrease substantially in EL/EL cases, indicates that the models are highly prone to confusion between this type of case and the other two categories. Furthermore, a clear performance gap exists between DeepSeek-R1-32B (FT) and DeepSeek-R1-32B (Base), which further verifies that domain-specific fine-tuning can effectively improve the accuracy and stability.



Fine-tuning provides a more significant performance improvement for LLMs compared with using RAG only. Compared with the general base model configuration, all evaluated models achieve clear increases in accuracy and F1-score after applying fine-tuning. This demonstrates that domain-specific fine-tuning enables the models to better learn liability-related semantic features, traffic violations, and responsibility allocation rules from accident records. The improvement is particularly evident in large-scale models such as DeepSeek-R1-32B and GPT-oss-20B.The combined FT+RAG configuration achieves the best overall performance. Fine-tuning equips the model with specialized reasoning capability, while retrieval-augmented generation supplements external knowledge during inference.



We comparison between DeepSeek-R1-32B(FT) using General Cases + Judicial Cases, DeepSeek-R1-32B(base), and three existing liability determination models reported in prior studies. It should be noted that the models marked with * were adopted from previous studies. Differences exist in datasets, task definitions, sample composition, feature settings, and experimental schemes among these models. Therefore, the results in this table do not represent a strictly controlled comparison and should only be used as a reference for understanding the relative performance of different algorithms. Compared with traditional machine learning and deep learning methods, the proposed DeepSeek-R1-32B (FT) demonstrates superior performance across all three liability determination scenarios, especially in complex accident cases/

Models/Accuracy	FL/ZL	PL/SL	EL/EL	Overall
DeepSeek-R1:32B (Base)	82.8%	73.2%	67.7%	74.6%
* LR+DNN (Huang et al., 2025)	91.0%	78.0%		81.0%
*Self-Organizing Feature Maps and Data Mining Models (Liu, 2007)				79.5%
*AMFormer Framework (Wang et al., 2025)				93.5%
DeepSeek-R1-32B(FT)	94.8%	93.5%	91.5%	93.7%

04 DISCUSSION

Further human review after obtaining the model's recommendation is essential, especially for equal liability cases. Such cases usually involve more complex accident scenarios and traffic violations, well the model also shows relatively lower accuracy in these cases. Potential risks include incorrect responsibility recommendations, excessive user reliance on the model, insufficient capability in handling difficult cases, and privacy or security issues caused by confidential accident records. Future research will expand accident datasets from different regions and scenario types, improve the model's cross-regional adaptability, optimize uncertainty evaluation functions, and establish a human-machine collaborative framework to ensure compliant and stable deployment.

05 CONCLUSIONS

To sum up, this study presents a LLM designed for determining traffic accident liability. The model integrates fine-tuning on a high-quality accident case dataset with embedded RAG techniques, allowing it to thoroughly analyze accident descriptions provided by users. It can identify violation-related behaviors and causal factors, and automatically carry out the full process—from extracting violations and analyzing behaviors to reasoning about liability and generating the final decision. This method achieves accurate traffic accident liability determination while maintaining clear interpretability. Several key conclusions can be drawn

CONTACT

[Qiangqiang Shangguan] [qiangshangguan@tongji.edu.cn]

QR

REFERENCES / ACKNOWLEDGEMENTS

This study was jointly supported by the National Natural Sci- ence Foundation of China (No. 52472364, No. 52372338), the Shanghai Rising-Star Program (No. 24YF2748100), and the Fundamental Research Funds for the Central Universi- ties.