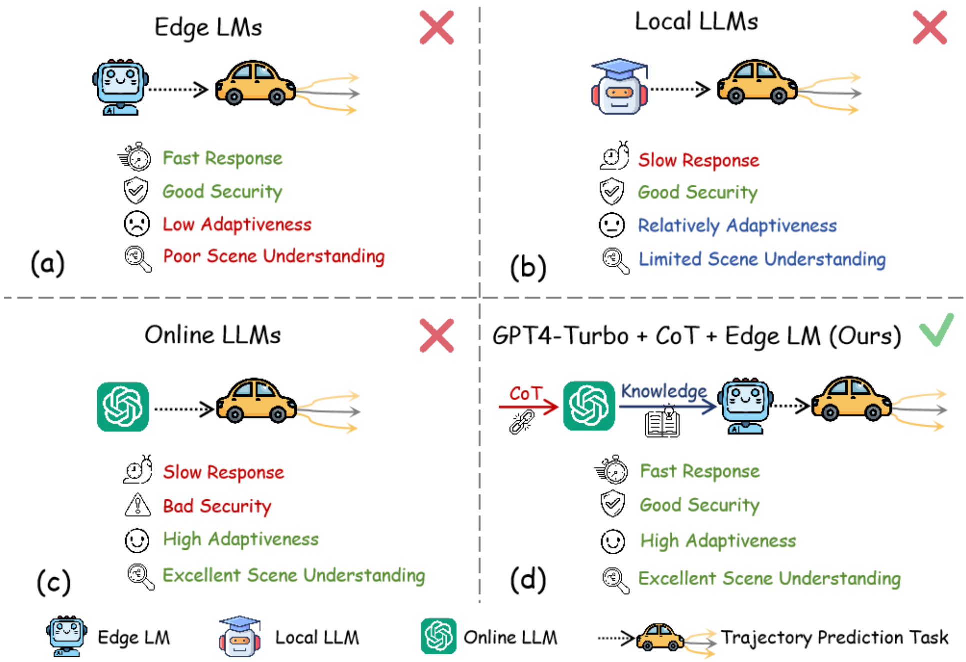
# LLM-TP: Enhancing Trajectory Prediction for Autonomous Driving with LLMs and Chain-of-Thought Prompting




## 1. Abstract

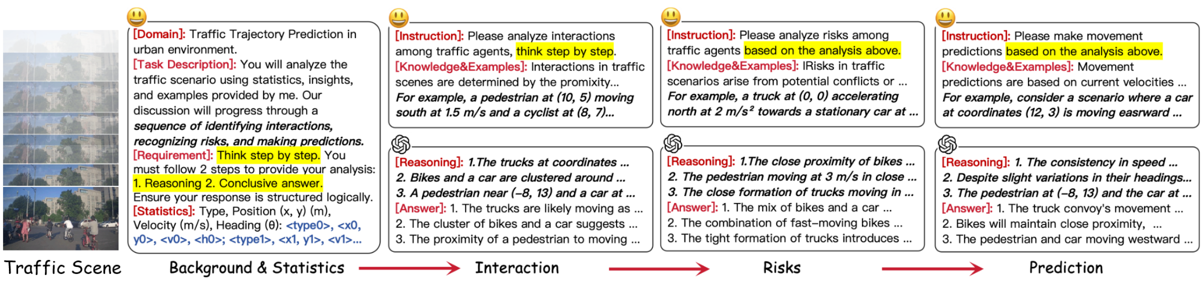

Accurate trajectory prediction is essential for the safety of autonomous driving. In this study, we introduce the LLM-TP (Large Language Model for Trajectory Prediction), a novel approach that enhances trajectory prediction through a chain-of-thoughts reasoning process tailored to specific traffic scenarios. Utilizing the power of LLMs, LLM-TP generates instructive texts that significantly improve the model's understanding of complex traffic environments, thereby boosting prediction accuracy and robustness. We also present two new datasets, Highway-Text and Urban-Text, specifically designed for fine-tuning lightweight language model in the generation of context-specific instructional texts. This approach not only improves prediction but also addresses inference cost concerns through efficient textualized data handling. Incorporating a novel spatial encoding technique and an uncertainty-aware

method within LLM-TP, we achieve more precise trajectory predictions. Our comprehensive evaluations on real-world datasets—NGSIM, HighD, MoCAD, and ApolloScape—demonstrate that LLM-TP performs at state-of-the-art levels, surpassing existing models. In addition, the exceptional performance of LLM-TP underscore its efficacy and viability. Compared to existing models, LLM-TP achieved improvements in prediction accuracy of `12.1%`, `23.1%`, `15.7%`, and `5.1%` on the NGSIM, HighD, MoCAD, and ApolloScape datasets, respectively. These results highlight the model's superiority in handling complex traffic scenarios efficiently.

# 2. Dataset

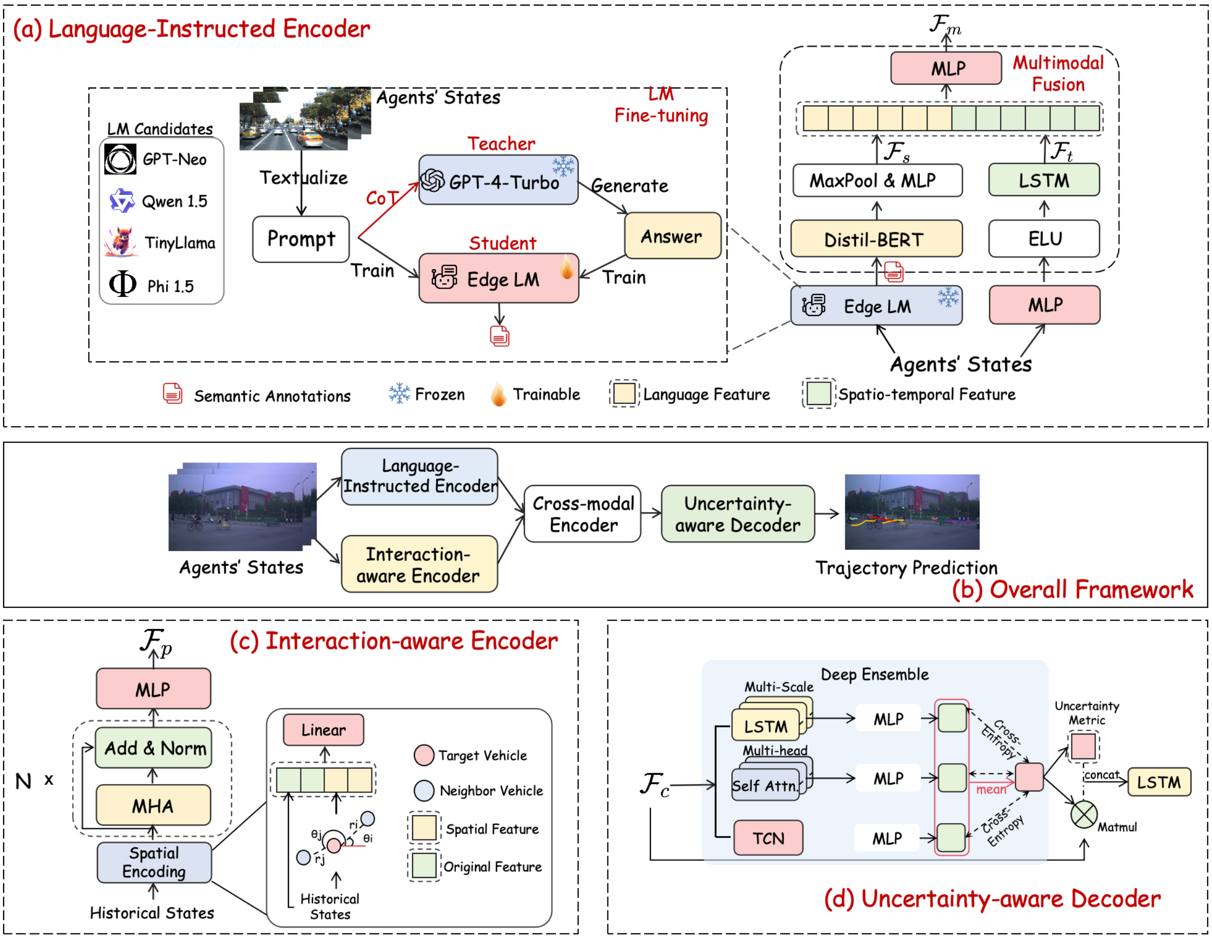
This study contributes to the field of trajectory prediction by introducing two scene description dataset: ***Highway-Text*** and ***Urban-Text***. They collectively encompass over `10` million words describing various traffic scenarios. The ***Highway-Text*** dataset contains scene descriptions from `4,327` traffic scenarios derived from the Next Generation Simulation (NGSIM) dataset and `2,279` scenarios from the Highway Drone Dataset (HighD). Meanwhile, the ***Urban-Text*** dataset includes multi-agent scene descriptions from `3,255` samples in Macao Connected Autonomous Driving (MoCAD) and `2,176` samples from ApolloScape, covering diverse environments such as campus roads, urban roads, intersections, and roundabouts. Both datasets are divided into training (`70%`), validation (`10%`), and testing (`20%`) sets. To our knowledge, these datasets are the first in the field to leverage the linguistic capabilities of the LLM GPT4-Turbo, which utilizes regularized **CoT prompting** to generate detailed semantic descriptions. These descriptions encompass interaction analysis, risk assessment, and movement predictions, offering a comprehensive semantic understanding of traffic scenarios.

A detailed presentation of our CoT prompting is given below.



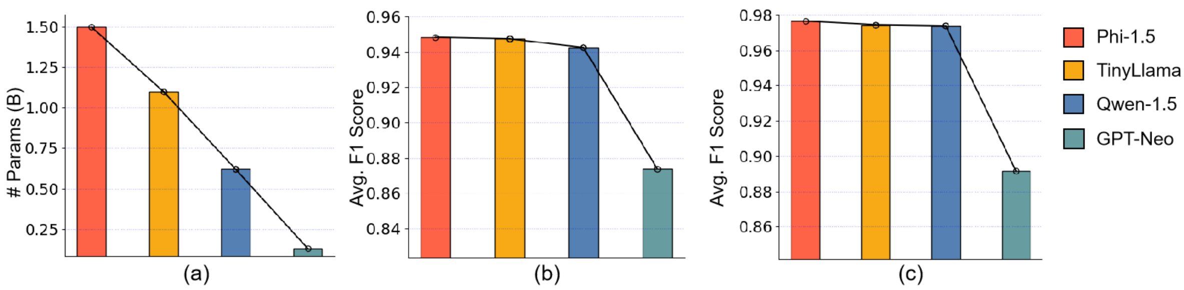
# 3. Methodology Overview

The structure of the proposed framework is illustrated below.

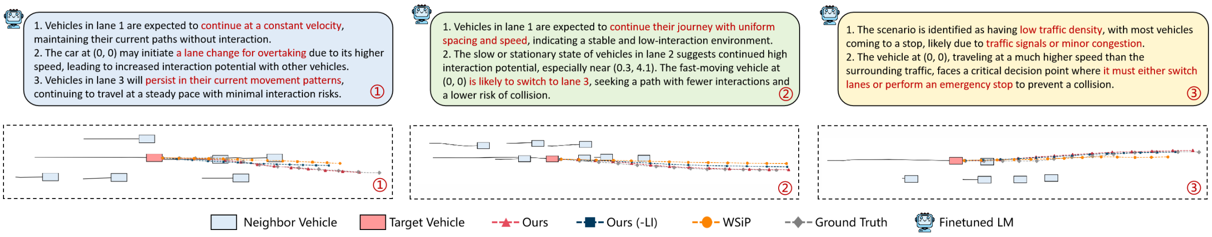



# 4. Experiment

Comparison of four LMs: Parameter Count (a) and Performance on Urban-Text (b) and Highway-Text (c). Metric: F1 Score from BERT Score.




Qualitative results compare the LLM-TP on the NGSIM dataset against its variant without the Language-Instructed Encoder and WSiP. Semantic annotations provided by the fine-tuned LM for each scenario is also displayed on the top of the visualization.




# 5. Contact

If you have any questions, feel free to contact `Hanlin Kong` (hanlinkong@foxmail.com).