[CVPR'24] Driving Everywhere with Large Language Model Policy Adaptation

1. Link:
    1. paper: https://openaccess.thecvf.com/content/CVPR2024/papers/Li_Driving_Everywhere_with_Large_Language_Model_Policy_Adaptation_CVPR_2024_paper.pdf
    2. youtube: https://www.youtube.com/watch?v=fQ3HJbEkP4U
    3. bilibili(w/less informed): https://www.bilibili.com/video/BV16y411h7eG/?spm_id_from=333.999.0.0&vd_source=0340bcb393969ab1f8766419a9956fcf
    4. git(empty, 24.08): https://github.com/Boyiliee/llada
2. Arthurs: Boyi Li, ..., Marco Pavone team (Nvidia, Stanford)
3. TL;DR: A LLM-as-planner framework that use LLM to facilitate traffic-rule interpretations, capable for serving as a driver assitant or a planner-checker in the AD system. The inputs of the framework are ego state, observation description and traffic handbook, the output of is one-step high level plan.
4. [jintian] questions
    1. The idea of providing guidance to drivers is great; however, the questionare-based experiment only test the validness of LLVDA, does this function require time sensitivity? (e.g. warning the driver not turning right before he do so.) If so, how does the module address it?

Existing problems

Traffic Rules in AV Planning.

1. Expressing the entire traffic law is not scalable.
2. Adapting to new traffic rules is difficult.

LLMs for Robotic Reasoning.

1. However, the key difference is that our method focuses on policy adaptation via LLMs rather than the wholesale replacement of modules with LLMs.

LLMs for Autonomous Driving.

1. Traditional perception and planning modules in AV system are generally non-adaptive, preventing AVs from generalizing to any in-the-wild domain.
2. For solving out-of-domain generalization and detection problems, majority of LLM works focus on low-level tasks.

Contribution

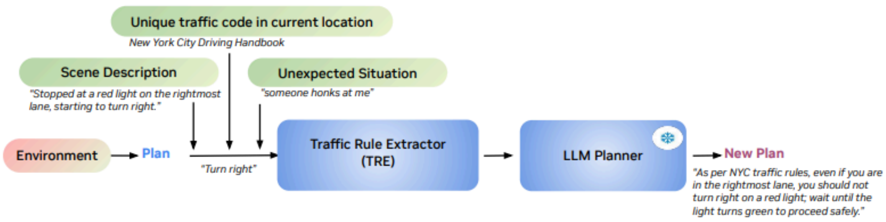

Figure 2. Overview of LLaDA. In this illustration, the driver learned how to drive in California but now needs to drive in New York City. However, the road situation, traffic code, and unexpected situations are different. In our system, we consider three inputs: initial plan ("*Turn right*"), unique traffic code in current location (*New York City Driving Handbook*), and unexpected situation ("*someone honks at me*"). We will feed these three inputs into a Traffic Rule Extractor (TRE), which aims to organize and filter the inputs and feed the output into the frozen LLMs to obtain the final new plan. In this paper, we set GPT-4 as our default LLM.

1. A training-free mechanism to assist human drivers and adapt autonomous driving policies to new environments by distilling leveraging the zero-shot generalizability of LLMs.

Details

1. together with llm planner

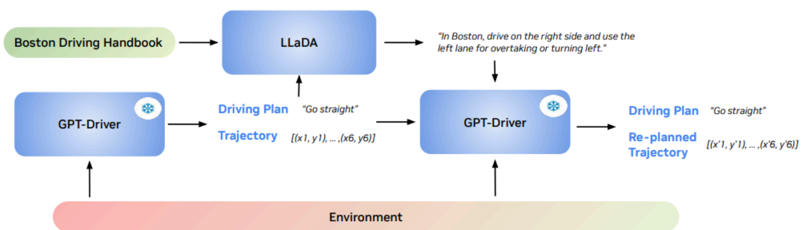

Figure 4. Combining LLaDA with GPT-Driver for motion planning on the nuScenes dataset.

2. questionare design

Location: Germany
Scenario: Driving on the Autobahn in the leftmost lane with heavy traffic.
Unexpected Situation: An emergency vehicle is approaching from behind.

Relevant Local Law: Emergency vehicles (ambulances, police cars, fire trucks, and other vehicles identifiable by their flashing blue lights and multitone warning signals) have special right-of-way privileges. When emergency vehicles approach an autobahn traffic jam, drivers are required to move their vehicles to the extreme right or left, depending on the lane occupied, to permit the emergency vehicle to pass through the center of the congestion. On three-lane autobahns, clearance must be made between left and center lanes

Driving Assistant Instructions:
The driver should immediately and safely move to the rightmost lane to allow the emergency vehicle to pass, as per German traffic regulations.

Does the instruction follow the relevant local law? *

- Yes
- No

How useful is the instruction? *

- Not at all useful
- Slightly useful
- Moderately useful
- Very useful
- Extremely useful