

Driving Everywhere with Large Language Model Policy Adaptation

Boyi Li¹ Yue Wang^{1,2} Jiageng Mao² Boris Ivanovic¹ Sushant Veer¹ Karen Leung^{1,3} Marco Pavone^{1,4}
¹NVIDIA ²University of Southern California ³University of Washington ⁴Stanford University

Abstract

Adapting driving behavior to new environments, customs, and laws is a long-standing problem in autonomous driving, precluding the widespread deployment of autonomous vehicles (AVs). In this paper, we present LLaDA, a simple yet powerful tool that enables human drivers and autonomous vehicles alike to drive everywhere by adapting their tasks and motion plans to traffic rules in new locations. LLaDA achieves this by leveraging the impressive zero-shot generalizability of large language models (LLMs) in interpreting the traffic rules in the local driver handbook. Through an extensive user study, we show that LLaDA’s instructions are useful in disambiguating in-the-wild unexpected situations. We also demonstrate LLaDA’s ability to adapt AV motion planning policies in real-world datasets; LLaDA outperforms baseline planning approaches on all our metrics. Please check our website for more details: [LLaDA](#).

1. Introduction

Despite the rapid pace of progress in autonomous driving, autonomous vehicles (AVs) continue to operate primarily in geo-fenced areas. A key inhibitor for AVs to be able to drive everywhere is the variation in traffic rules and norms across different geographical regions. Traffic rule differences in different geographic locations can range from significant (e.g., left-hand driving in the UK and right-hand driving in the US) to subtle (e.g., right turn on red is acceptable in San Francisco but not in New York city). In fact, adapting to new driving rules and common practice is difficult for humans and AVs alike; failure to adapt to local driving norms can lead to unpredictable and unexpected behaviors which may result in unsafe situations [15, 35]. Studies have shown that tourists are more susceptible to accidents [25, 26] that can sometimes result in injury or death [28]. This calls for a complete study of policy adaption in current AV systems.

At the same time, LLMs have recently emerged as front-runners for zero- or few-shot adaptation to out-of-domain data in various fields, including vision and robotics [6, 16, 17]. Inspired by these works, our goal is to build



Figure 1. LLaDA enables drivers to obtain instructions in any region all over the world. For instance, the driver gets a driver’s license in California, USA, our system enables providing prompt instructions when the driver drives in different regions with different situations.

a **Large Language Driving Assistant (LLaDA)** that can rapidly adapt to local traffic rules and customs (Figure 1). Our method consists of three steps: First, we leverage existing methods to generate an executable policy; second, when presented with an unexpected situation described in natural language (either by a human prompt or a VLM such as GPT-4V [24] or LINGO-1 [32]), we leverage a Traffic Rule Extractor (TRE) to extract informative traffic rules relevant to the current scenario from the local traffic code; finally, we pass the TRE’s output along with the original plan to a pre-trained LLM (GPT-4V [24] in this paper) to adapt the plan accordingly. We test our method on the nuScenes [2] dataset and achieve improvements in motion planning under novel scenarios. We also provide extensive ablation studies and visualizations to further analyze our method.

Contributions. Our core contributions are three-fold:

1. We propose LLaDA, 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.
2. LLaDA can be immediately applied to any autonomous driving stack to improve their performance in new locations with different traffic rules.
3. Our method achieves performance improvements over previous state-of-the-arts, as verified by user studies and experiments on the nuScenes dataset.

2. Related Works

Traffic Rules in AV Planning. Researchers have explored the possibility of embedding traffic rules in the form of metric temporal logic (MTL) formulae [19], linear temporal logic (LTL) formulae [7, 14, 18], and signal temporal logic (STL) formulae [29, 33]. Expressing the entire traffic law as logic formulae is not scalable due to the sheer number of rules and the exceptions that can arise. Furthermore, adapting to traffic rules in a new region still requires the cumbersome encoding of new traffic rules in a machine-readable format. This challenge was highlighted in [20], where the use of a natural language description of traffic rules was proposed as a potential solution. There is a dearth of literature on directly using the traffic rule handbook in its natural language form for driving adaptation to new locations, and it is precisely what we aim to achieve in this paper.

LLMs for Robotic Reasoning. Recently, many works have adopted LLMs to tackle task planning in robotics. These methods generally leverage LLMs’ zero-shot generalization and reasoning ability to design a feasible plan for robots to execute. Of note, PaLM-E [4] develops an embodied multi-modal language model to solve a broad range of tasks including robotic planning, visual question answering, and captioning; this large model serves as a foundation for robotic tasks. VLP [5] further enables visual planning for complex long-horizon tasks by pretraining on internet-scale videos and images. Code-As-Policies [17] re-purposes a code-writing LLM to generate robot policy code given natural language commands; it formulates task planning as an in-context code generation and function call problem. ITP [16] further proposes a simple framework to perform interactive task planning with language models, improving upon Code-As-Policies. Inspired by these works, our method also leverages LLMs for autonomous driving. 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. Most autonomous driving pipelines consist of perception, prediction, planning, and control, which have been significantly advanced by machine learning and deep neural networks in recent years. Despite such tremendous progress, both perception and planning are generally non-adaptive, preventing AVs from generalizing to any in-the-wild domain. Recent works leverage foundation models to provide autonomous driving pipelines with common sense reasoning ability. Wang et al. [31] proposes a method to extract nuanced spatial (pixel/patch-aligned) features from Transformers to enable the encapsulation of both spatial and semantic features. GPT-Driver [21] finetunes GPT-3.5 to enable motion planning and provide chain-of-thought reasoning for autonomous driving. DriveGPT4 [34] further formulates driving as an end-to-end visual question answering prob-

lem. Most recently, MotionLM [27] represents continuous trajectories as sequences of discrete motion tokens and casts multi-agent motion prediction as a language modeling task over this domain. Our work also leverages LLMs for policy adaption, however, we do not fine-tune or train a new foundation model. Instead, our method capitalizes on GPT-4 to perform direct in-context reasoning.

In parallel, there has been a plethora of literature on AV out-of-domain (OoD) generalization and detection [1, 8–10, 12, 13, 22, 30]. However, the vast majority of such works focus on low-level tasks (e.g., transferring perception models to data from different sensor configurations [1], adapting prediction methods to behaviors from different regions [13], etc.) and less on higher-level semantic generalization [6], which our work focuses on.

3. Driving Everywhere with Large Language Model Policy Adaptation

In this section, we will introduce our method, LLaDA, for adapting motion plans to traffic rules in new geographical areas and discuss all its building blocks.

LLaDA receives four inputs, all in the form of natural language: (i) a nominal execution plan, (ii) the traffic code of the current location, (iii) a description of the current scene from the ego’s perspective, and (iv) a description of any “unexpected” scenario that may be unfolding. LLaDA ingests these four inputs and outputs a motion plan – also represented in natural language – that addresses the scenario by leveraging the local traffic rules. The nominal execution plan can be generated by a human driver. Similarly, the scene description and the unexpected scenario description can be generated by a human or a VLM. The unique traffic code in the current location is the text of the entire driver handbook that describes the rules of the road for that location. Under normal circumstances, the unexpected situation input defaults to `normal status`; however, if something out-of-the-ordinary unfolds, such as the ego vehicle getting honked or flashed at, or if the ego driver notices something unusual in the environment (e.g., an animal on the road), the appropriate text description of the scenario can be supplied to LLaDA. To make the role of LLaDA more concrete, consider an example: An AV is operating in New York City (NYC) and the nominal motion plan for the vehicle is to turn right at a signalized intersection with a red light. The AV was honked at by cross traffic which is unexpected. LLaDA will take these inputs along with NYC’s driver manual and adapt the motion plan to `no right turn on a red light` because NYC traffic law prohibits right turns on red [23]. In the remainder of this section, we will discuss the building blocks of LLaDA, illustrated in Figure 2.

Traffic Rule Extractor. Passing the entire driver handbook to the LLM is superfluous as we only need the traffic

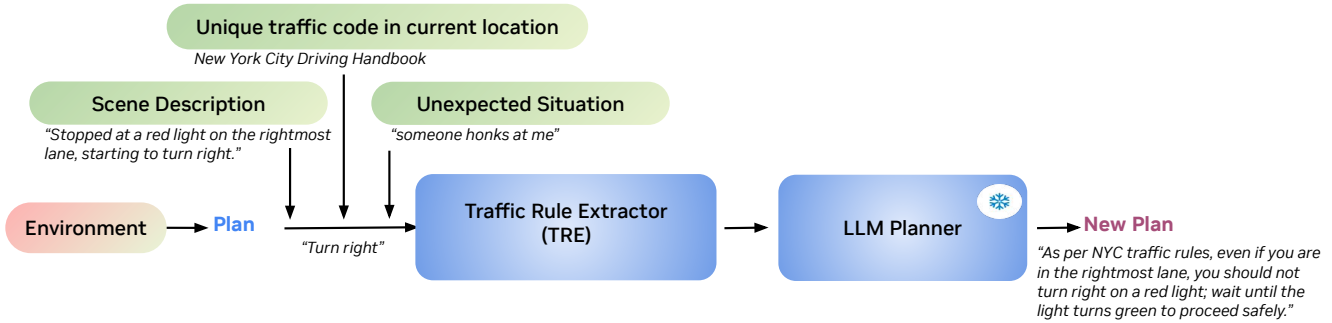


Figure 2. Overview of LLADA. In this illustration, the driver learned how to drive in California but now needs to drive in New York City. However, the road situation, traffic code, and unexpected situations are different. In our system, we consider three inputs: initial plan (“Turn right”), unique traffic 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.

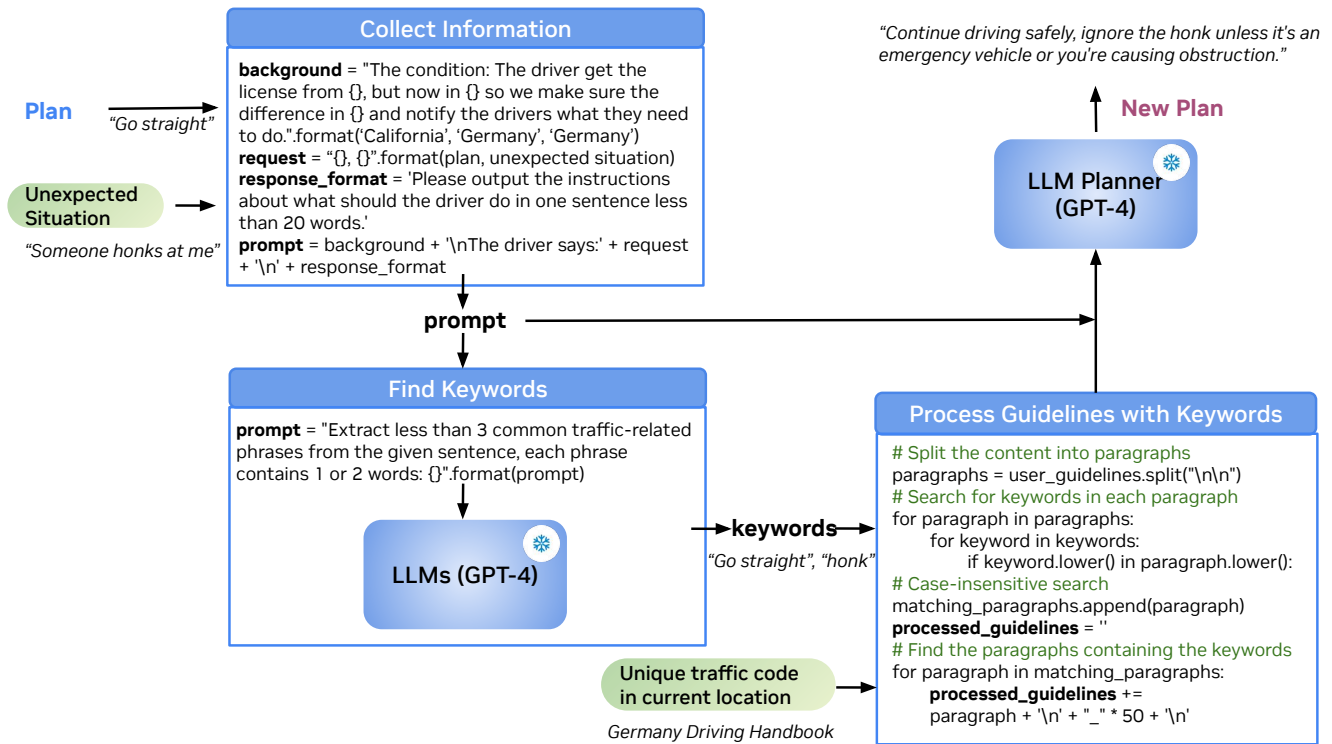


Figure 3. Details of Traffic Rule Extractor (TRE). As is shown in the figure, we first organize the information (such as locations, “Turn right” and “someone honks at me”) into a prompt. Then we feed the prompt to find the one or two keywords using GPT-4. To guarantee the search quality, each keyword contains one or two words. Then we find the key paragraphs that contain extracted keywords in the unique traffic code. In this way, we could filter out the necessary information and only organize the valuable material into GPT-4 to obtain the final new plan.

rules relevant to the current scenario that the vehicle finds itself in. In fact, extraneous information of the traffic code can hurt the LLM Planner’s performance. To achieve this *task-relevant* traffic rule extraction we use the Traffic Rule Extractor (TRE). TRE uses the nominal execution plan and the description of the unexpected scenario to extract key-

words in the traffic code of the current location, which are further used to extract paragraphs that comprise these keywords. We use Traffic Rule Extractor (TRE) to identify the most relevant keywords and paragraph extraction; see **Figure 3** for an illustration of TRE’s operation. We could observe that TRE is simple yet efficient in extracting key

paragraphs from the unique traffic code, it first generates a prompt and finds keywords in the organized prompt using GPT-4. Then we find the keywords in the unique traffic code in the current location. By organizing the processed guidelines and prompt, we can obtain a new plan accurately by using GPT-4 twice. After obtaining the relevant paragraph, we input the organized information from TRE into an LLM (GPT-4) to obtain the final new plan, referred to as the LLM Planner.

4. Applications of LLaDA

LLaDA is a general purpose tool for seamlessly adapting driving policies to traffic rules in novel locations. We see two main applications that can benefit from LLaDA:

Traffic Rule Assistance for Tourists. Standalone, LLaDA can serve as a guide for human drivers in new locations. We envision an interface wherein a human driver, when encountered with an unexpected scenario, can query LLaDA in natural language via a speech-to-text module on how to resolve it. As described in Section 3, LLaDA can take this natural language description of the scene, the unexpected scenario, and the nominal execution plan and provide a new plan which adheres to the local traffic laws. It is worth pointing out that in its current form, LLaDA cannot provide plan corrections unless queried by the human driver. This limits its usability to scenarios where the human driver becomes aware that they are in an unexpected scenario. Extending LLaDA to automatically provide plan corrections requires the development of an *unexpected scenario detector and translator*, which is beyond the scope of this current work and will be explored as part of our future work. We conducted a survey to garner human feedback about the usefulness and accuracy of LLaDA in some challenging traffic rule scenarios – the results are discussed in Section 5.

AV Motion Plan Adaptation. We can also leverage LLaDA’s traffic law adaptation ability in an AV planning stack to automatically adapt AV plans to the rules of a new geographical location. This can be achieved by interfacing LLaDA with any motion planner capable of generating high-level semantic descriptions of its motion plan (e.g., GPT-driver [21]) and a VLM (e.g., GPT-4V) that can translate the scene and the unexpected scenario into their respective textual descriptions. LLaDA then adapts the nominal execution plan and communicates it to a downstream planner that updates the low-level waypoint trajectory for the AV. Our approach for using LLaDA to adapt AV motion plans is summarized in Figure 4. We demonstrate the benefits that LLaDA can deliver to AV planning in our experiments where a nominal planner trained in Singapore is deployed in Boston; more details regarding the experiments are provided in Section 5.

5. Experiments

5.1. Implementation Details

Since LLaDA takes advantage of large pre-trained language models, our method is training-free and easily be applied to any existing driving system. LLaDA could be run with a single CPU. In this paper, we assume the driver obtains driver’s license from California as the default setting.

5.2. LLaDA Examples

We show a full set of functions of LLaDA in our [website](#). LLaDA enables the system to provide the most updated instructions based on local traffic rules, we show the basic functions of LLaDA and display how it works when the drivers are in different places, in diverse unexpected situations, or with diverse plans under various environments. We could observe that LLaDA is robust to distinct conditions. We also notice that without the driving handbook, the model cannot provide accurate information. We assume this is because GPT-4 may not be able to provide detailed instructions without the context or complex prompt tuning, while LLaDA could successfully alleviate this problem and generate reasonable instructions with emphasis on the specific local traffic rule and driver’s request.

5.3. Inference on Random Nuscenes/Nuplan Videos

Nuscenes [2] and Nuplan [3] datasets are two of the most used dataset for autonomous driving. Nuscenes is the first dataset to carry the full autonomous vehicle sensor suite and Nuplan is the world’s first closed-loop ML-based planning benchmark for autonomous driving. However, Nuscenes only contains 12 simple instructions such as “accelerate” and Nuplan only has 74 instructions (scenario types) such as “accelerating at stop sign no crosswalk”, which may not provide constructive and efficient instructions for drivers in different locations. LLaDA could successfully address this problem and can be applied to random videos. We first show Nuscenes example in Figure 5. We also show Nuplan examples in Figure 6. It is obvious that LLaDA works for random videos under diverse scenarios, achieving driving **everywhere** with language policy.

5.4. Challenging Situations

To further verify the efficiency of LLaDA, we consider several challenging cases and compare the results with and without our approach. Also, since GPT-4 could translate different languages at the meanwhile, LLaDA is able to process different language inputs and output the corresponding instructions (See row 5). We display the results in Table 1. In example 1, in NYC there is no right-turn on red, which is allowed in San Francisco. In example 2, LLaDA can point out something relating to Rettungsgasse (move to the right). This is because in the US the rule is that everyone

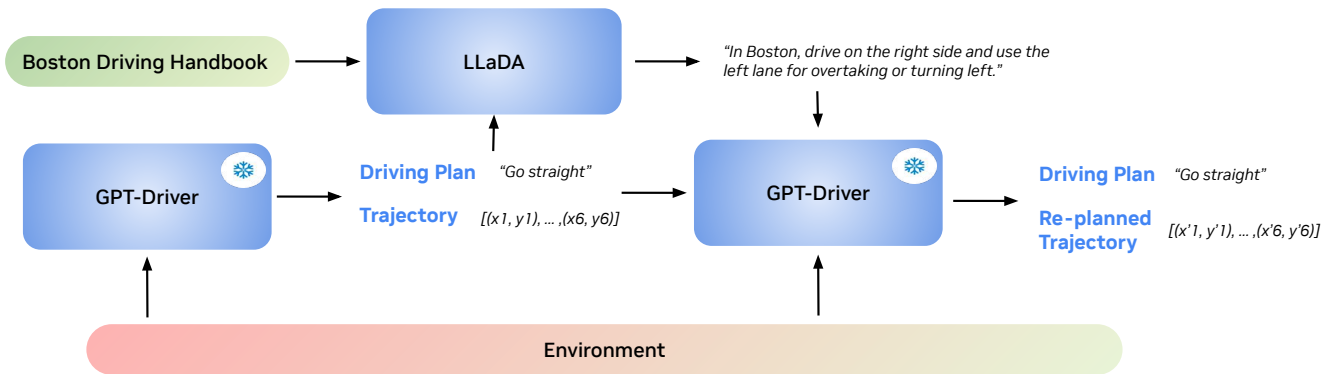


Figure 4. Combining LLaDA with GPT-Driver for motion planning on the nuScenes dataset.

Plan	Unique Traffic Code	Unexpected Situation	Outputs (New Plan)
 Change to the left lane	Singapore Driving Handbook	normal status	Slow down, signal, and carefully change to the left lane if it's safe and unoccupied.
 Go straight with constant speed		found trash on the road	Stay in your lane, slow down, and safely navigate around the trash without changing lanes or stopping
 Turn left with constant speed		normal status	Continue driving straight at a consistent speed, observing local speed limits and traffic signs.
		a car suddenly appears	Slow down, check mirrors, and prepare to brake if necessary.
		normal status	Signal in good time, keep to the left in Singapore, and do not swing out to the right while turning left.
		a car suddenly appears	Slow down, signal your intention, and turn left only when it's safe to do so.

Figure 5. Inference on a random Nuscenes video in Singapore. Since different from most countries, Singapore has a right-hand traffic system, which requires distinct behavior in comparison with California’s (left-hand traffic system). For example, when “Turn left with constant speed” and “normal status”, LLaDA suggest the driver to “keep to the left in Singapore”.

pulls over to the right, but this is not standard in Germany. In example 3, LLaDA is able to point out that we should overtake the slow car safely from the right lane, since overtaking on the left is illegal in London. For example 4, our system could point out that an unprotected right in England (left-driving system) requires checking the traffic coming at you as you will have to cut through upcoming traffic. Both should mention checking for pedestrians. For example 5, since the driver is on the Autobahn, where being in the left lane typically requires driving at very high speeds compared to American or many other countries’ highway speed limits. On the German Autobahn, the absence of a speed limit means that drivers instead adopt different “speed zones” per

lane, with the leftmost being the fastest and the rightmost being the slowest. For example 6, Amish communities exist in the US and Canada (primarily in the northeast USA), and they frequently have horse-pulled carriages on roads. So our system successfully provides the instructions to give right-of-way to horses.

5.5. Evaluator-based Assessment.

We conducted an evaluator-based assessment to further validate the usefulness of videos generated with LLaDA. Here we show the corresponding questionnaire, we list the questions in [this Google Form](#). We provided location, scenario, unexpected situation, relevant local law as conditions, and

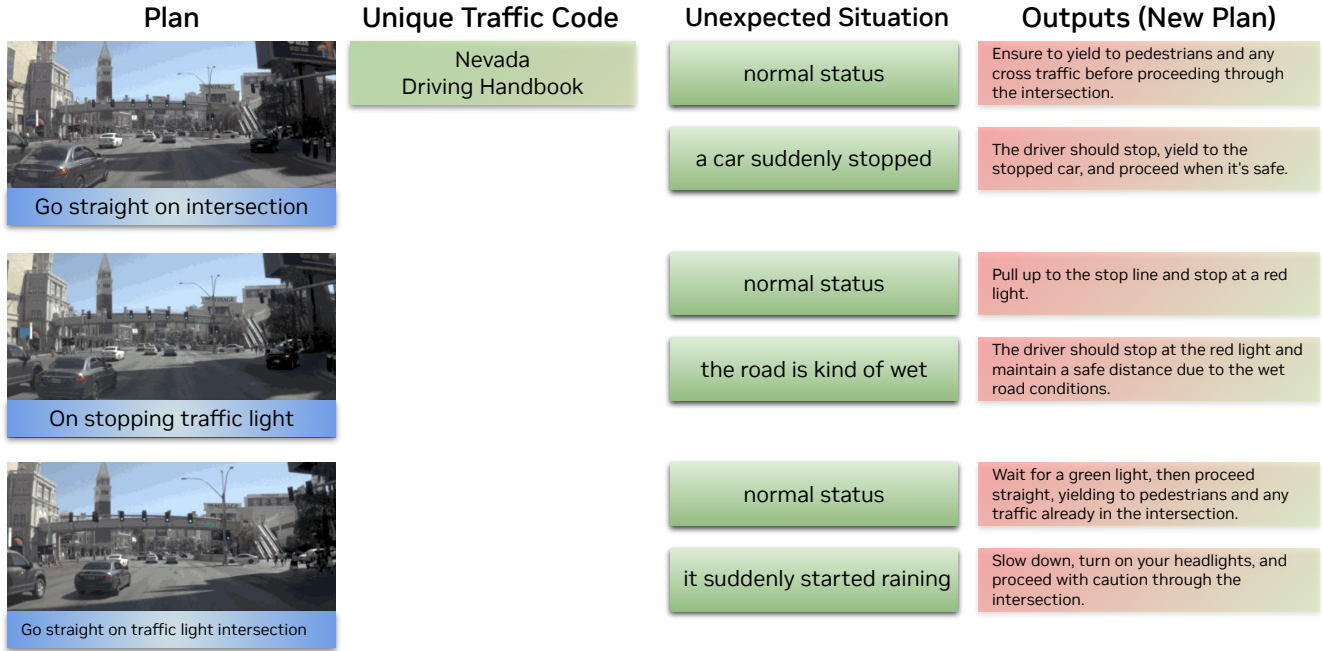


Figure 6. Inference on a random NuPlan video in Las Vegas. For this case, we select Nevada driving handbook, the illustration shows that LLaDA works very well in NuPlan and can provide accurate new plan with various unexpected situations.

Example	Original Location	Target Location	Plan	Unexpected Situation	Outputs (New Plan)
1	San Francisco NYC	NYC San Francisco	Turn right on red Turn right on red	normal status normal status	Do not turn right on red in NYC unless a sign permitting it is posted. Stop completely, yield for pedestrians and turn right if there's no "No Turn on Red" sign.
2	California	Germany	Drive straight on the highway	an emergency vehicle is approaching from behind	Move to the right and allow the emergency vehicle to pass.
3	NYC	London	Drive straight on the highway	the car in front drives very slowly, we are in the middle lane of a 3 lane highway	Overtake the slow car safely from the right lane, as overtaking on the left is illegal in London.
4	California	Singapore	Unprotected right	normal status	Yield to all other traffic and pedestrians before making your right turn.
5	California	Germany	Drive straight on the highway, in the leftmost lane	I keep getting honked at by cars behind me	Move to the right lane, the leftmost lane in Germany is for overtaking and faster vehicles.
6	California	Ontario	Driving on a rural two-lane road	there's a horse pulling a carriage	The driver should slow down, pass the carriage cautiously, and give plenty of space to the horse.

Table 1. Inference on various challenging examples. We compare the outputs with and without LLaDA, and it is obvious that LLaDA significantly improves the plan and successfully handles various difficult tasks under diverse scenarios and environments.

LLaDA output as driving assistant instructions. Here is the relevant local law that indicates what we want to pay attention to while driving. We asked two questions to 24 participants about each of the 8 cases. 54.2% participants have more than 10 years of driving experience and 20.8% participants have 5-10 years driving experience. Also, 75% participants are from the United States. In our assessment, we ask two questions: “Does the instruction follow the relevant local law?” and “How useful is the instruction?”. The results show that 70.3% participants think the instructions strictly follow the relevant local law, and 82.8% participants find the instructions are very or extremely helpful for them. This highlights that LLaDA brings about a significant enhancement in the performance of baseline video diffusion models in both the alignment and visual quality.

5.6. Comparison on Motion Planning

We conduct experiments on the nuScenes dataset to validate the effectiveness of LLaDA in motion planning. NuScenes consists of perception and trajectory data collected from Singapore and Boston, which have different traffic rules (e.g., driving side difference). Specifically, we first utilize GPT-Driver [21] to generate an initial driving trajectory for a particular driving scenario, and then we leverage LLaDA to generate guidelines for GPT-Driver to re-generate a new planned driving trajectory. Since LLaDA provides country-specific guidelines, we fine-tuned the GPT-Driver on the *Singapore* subset of the nuScenes dataset and evaluated the performances of GPT-Driver and LLaDA on the *Boston* subset of the nuScenes validation set. We follow [11, 21] and leverage L2 error (in meters) and collision rate (in percentage) as evaluation metrics. The average L2 error is computed by measuring each waypoint’s distance in the planned and ground-truth trajectories. It reflects the prox-



Figure 7. Visualization of the motion planning results on the nuScenes Boston subset. Ground truth trajectories are in **green**. Trajectories generated by GPT-Driver are in **red**. Re-planned trajectories by LLaDA are in **purple**.

imity of a planned trajectory to a human driving trajectory. The collision rate is computed by placing an ego-vehicle box on each waypoint of the planned trajectory and then checking for collisions with the ground truth bounding boxes of other objects. It reflects the safety of a planned trajectory. We follow the common practice in previous works and evaluate the motion planning result in the 3-second time horizon. **Table 2** shows the motion planning results. With the guidelines provided by LLaDA, GPT-Driver could adapt its motion planning capability from Singapore to Boston and reduce planning errors.

5.7. Ablation Study on Potential Safety Issues

There might be concerns that since LLaDA is based on LLMs, it might generate prompts that might provide dangerous instructions. To alleviate this concern, we evaluate LLaDA on three different critical cases with diverse countries and unexpected situations: 1) when facing a stop sign, whether LLaDA suggests “stop” or not. 2) When facing the red light, whether LLaDA suggests “stop”/specific safe notifications or not. 3) When it rains heavily, whether LLaDA suggests “slow down” or not. 4) When a pedestrian walks across the street, whether LLaDA suggests to “yield to the pedestrian” or not. For each case, we evaluate 50 random examples and report the average score of each case. For each example, if the answer will cause potentially dangerous behavior, we will treat it as an error. We observe that LLaDA achieves 0% error rate for all 4 cases. In case “stop sign”, where all the instructions suggest “Wait at the stop sign until it’s safe”. In case “red light”, where all the instructions suggest “come to a complete stop at the red light” or “wait until it turns green”. In case “it rains heavily”, where all the instructions suggest “turn on the headlight” and “reduce speed”. In case “A pedestrian walks across

the street”, where all the instructions suggest “yield to the pedestrian”. We didn’t notice any potentially harmful instructions that might cause dangerous behavior, which ensures user safety by directing them toward appropriate behavior.

5.8. Combining with GPT-4V

Our approach can be combined with different systems to improve the functionality to provide accurate information for humans. In this section, we study the case by combining LLaDA with GPT-4V. GPT-4V is the vision branch of GPT-4 which can output corresponding captions to describe the scenes based on a visual input. We randomly pick two scenes from Youtube ¹ and ask GPT-4V to provide additional captions, we add these additional captions to the user’s request. We show an example in **Figure 8**, it could be observed that LLaDA could process the information GPT-4V very well and provide accurate instructions based on given captions.

6. Conclusion, Limitations, and Future Work

Conclusion. In this work, we proposed LLaDA, an LLM-powered framework that adapts nominal motion plans by a human driver or an AV to local traffic rules of that region. The modularity of LLaDA facilitates its use for human driver assistance as well as AV plan adaptation. To our knowledge, LLaDA is the first to propose traffic rule-based adaptation via LLMs. Our results show that human drivers find LLaDA to be helpful for driving in new locations, and LLaDA also improves the performance of AV planners in

¹Source:<https://www.youtube.com/watch?v=Boh66Pjjiq0> and https://www.youtube.com/watch?v=xn_mSqTrOUo

Method	L2 (m) ↓				Collision (%) ↓			
	1s	2s	3s	Avg.	1s	2s	3s	Avg.
GPT-Driver [21]	0.27	0.59	1.04	0.63	0.22	0.45	1.07	0.58
GPT-Driver [21] + LLaDA Re-Planning (Ours)	0.27	0.58	1.02	0.62	0.22	0.41	1.04	0.56

Table 2. Motion planning results. With the guidelines provided by LLaDA, GPT-Driver could adapt its motion planning capability from Singapore to Boston and reduce planning errors.





Plan	GPT-4V Scene Description	Unique Traffic Code	Unexpected Situation	Outputs (New Plan)
 Already in the left lane and go straight	A bridge stretches ahead with light traffic, leading towards a modern city skyline under a clear sky.	China Driving Handbook	normal status a car suddenly appears from right	Continue straight in the left lane, obeying local speed limits and traffic signals. Slow down, check your mirrors, and yield to the car coming from the right.
 Go straight with constant speed	Driving steadily, a car approaches a well-lit tunnel, bordered by concrete walls, under a clear blue sky.		normal status a big truck suddenly appears	Continue driving straight at a consistent speed, observing local speed limits and traffic signs. Slow down, maintain distance from the truck, and proceed with caution.
 Draai linksaf met constante snelheid	A bustling city street in Amsterdam is filled with pedestrians, vehicles, and cyclists, with the prominent "De Bijenkorf" department store in the background.	Netherlands Driving Handbook	normal status the weather turns cloudy	The driver should turn left at a constant speed. Turn left at a constant speed, be cautious due to cloudy weather.
 Vertragen en omdraaien	Construction workers operate machinery amid roadwork while pedestrians navigate the busy surrounding streets of a European city.		normal status a car suddenly appears	The driver should slow down and turn around. Slow down, signal, check mirrors and blind spots, then carefully turn around when safe.

Figure 8. Combining LLaDA with GPT-4V and inference on random YouTube scenes. In the first two rows, we show the scenes of Chongqing in China. In the rest two rows, we show the scenes of the crowded Amsterdam Center in Netherlands. As for the plans, we input Dutch instead of English to verify the capability of LLaDA to interact with drivers from various cultures automatically without mentioning the language type of the input. “Draai linksaf met constante snelheid” represents “Turn left with constant speed” and “Vertragen en omdraaien” represents “decelerate and turn around”. We asked the system to output English instructions, we noticed that the system extracts keywords in English and is able to provide accurate plans.

new locations.

Limitations. Though LLaDA provides various benefits, it also suffers from two limitations: first, since LLaDA requires running an LLM in the control loop, the runtime for LLaDA is not yet conducive for closed-loop use in an AV planning stack – this limitation is shared by all LLM-based motion planners. Second, as discussed in our results earlier, LLaDA is sensitive to the quality of scene descriptions. Although GPT-4V can provide such descriptions, they are sometimes not sufficiently accurate. This limitation points towards the need to develop an AV-specific foundation model that can provide AV-centric scene descriptions.

Broader Impact. As a human driver assistant, we hope that LLaDA would reduce the number of road accidents induced by tourists oblivious of the local traffic rules. As a

policy adapter for AVs, we expect LLaDA to pave the way towards easier traffic rule adaptation for AVs allowing them to expand their operations beyond geo-fenced regions.

Future Work. For future work, there are various directions we are excited to pursue: first, we will explore improving GPT-4V’s scene descriptions by fine-tuning it on AV datasets. Second, we will explore the development of an unexpected scenario detector which will allow us to use LLaDA only when it is needed, thereby significantly alleviating the computational burden involved in running an LLM-based module in the control loop. Finally, we will work towards furnishing safety certificates for the LLM outputs by leveraging recent developments in uncertainty quantification and calibration techniques for ML, such as conformal prediction and generalization theory.

References

- [1] Alexander Amini, Tsun-Hsuan Wang, Igor Gilitschenski, Wilko Schwarting, Zhijian Liu, Song Han, Sertac Karaman, and Daniela Rus. Vista 2.0: An open, data-driven simulator for multimodal sensing and policy learning for autonomous vehicles. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 2419–2426. IEEE, 2022. 2
- [2] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. 1, 4
- [3] Holger Caesar, Juraj Kabzan, Kok Seang Tan, Whye Kit Fong, Eric Wolff, Alex Lang, Luke Fletcher, Oscar Beijbom, and Sammy Omari. nuplan: A closed-loop ml-based planning benchmark for autonomous vehicles. *arXiv preprint arXiv:2106.11810*, 2021. 4
- [4] Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. PaLM-E: An embodied multimodal language model. In *Int. Conf. Mach. Learn.*, 2023. 2
- [5] Yilun Du, Mengjiao Yang, Pete Florence, Fei Xia, Ayzaan Wahid, Brian Ichter, Pierre Sermanet, Tianhe Yu, Pieter Abbeel, Joshua B. Tenenbaum, Leslie Kaelbling, Andy Zeng, and Jonathan Tompson. Video language planning, 2023. 2
- [6] Amine Elhafsi, Rohan Sinha, Christopher Agia, Edward Schmerling, Issa AD Nesnas, and Marco Pavone. Semantic anomaly detection with large language models. *Autonomous Robots*, pages 1–21, 2023. 1, 2
- [7] Klemens Esterle, Luis Gressenbuch, and Alois Knoll. Formalizing traffic rules for machine interpretability. In *Proceedings of the Connected and Automated Vehicles Symposium (CAVS)*, pages 1–7. IEEE, 2020. 2
- [8] Alec Farid, Sushant Veer, Boris Ivanovic, Karen Leung, and Marco Pavone. Task-relevant failure detection for trajectory predictors in autonomous vehicles. *arXiv preprint arXiv:2207.12380*, 2022. 2
- [9] Angelos Filos, Panagiotis Tigkas, Rowan McAllister, Nicholas Rhinehart, Sergey Levine, and Yarin Gal. Can autonomous vehicles identify, recover from, and adapt to distribution shifts? In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 3145–3153. PMLR, 2020.
- [10] Dan Hendrycks, Steven Basart, Mantas Mazeika, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling out-of-distribution detection for real-world settings. *arXiv preprint arXiv:1911.11132*, 2019. 2
- [11] Shengchao Hu, Li Chen, Penghao Wu, Hongyang Li, Junchi Yan, and Dacheng Tao. St-p3: End-to-end vision-based autonomous driving via spatial-temporal feature learning. In *European Conference on Computer Vision (ECCV)*, 2022. 6
- [12] Masha Itkina and Mykel Kochenderfer. Interpretable self-aware neural networks for robust trajectory prediction. In *Conference on Robot Learning*, pages 606–617. PMLR, 2023. 2
- [13] Boris Ivanovic, James Harrison, and Marco Pavone. Expanding the deployment envelope of behavior prediction via adaptive meta-learning. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2023. 2
- [14] Jesper Karlsson and Jana Tumova. Intention-aware motion planning with road rules. In *Proceedings of the International Conference on Automation Science and Engineering (CASE)*, pages 526–532. IEEE, 2020. 2
- [15] Markus Koppenborg, Peter Nickel, Birgit Naber, Andy Lungfiel, and Michael Huelke. Effects of movement speed and predictability in human–robot collaboration. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 27(4):197–209, 2017. 1
- [16] Boyi Li, Philipp Wu, Pieter Abbeel, and Jitendra Malik. Interactive task planning with language models. *arXiv preprint arXiv:2310.10645*, 2023. 1, 2
- [17] Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. Code as policies: Language model programs for embodied control. In *arXiv preprint arXiv:2209.07753*, 2022. 1, 2
- [18] Sebastian Maierhofer, Anna-Katharina Rettinger, Eva Charlotte Mayer, and Matthias Althoff. Formalization of interstate traffic rules in temporal logic. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, pages 752–759. IEEE, 2020. 2
- [19] Kumar Manas and Adrian Paschke. Legal compliance checking of autonomous driving with formalized traffic rule exceptions. In *Proceedings of the Workshop on Logic Programming and Legal Reasoning in conjunction with 39th International Conference on Logic Programming (ICLP)*, 2022. 2
- [20] Kumar Manas, Stefan Zwicklbauer, and Adrian Paschke. Robust traffic rules and knowledge representation for conflict resolution in autonomous driving. In *Proceedings of the 16th International Rule Challenge and 6th Doctoral Consortium @ RuleML+RR*, 2022. 2
- [21] Jiageng Mao, Yuxi Qian, Hang Zhao, and Yue Wang. GPT-Driver: Learning to drive with GPT. *arXiv preprint arXiv:2310.01415*, 2023. 2, 4, 6, 8
- [22] Rowan McAllister, Gregory Kahn, Jeff Clune, and Sergey Levine. Robustness to out-of-distribution inputs via task-aware generative uncertainty. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 2083–2089. IEEE, 2019. 2
- [23] New York City Department of Transportation. Right turn on red in staten island. <https://portal.311.nyc.gov/article/?kanumber=KA-01354>. 2
- [24] OpenAI. Gpt-4 technical report, 2023. 1
- [25] Andreas Psarras, Theodore Panagiotidis, and Andreas Anagnostidis. Covid-19, tourism and road traffic accidents: Evidence from greece. *Journal of Transportation Safety & Security*, pages 1–21, 2023. 1

- [26] Jaume Rosselló and Oscar Saenz-de Miera. Road accidents and tourism: the case of the balearic islands (spain). *Accident Analysis & Prevention*, 43(3):675–683, 2011. 1
- [27] Ari Seff, Brian Cera, Dian Chen, Mason Ng, Aurick Zhou, Nigamaa Nayakanti, Khaled S. Refaat, Rami Al-Rfou, and Benjamin Sapp. Motionlm: Multi-agent motion forecasting as language modeling, 2023. 2
- [28] Bay City News Service and Embarcadero Staff. Australian tourist charged with manslaughter in wrong-way crash that killed two La Honda residents. <https://www.paloaltoonline.com/news/2023/09/07/australian-tourist-charged-with-manslaughter-in-wrong-way-crash-that-killed-two-la-honda-residents>, 2023. 1
- [29] Sushant Veer, Karen Leung, Ryan Cosner, Yuxiao Chen, Peter Karkus, and Marco Pavone. Receding horizon planning with rule hierarchies for autonomous vehicles. *arXiv preprint arXiv:2212.03323*, 2022. 2
- [30] Sushant Veer, Apoorva Sharma, and Marco Pavone. Multi-predictor fusion: Combining learning-based and rule-based trajectory predictors. *arXiv preprint arXiv:2307.01408*, 2023. 2
- [31] Tsun-Hsuan Wang, Alaa Maalouf, Wei Xiao, Yutong Ban, Alexander Amini, Guy Rosman, Sertac Karaman, and Daniela Rus. Drive anywhere: Generalizable end-to-end autonomous driving with multi-modal foundation models. *arXiv preprint arXiv:2310.17642*, 2023. 2
- [32] Wayve. Lingo-1: Exploring natural language for autonomous driving. <https://wayve.ai/thinking/lingo-natural-language-autonomous-driving/>, 2023. 1
- [33] Wei Xiao, Noushin Mehdipour, Anne Collin, Amitai Y Bin-Nun, Emilio Frazzoli, Radboud Duintjer Tebbens, and Calin Belta. Rule-based optimal control for autonomous driving. In *Proceedings ACM/IEEE Int. Conf. on Cyber-Physical Systems*, pages 143–154, 2021. 2
- [34] Zhenhua Xu, Yujia Zhang, Enze Xie, Zhen Zhao, Yong Guo, Kenneth KY Wong, Zhenguo Li, and Hengshuang Zhao. Drivegpt4: Interpretable end-to-end autonomous driving via large language model. *arXiv preprint arXiv:2310.01412*, 2023. 2
- [35] Yu Zhang, Sarath Sreedharan, Anagha Kulkarni, Tathagata Chakraborti, Hankz Hankui Zhuo, and Subbarao Kambhampati. Plan explicability and predictability for robot task planning. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2017. 1