3. Research Goal

In this project, we aim to develop an AI-based motion planner utilizing LLMs to guide an autonomous vehicle in navigation tasks. The main research goal is to **Design an LLM-Based Motion Planner for Safe Decision-Making**. To achieve this goal, we seek answers to the following research questions:

- *Vehicle Perception:* How can we effectively interpret environmental observations so that the LLM can accurately understand the driving environment.
- *LLM Integration:* How to integrate an LLM and adapt it into the motion planning task?
- *Real-world Evaluation:* How to effectively evaluate the performance of the proposed motion planner in real-world driving environments?

We propose the following research tasks and present technical details to tackle the above research questions.

4. Research Plan

Figure 1 presents an overview framework of the proposed LLM-based motion planner, which aims to generate safe and effective driving actions $\mathcal{T}$ based on current observations $\mathcal{O}$ and prior successful experiences $\mathcal{H}$. The output of the motion planner will also update the memory module. Mathematically, our goal is to find a function F (i.e., LLM-based motion planner) such that

$$\mathcal{T} = F(\mathcal{O}, \mathcal{H}).$$

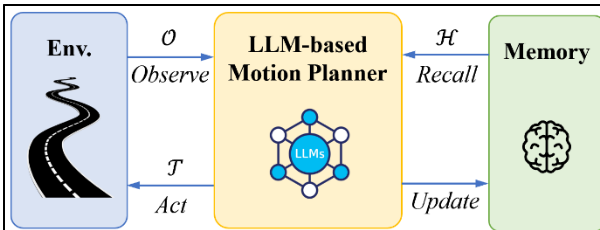

Figure 1 An Overview Framework of the Proposed Motion Planner

3.1 Task 1: Encode the Driving Environment via a *Scene Describer*

To enable an LLM effectively understand and interpret the current driving environment into textual descriptions, we propose a design of a *Scene Describer* to transcribe real-time driving scenarios into structured descriptive texts. The *Scene Describer* utilizes a standardized sentence structure to articulate details about the objects surrounding the ego-vehicle and their potential impacts on it.

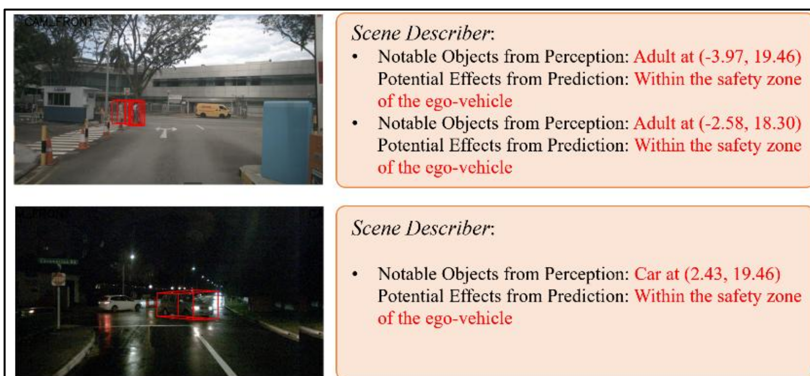

Figure 2 Example outputs from the *Scene Describer.*

Figure 2 presents two example outputs from the *Scene Describer*. On the left are visual observations from the vehicle's front-view camera, while the corresponding textual descriptions on the right are generated by the *Scene Describer*. In these examples, notable objects in the vicinity of the ego-vehicle are identified using a pre-trained object detector, i.e., Mask R-CNN [6]. Their future movements and potential interactions with the ego-vehicle are then estimated using a motion predictor, i.e., UniAD [4]. By detailing the predicted trajectories and behaviors of surrounding objects, the *Scene Describer* provides the LLM with the necessary context to reason about ***potential interferences*** with the ego-vehicle. This process ensures that the motion planner can make informed decisions to maintain safety and efficiency in navigation, effectively bridging the gap between raw sensory data and the high-level reasoning required for autonomous driving.

Expected Outcomes from Task 1: **(i)** an effective computer vision model to detect objects surrounding the ego-vehicle; **(ii)** a *Scene Describer* that translates visual observations into textual descriptions.

3.2 Task 2: Integrate LLM into the Motion Planning Task

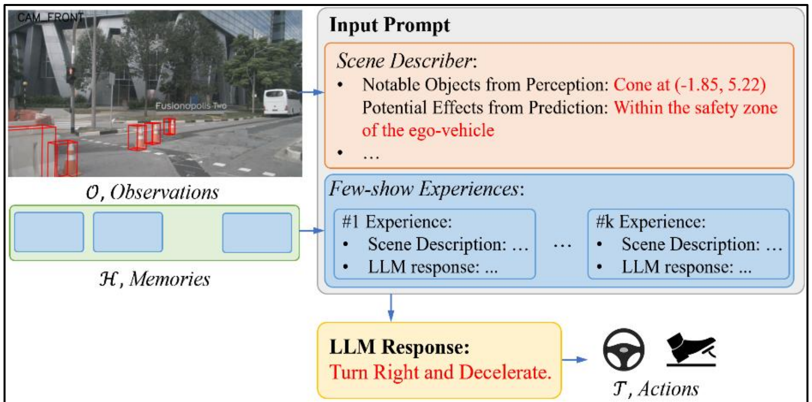

Figure 3 Detailed configuration of the LLM-based motion planner.

The core of this task involves adapting a pre-trained LLM, such as GPT-4, to handle the motion planning needs of autonomous driving. As shown in **Figure 3**, the **Input Prompt** comprises the current visual observation's description (generated by the *Scene Describer*) alongside few-shot experiences from the memory module, where each experience contains a textual scene description coupled with the LLM's response. The design of **Input Prompt** follows the Chain-of-Thought reasoning [7], where the LLM is guided to reason the current driving situation and produce a meta-action description that specifies the required driving directions (left or right) and speed adjustments (accelerate or decelerate). Then, the generated instructions are used to control the vehicle's actuators (i.e., steering wheel and gas/brake pedals). Successful actions are archived in the memory module for future reference, enriching the dataset used for further training and enhancing the LLM's performance over time.

Expected Outcomes of Task 2: **(i)** An effective motion planner centered on LLMs to dynamically control the vehicle to reach the destination safely.

3.3 Task 3: Implementation of the Proposed Motion Planner in an Indoor Testbed

Figure 4 An example indoor testbed.

The implementation and evaluation of our proposed motion planner will take place in an indoor testbed, developed at the PI's lab. Within this testbed, the PI will simulate an urban traffic scenario using mobile robots that represent autonomous vehicles (see **Figure 4**). Among these, one robot will be implemented with the proposed LLM-based motion planner, while the remaining robots will follow pre-defined driving routes to circle inside the testbed area. The mobile robot with the proposed motion planner is tested to loop inside the testbed, avoiding collisions with the other robots or obstacles.

The performance of our LLM-based motion planner will be compared with state-of-the-art motion planning algorithms [4, 8] from various metrics, such as the average collision rate and completion time of one loop. The results from these experimental trials will provide unique contributions towards the LLM-based motion planner research since most of existing LLM-based approaches were evaluated with simulators and existing datasets. This task not only tests the practical application of our theoretical model but also provides a robust platform for further enhancements based on real-world data and performance metrics.

Expected Outcomes of Task 3: **(i)** an indoor testbed that simulates real-world traffic situations; **(ii)** research papers that summarize the proposed LLM-based motion planner and experimental results in the indoor testbed.

3.4 Evaluation and Project Timeline

The success of this project will be evaluated based on the following criteria: **(i)** Successful demonstrations of the proposed motion planner in an indoor testbed, where the system effectively navigates dynamic