Drive as You Speak: Enabling Human-Like Interaction with Large Language Models in Autonomous Vehicles

Can Cui\(^1\)*, Yunsheng Ma\(^1\), Xu Cao\(^2\), Wenqian Ye\(^2,3\), Ziran Wang\(^1\)
1 Purdue University, West Lafayette, IN, USA, 47907
2 PediaMed.AI Lab, Shenzhen, China, 518048
3 University of Virginia, Charlottesville, VA, USA, 22903
\{cancui\}@purdue.edu

Abstract

The future of autonomous vehicles lies in the convergence of human-centric design and advanced AI capabilities. Autonomous vehicles of the future will not only transport passengers but also interact and adapt to their desires, making the journey comfortable, efficient, and pleasant. In this paper, we present a novel framework that leverages Large Language Models (LLMs) to enhance autonomous vehicles’ decision-making processes. By integrating LLMs’ natural language capabilities and contextual understanding, specialized tools usage, synergizing reasoning, and acting with various modules on autonomous vehicles, this framework aims to seamlessly integrate the advanced language and reasoning capabilities of LLMs into autonomous vehicles. The proposed framework holds the potential to revolutionize the way autonomous vehicles operate, offering personalized assistance, continuous learning, and transparent decision-making, ultimately contributing to safer and more efficient autonomous driving technologies.

1. Introduction

Recently, Large Language Models (LLMs) have attracted significant attention. The key to their success lies in their remarkable ability to process a wide range of word-based inputs, including prompts, questions, dialogues, and vocabulary spanning diverse domains, resulting in significant and coherent textual outputs. LLMs serve as vast stores of abundant information and knowledge acquired from numerous texts, much like the human brain. Considering the LLM’s ability to emulate the human brain functions, it prompts us to ask: could we leverage the impressive capabilities of LLMs to revolutionize the future of autonomous driving?

Imagine a situation where you’re sitting in an autonomous vehicle and you desire to safely overtake another vehicle. All you have to do is utter the command: “Overtake the vehicle in front of me.” At that point, the LLMs would swiftly assess the existing conditions and safety listen, and ask questions before reasoning, providing you with informed guidance on the feasibility and recommended actions for executing the maneuver. Furthermore, in the context of fully autonomous vehicles, the LLMs’ capabilities could even extend to taking charge of the vehicle and executing the instructed commands.

While LLMs have the potential to greatly enhance convenience and improve the driving experience for drivers, a significant challenge arises: LLMs lack understanding of information about the driving environment. Unlike humans, LLMs lack the inherent ability to perceive the physical environment. In other words, these models do not possess the capability to visually perceive and interact with the world around them [2]. It renders LLMs challenged in making sound decisions for the current situation, potentially leading to suboptimal outcomes or even hazardous consequences.

To address the challenge above, we present a perspective where LLMs can serve as the decision-making “brain” within autonomous vehicles. Complementing this, various tools within the autonomous vehicle ecosystem, including the perception module, localization module, and in-cabin monitor, function as the vehicle’s sensory “eyes.” This configuration enables LLMs to overcome the inherent limitation of not directly accessing real-time environmental information. By receiving processed data from the perception module, LLMs can facilitate informed decision-making, resulting in significant enhancements to the performance of the autonomous vehicle. Additionally, the vehicle’s actions and controller function as its “hands,” executing instructions derived from the LLM’s decision-making process.

When comparing autonomous vehicles with and without integrated LLMs, it becomes evident that the latter offers a diverse array of compelling advantages. These advantages extend across various aspects of functionality and perfor-
2. Perspective: the Role of LLMs in Advancing Autonomous Vehicles

As previously established in the earlier section, we’ve established that the LLMs serve as the “brain” in the autonomous driving system, facilitating driver interaction and decision-making, while the useful sensory tools and actuation function as the vehicle’s “eyes” and “hands” respectively. To be more specific, When a driver requests a particular operation, the LLM prompts the related modules to provide data that has been processed to extract relevant information from the environment. By integrating the linguistic analysis of LLMs with the processed sensory inputs from the selected modules, the LLM can then make well-informed decisions. If the command is deemed both feasible and safe based on the prior analysis, the LLMs will transmit the corresponding instructions to the vehicle's controller. This includes components such as the steering wheel, throttle pedal, braking, and other control elements, enabling them to execute the necessary operations. Alternatively, if the operation is deemed inappropriate, the LLMs will provide drivers with a detailed explanation as to why the requested action is not suitable for execution.

Revisiting the example at the beginning of this paper, when drivers issue the command to overtake the vehicle ahead, the LLMs come into play by querying the perception module for pertinent processed information. This includes details such as the distance and speed of the target vehicle, the velocity of the ego vehicle, road conditions of potential lanes, the presence of other vehicles and their distances on those lanes, and other useful navigation information from the map system. Through an analysis of the provided data and the given command, the LLMs make a decision regarding whether to execute the driver’s request. If the decision is affirmative, the LLMs subsequently communicate instructions to the controller, guiding the next course of action.

Having explored this intricate interaction between LLMs and the autonomous vehicle’s decision-making process, we shift our focus to a broader context and propose the concept of a human-centric LLMs integrated framework for autonomous vehicles based on our prior work of the mobility digital twin [30]. As shown in Fig. 1, the physical world comprises human drivers, vehicles, and traffic objects. In the physical world, human drivers are the central agents in the physical world, sending commands and instructions to LLMs as they navigate roadways. The traffic environment contains various elements including vehicles, pedestrians, traffic lights, road conditions, and traffic cones, all of which contribute to the complexity of movement and interactions on the road. The vehicle, directed by the LLMs, operates within this ecosystem, executing the commands it receives from either drivers or LLMs through controllers and actuators.

The virtual world includes LLMs, memory, and essential tools which include the perception module, localization module, and in-cabin monitor. The perception module acquires raw input from sensors, including external cameras, LIDARs, and radars [8, 9], and processes this data into a format suitable for the LLMs. The localization module employs GNSS data to determine the vehicle’s precise location. Within the vehicle, the in-cabin monitor employs internal cameras, thermometers, and other sensors to vigilantly observe the in-cabin environment, preempting distractions, extreme temperatures, or uncomfortable conditions. At the core of the entire framework lies the LLMs, serving as its central intelligence. They receive commands from drivers, subsequently initiating queries to pertinent modules for related information. Furthermore, the memory section acts as a repository, storing historical operations and drivers’ preferences, enabling continuous learning and enhancement for the LLMs. This repository of experiences equips the LLMs to make analogous decisions when confronted with similar situations, bolstering the system’s adaptability and performance over time. The memory also houses maps and local law information, empowering the LLMs to make even wiser decisions adaptable to a variety of scenarios.

3. Review: Can LLMs Really Do This?

Through a comprehensive review of both theoretical underpinnings and real-world implementations, we seek to
Figure 1. The human-centric LLM-integrated framework for autonomous vehicles.

address the fundamental question: Can LLMs really contribute to the improvement of autonomous driving by actively participating in the decision-making framework? By examining the current state of research and analyzing use cases, this section aims to provide a thorough assessment of the extent to which LLMs can bring to the landscape of human-centric autonomous driving.

3.1. Adaptive Techniques and Human-Centric Refinements for LLMs

Parameter-efficient fine tuning (PEFT) is a crucial technique used to adapt pre-trained language models (LLMs) to specialized downstream applications [6, 10, 12, 14, 18]. Hu et al. [14] proposed utilizing low-rank decomposition matrices to reduce the number of trainable parameters needed for fine-tuning language models. Lester et al. [18] explore prompt tuning, a method for conditioning language models with learned soft prompts, which achieves competitive performance compared to full fine-tuning and enables model reuse for many tasks. These PEFT techniques offer valuable tools for adapting LLMs to autonomous driving tasks.

Reinforcement Learning from Human Feedback (RLHF) [1, 21, 23, 24, 26] has emerged as a key strategy for fine-tuning LLM systems to align more closely with human preferences. Ouyang et al. [21] introduce a human-in-the-loop process to create a model that better follows instructions. Bai et al. [1] propose a method for training a harmless AI assistant without human labels, providing better control over AI behavior with minimal human input. These approaches hold significant promise for developing LLMs for autonomous driving applications, as they can contribute in two dimensions. Firstly, they can ensure that LLMs avoid making decisions that may be illegal or unethical. Secondly, these methodologies enable LLMs to continually adapt and align their decision-making processes with user preferences, enhancing personalization and trust in autonomous vehicles.

LLM-based autonomous driving applications can also benefit from advanced prompting techniques [3, 16, 29, 31, 32]. Chain-of-thought prompting [31] improves LLMs’ ability to perform complex reasoning. Gao et al. [13] propose an approach that uses LLMs to read natural language problems and generate programs as intermediate reasoning steps. Yao et al. [32] present a new prompting technique that allows LLMs to make decisions about how to interact with external APIs. These methods provide a solid foundation for the development of LLMs for autonomous driving applications with two significant benefits. Firstly, they greatly enhance LLMs’ reasoning capabilities, particularly in complex, multi-step scenarios. Secondly, these techniques improve the adaptability and versatility of LLMs, key attributes for autonomous driving systems interfacing with various tools and data sources.
3.2. Advancements in LLMs: Implications for Autonomous Driving Decision-Making

Recent Research has shown that LLMs can perform well in most commonsense tasks [4], which means it has the potential to make wise and feasible decisions in autonomous driving scenarios. The utilization of LLMs in the context of autonomous driving presents a captivating and potentially transformative direction for research. Recent investigations have brought light on the diverse ways in which LLMs can profoundly impact the landscape of autonomous vehicles. For instance, the study conducted by [19] highlights the promise of AI-infused with legal knowledge, offering the potential to avert legal transgressions in autonomous driving scenarios, thereby contributing to the establishment of a safer AI-driven environment. Additionally, [33] demonstrates that LLMs possess the capability to learn from local laws and accident reports, and effectively contribute to reducing accident rates, thus enhancing the safety of autonomous driving.

The application of LLMs to decision-making in autonomous driving is notably explored by [5]. Their research introduces the PaLM model, demonstrating that LLMs exhibit a capacity to effectively tackle intricate reasoning tasks and, intriguingly, surpass the performance of an average human. Such a finding carries significant implications, hinting at LLMs’ remarkable ability to navigate complex scenarios, make astute judgments, and potentially lay the groundwork for optimal decision-making in autonomous vehicles.

The work highlighted in [22] demonstrates the utilization of large language models to effectively store experiences in natural language, forming a foundational approach for integrating historical data into our architecture. The adaptive capabilities of LLMs are showcased in various ways. [17] underscores LLMs’ proficiency in zero-shot reasoning, enabling them to deal with novel and unfamiliar situations, a vital feature for autonomous vehicles operating in dynamic environments. The study by [6] exemplifies that LLMs can be fine-tuned to exhibit enhanced performance, particularly in tasks with limited training data.

Additionally, LLMs have shown great potential in both transportation and robotics areas, as highlighted by [7, 34], and [27] respectively. They reveal LLMs’ prowess in tasks such as zero-shot planning and interactive conversations, even facilitating interaction with perception-action-based API libraries, an attribute that aligns with the demands of autonomous vehicles.

Furthermore, the work [28] demonstrates LLMs’ potential for continuous learning, which is of paramount importance for adapting to evolving road conditions and enhancing performance over time.

The investigation from [11] introduces embodied language models capable of assimilating real-world sensor data, thus bridging the gap between perception and language. This development lays the foundation for potential advancements in autonomous vehicles, where LLMs could process sensory inputs, comprehend their surroundings, and consequently make more informed decisions. Building on these insights, additional studies [4], [35], [32], [25], and [15] have further enriched our understanding of LLMs’ capabilities, underscoring their potential in decision-making, reasoning, and synergizing reasoning and acting.

4. Experiment: Decision-Making and Motion Planning with ChatGPT-4

To gain a deeper understanding of the practical capabilities of LLMs in the context of autonomous driving tasks, we embark on an insightful exploration involving real-world decision-making scenarios. This comprehensive case study serves as a compelling demonstration of how LLMs can effectively enhance autonomous vehicles by harnessing the potential of ChatGPT-4 [20] to replicate decision-making processes. Our investigation is structured in two distinct phases. Initially, we pose autonomous driving concept-related queries to GPT-4, which unveils its grasp of how language models can be seamlessly integrated into autonomous driving. Subsequently, we design and present genuine real-world situations to assess the decision-making proficiency of LLMs. This section covers an in-depth understanding of this case study, including a detailed conversation with GPT-4 that highlights our findings. This analysis serves to underscore the practical implications of leveraging LLMs for enhanced autonomous driving.
In our exploration with ChatGPT, we first asked some general conceptual questions regarding the LLMs in autonomous vehicles and aimed to find the true potential of LLMs in advancing the future of autonomous driving.

Figure 3. Experiment illustrating LLM-assisted decision-making and motion planning in a complex driving scenario. The Ego vehicle and its trajectory are marked orange; The vehicle ahead in the current lane and its trajectory are blue; The vehicles in adjacent lanes and their trajectories are green.
The responses indicated a profound ability of the LLMs to bridge the interaction between the vehicle and its passengers. From the responses, it’s evident that LLMs can explain complex driving scenarios, decisions made by the vehicle, and even the technical of various autonomous modules. An especially significant observation was the LLMs’ strength in processing vast volumes of data, and then converting these into real-time, understandable feedback. Such feedback isn’t just about driving status but relates to the core autonomous functionalities, including the perception module’s utilization and the motion planning’s choices. Furthermore, the model demonstrated an enhanced capacity for vehicle-to-vehicle communications and, critically, troubleshooting. This capability not only fosters trust but can also develop the user experience by explaining the complex decisions of autonomous operations.

As we can see in Figure 3, we simulated a real-world driving scenario where the autonomous vehicle is equipped with Large Language Models (LLMs) to assist in decision-making and motion planning. The vehicle was on a two-lane Indiana highway, traveling east to west at 96 km/h. It was behind another vehicle moving at the same speed but only 8 meters away, a distance less than optimal for safety. On the adjacent left lane, two other vehicles were noted: one 30 meters ahead moving at 112 km/h, and another 40 meters behind at 104 km/h. The driver was highly attentive, and one passenger was wearing a seatbelt.

The LLMs were tasked with processing this multilayered data sourced from the perception module (vehicle speeds and distances), the localization module (road and environmental conditions), and the in-cabin monitoring system (driver’s attention level and safety measures like seatbelts). The LLMs formulated a comprehensive 9-step motion plan that prioritized safety while efficiently executing the driver’s command to overtake the front vehicle.

In the experimental scenario, the Large Language Models (LLMs) showcased their advanced reasoning ability by not just collecting and analyzing data but also applying layers of context-sensitive reasoning. The LLMs evaluated the speeds and distances of surrounding vehicles, the driver’s state of attention, and even the traffic conditions to determine the safest and most efficient trajectory for overtaking. This capability to reason in real-time, considering multiple factors dynamically, significantly contributes to road safety and operational efficacy. The LLMs didn’t merely follow pre-defined rules but adapted their decision-making to the unique circumstances, highlighting their potential for enhancing the future of autonomous driving.

Additionally, the language interaction capabilities of the LLMs proved crucial for trust-building. When the driver commanded to “overtake the vehicle in front,” the LLMs assessed various factors and communicated their reasoning to the driver. This transparent interaction not only enhanced safety but also instilled greater confidence in the vehicle’s autonomous capabilities.

LLMs also can access previous data and user preferences from the memory module, which allows for a more personalized driving experience. In the context of the experiment, for instance, the system could recall the driver’s typical comfort levels with overtaking speeds, following distances, and lane preferences. This information could then influence how the LLMs interpret and execute a command like “overtake the front vehicle,” ensuring that the action aligns with the driver’s past behavior and comfort zones. As a result, the LLMs’ capacity for memory-driven personalization not only improves user satisfaction but also can contribute to safer, more predictable autonomous driving scenarios.

Another crucial advantage is enhanced transparency and trust. When the vehicle makes a complex decision, such as overtaking another vehicle on a high-speed, two-lane highway, passengers and drivers might naturally have questions or concerns. In these instances, the LLMs don’t just execute the task but also articulate the reasoning behind each step of the decision-making process. By providing real-time, detailed explanations in understandable language, the LLMs demystify the vehicle’s actions and underlying logic. This not only satisfies the innate human curiosity about how autonomous systems work but also builds a higher level of trust between the vehicle and its occupants.

Moreover, the advantage of “zero-shoting” was particularly evident during the complex overtaking maneuver on a high-speed Indiana highway. Despite the LLMs not having encountered this specific set of circumstances before—varying speeds, distances, and even driver alertness—it was able to use its generalized training to safely and efficiently generate a trajectory for the overtaking action. This ensures that even in dynamic or rare scenarios, the system can make sound judgments while keeping users informed, hence building trust in autonomous technology.

5. Conclusion

In conclusion, our paper has provided a comprehensive framework for integrating Large Language Models (LLMs) into the ecosystem of autonomous vehicles. We highlighted how LLMs offer advanced reasoning capabilities that can make autonomous systems more flexible and responsive to complex, real-world scenarios. Additionally, by leveraging the capabilities of LLMs, we can enrich the human-vehicle interaction, providing a more reliable, intuitive, and responsive interface. Unlike traditional autonomous systems, which lack the capacity for language understanding, LLMs can handle complex requests, offer real-time feedback and comprehensive explanations, and assist in decision-making during complex or rare driving scenarios. This suggests a future where LLMs can significantly enhance efficiency, safety, and user-centric design in autonomous vehicles.
References


