Hybrid-Driving: An Autonomous Driving Decision Framework Integrating Large Language Models, Knowledge Graphs and Driving Rules

Jiabao Wang[1,2], **Zepeng Wu**[1,2], **Qian Dong**[1], **Lingzhong Meng**[1], **Yunzhi Xue**[1], **Yukuan Yang**[1,†]

[1]Institute of Software, Chinese Academy of Sciences, Beijing 100190, China.
[2]University of Chinese Academy of Sciences, Beijing 101408, China.
{wangjiabao2023, wuzepeng2023, dongqian, lingzhong, yunzhi, yangyukuan}@iscas.ac.cn

Abstract

Recent advancements have underscored the exceptional analytical and situational understanding capabilities of Large Language Models (LLMs) in autonomous driving decisions. However, the inherent hallucination issues of LLMs pose significant safety concerns when utilized as standalone decision-making systems. To address these challenges, we propose the Hybrid-Driving framework, which leverages LLMs' situational comprehension and reasoning abilities alongside the specialized driving expertise embedded in knowledge graphs and driving rules, thereby enhancing the safety, robustness, and reliability of autonomous driving decisions. To articulate driving experiences clearly, we introduce the Scenario Evolution Knowledge Graph (SEKG), which integrates scenario prediction and action risk analysis in autonomous driving. By delineating observation areas and defining Time-to-Collision (TTC) levels, we effectively control the number of driving scenario nodes and ensure scenario diversity. Based on the scenario evolution relationships within the SEKG, we predict scenarios and assess associated action risks. Additionally, we implement a rule-filtering mechanism to eliminate unreasonable actions and employ prompt engineering to integrate scenario information, optional actions, and SEKG-based action risk analysis into the LLMs for decision-making. Extensive experiments demonstrate that our approach substantially improves decision success rates compared to using LLMs alone ($\geq$37.5%), as well as surpasses the DiLu framework with LLMs and few-shot driving memory ($\geq$7.5%), and other reinforcement learning methods ($\geq$11%). These results validate the effectiveness of the Hybrid-Driving framework in enhancing LLM reliability for autonomous driving and advocate for its broader application of domain-specific knowledge across other fields.

Introduction

Recently, the rapid development of large language models (LLMs), exemplified by ChatGPT (Gabashvili 2023), has heralded a new era in the field of artificial general intelligence (AGI) (Zhang et al. 2024). These advancements have become the driving force behind a new wave of technological revolution. Owing to their powerful understanding, reasoning, and decision-making capabilities, the integration of LLMs with autonomous driving systems offers substantial advantages (Brown et al. 2020). Notably, the innovative notion of employing LLMs in autonomous driving decision-making processes has captured considerable interest (Gyevnar 2024).

Although the research landscape on LLMs in autonomous driving is rapidly expanding, these applications still confront substantial challenges, particularly regarding safety issues (Bogdoll et al. 2021). Fundamentally, LLMs are inherently probabilistic in nature. The quality of data and training methodologies imparts LLMs with traits such as opacity, poor interpretability, and ambiguous operational boundaries (Yang et al. 2024). These characteristics often culminate in hallucinations (Leiser et al. 2024), generating content that may not align with real-world facts (Lin et al. 2022). In the context of autonomous driving, where safety is paramount, the deployment of LLMs encounters a significant trust crisis due to the uncertainty and hallucination issues. Additionally, autonomous driving systems demand both versatility and specialized expertise (Wang et al. 2024; Chen et al. 2024). Specifically, such systems must adeptly navigate diverse traffic scenarios with general ability while embodying professional driving knowledge and experience without committing overt errors. Therefore, integrating LLMs with hallucination problems into the decision-making modules of autonomous driving systems, without the guidance of professional driving expertise, poses significant safety risks (Azamfirei et al. 2023). Thus, incorporating professional knowledge into autonomous driving decision-making processes to mitigate the hallucination issue of LLMs has emerged as an urgent research challenge that necessitates immediate attention (Lin et al. 2024).

Integrating driving expertise to enhance domain-specific capabilities for LLMs represents a cutting-edge approach in the field (Cui et al. 2024). However, there remains a paucity of works that comprehensively and accurately define, store and utilize driving expertise within LLM decision-making processes. Driving expertise is herein defined as the knowledge encompassing the appropriate action a vehicle should undertake in a given scenario, the associated risks in subsequent scenarios, and the driving regulations that must be adhered to prevent collisions. This expertise can be effectively represented by Knowledge Graphs (KGs), which systematically organize entities such as road conditions, vehicles and

† Corresponding author.

their corresponding actions, and scenario transitions (Calbimonte et al. 2023). The structured nature of KGs facilitates the storage of extensive driving expertise, including the prescribed actions and their associated risks, thereby augmenting the reasoning capabilities of LLMs. Additionally, driving rules can be distilled into a set of guidelines that vehicles must follow in typical scenarios to avoid improper actions. By leveraging KGs and driving rules, driving expertise can be accurately represented and effectively utilized in enhancing LLM performance in domain-specific decisions.

Based on these motivations, we propose a novel framework, Hybrid-Driving, which, to our knowledge, is the first to integrate KGs and driving rules with LLMs for autonomous driving decisions. By leveraging the complementary strengths of the general reasoning capabilities of LLMs and the domain-specific expertise embedded in KGs and rules, Hybrid-Driving aims to mitigate the hallucination problem of LLMs and enhance the robustness and reliability of decision-making processes in autonomous driving systems. The contributions of our work are summarized as follows:

- We present Hybrid-Driving, an autonomous driving decision framework that integrates LLMs with domain-specific driving expertise to mitigate hallucinations and enhance decision-making performance. Driving expertise is defined as the selection of appropriate vehicle actions tailored to specific scenarios, the evaluation of associated risks, and adherence to established driving regulations. To this end, a Scenario Evolution Knowledge Graph (SEKG) and a comprehensive set of driving rules are developed and integrated into the framework to guide the LLMs, ensuring more accurate and reliable decisions.

- To address the challenges posed by the exponential increase in scenario nodes and the resultant complexity in scenario matching, we propose a method that integrates scenario observation area division with a Time-to-Collision (TTC)-based approach for scenario node construction. To balance the number of scenario nodes with the diversity of scenario representations, the observation area is divided into 4 distinct subareas, with the TTC within each subarea categorized into 6 levels. This approach facilitates the efficient construction and utilization of SEKG, thereby enhancing scenario prediction and action risk analysis.

- The scenario information, action risks derived from SEKG, and optional actions filtered by driving rules are integrated and fed into LLMs for final action determination. Extensive experiments reveal that our approach significantly enhances decision success rates compared to LLMs used in isolation ($\geq$ 37.5%) and exceeds the performance of current state-of-the-art methods ($\geq$7.5%). These findings affirm the efficacy of the Hybrid-Driving framework in improving LLM reliability for autonomous driving and support its potential application of domain-specific knowledge across various fields.

Related Work

Research on LLMs for autonomous driving decisions falls into two main categories: those employing LLMs without external knowledge and those incorporating external knowledge. Approaches without external knowledge include LMDrive, which integrates LLMs with multi-modal sensor data for interpreting natural language instructions (Shao et al. 2024); DriveGPT4, which utilizes video and textual inputs for enhanced decision-making (Xu et al. 2023); DriveVLM-Dual, combining visual-language models with autonomous driving techniques (Tian et al. 2024); GPT-Driver enhances motion planning by using LLMs for precise trajectory planning and detailed decision-making explanations (Mao et al. 2023); and BEVGPT, which uses BEV images for accurate predictions (Wang et al. 2023). Recent advancements have also incorporated external knowledge to improve the security and reliability of autonomous driving decisions. Significant contributions in this domain include DiLu, which enhances causal reasoning with a driving memory module (Wen et al. 2024); RAG-Driver integrates retrieval-augmented learning for generating driving explanations (Yuan et al. 2024); and DriVLMe leverages embodied and social experiences for complex scenarios (Huang et al. 2024). Despite these advancements, existing methods only rely on static knowledge and a fixed repository of action decisions for specific scenarios, without incorporating dynamic predictions, risk assessments and driving rules for future scenarios.

Methodology

Overview

This research seeks to leverage the general reasoning capabilities of LLMs alongside domain-specific knowledge embedded in KGs and driving rules to enhance the decision performance in autonomous driving systems. Accordingly, we propose a novel framework, Hybrid-Driving, which integrates these complementary strengths to overcome the limitations of single-focus approaches and improve the safety and reliability of autonomous driving systems. The Hybrid-Driving framework comprises three primary modules, as illustrated in Figure 1.

Simulation and TTC-based Scenario Node Construction Module

The simulation and TTC-based scenario node construction module manages the simulation environment and generates scenario nodes. This process involves three key steps: observation area division, scenario parameter extraction, and TTC-based scenario node construction. These steps are crucial for ensuring accurate scenario nodes, which are vital for effective risk prediction and decision-making.

Observation Area Division: In practical driving scenarios, ego vehicles often need to focus only on background vehicles within a specific range that may pose a safety threat. Based on this finding, we define an observation rectangle area and concentrate solely on the vehicles within this region. As illustrated in Figure 1, we divide the observation

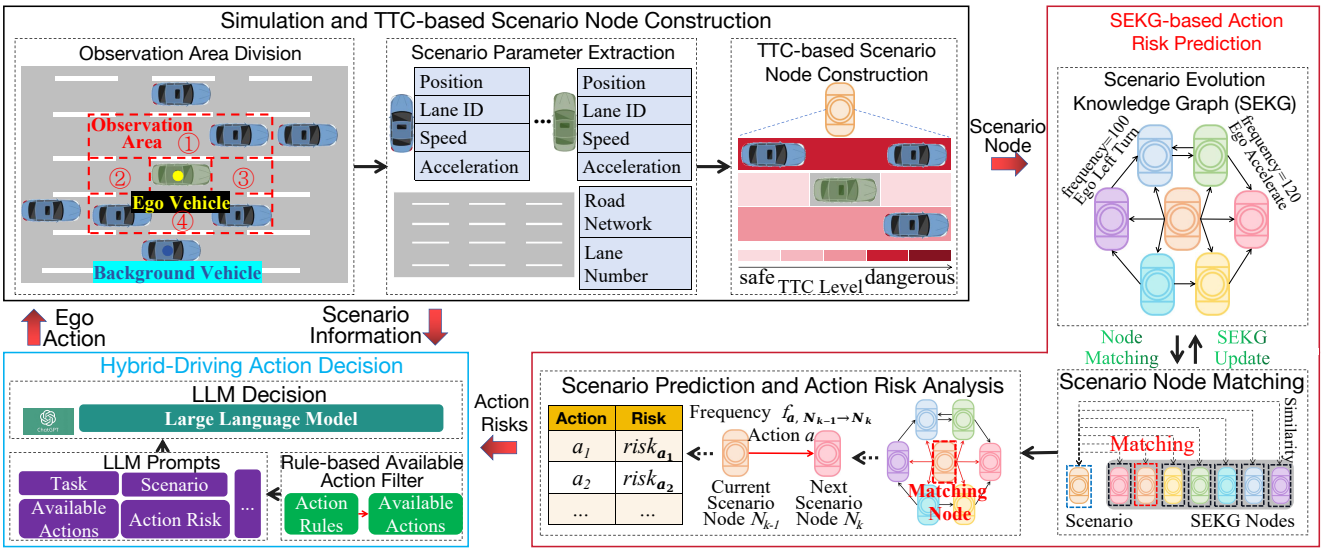


Figure 1: **The overall framework of Hybrid-Driving.** Hybrid-Driving comprises three primary modules: simulation and TTC-based scenario node construction, SEKG-based action risk prediction, and Hybrid-driving action decision. First, the simulation and TTC-based scenario node construction module collects the current scenario parameters of vehicles within the ego vehicle's observation area. It constructs a scenario node by calculating the TTC levels for the pre-segmented peripheral areas around the ego vehicle. Next, the SEKG-based action risk prediction module matches the constructed scenario node with nodes in a pre-established SEKG, developed by simulating typical driving models in a controlled environment and recording corresponding actions and scenario data. If a match is found, the system predicts the next scenario based on the evolution of the matched scenario and performs risk calculations for various actions. Additionally, the corresponding action frequency in the SEKG is updated. If no match is found, the current scenario node is added to the SEKG. Finally, the Hybrid-Driving action decision module evaluates the possible actions of the ego vehicle in the current scenario based on manually formulated empirical rules, eliminating unreasonable choices. It then inputs the current scenario, task, optional actions, and corresponding risk information into LLMs for final action decision. The chosen action is output to the simulation environment for execution.

area into 4 distinct subareas in this study: the left adjacent lane subarea (① in Figure 1), the rear subarea (② in Figure 1), the front subarea (③ in Figure 1), and the right adjacent lane subarea (④ in Figure 1). Limiting and subdividing this observation area offers two significant benefits: (1) Reduction of background information: By excluding vehicles that are far away and have less impact on the ego vehicle, the amount of background vehicles required for attention is reduced. (2) Avoiding SEKG node explosion: By dividing the area and selecting the TTC value of the vehicle posing the highest threat in each subarea as a feature element, the number of scenario nodes in the SEKG can be significantly reduced, avoiding an explosion in the number of scenario nodes.

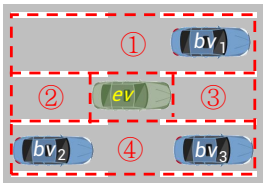
Scenario Parameter Extraction: Following the division of the observation area, we extract critical scenario parameters within the defined area. These parameters encompass road information, the positions, lane IDs, speeds, and accelerations of both background and ego vehicles. Subsequently, these parameters are employed to compute the TTC values and assess the threat levels of the various subareas, which are fundamental for the construction of scenario nodes.

TTC-based Scenario Node Construction: After the observation area division and scenario parameter extraction, the TTC value and its corresponding level are computed for each subarea, thereby facilitating the construction of the corresponding scenario node.

Firstly, the TTC value between the front vehicle and the rear vehicle is governed by

$$TTC_{f,r} = \begin{cases} \infty & \text{if } v_f \geq v_r \\ \dfrac{d}{|v_r - v_f|} & \text{otherwise} \end{cases} \quad (1)$$

where the subscripts f and r represent the front and rear vehicles, respectively. v_f and v_r are their speeds, and d is the distance between them.



Subarea	Subarea TTC
① 1 bv	$S_1 = TTC_{bv_1,ev}$
②③ 0 bv	$S_2 = S_3 = -1$
④ ≥1 bvs	$S_4 = min(TTC_{bv_i,ev})$

Figure 2: The subarea TTC value computation.

Consider the scenario depicted in Figure 2 as an example. For subareas that contain only one background vehicle,

such as subarea ①, the subarea TTC value is equal to the TTC value between the ego vehicle (ev) and the background vehicle (bv), which is

$$S_1 = TTC_{bv,ev} \tag{2}$$

where bv and ev represent the background vehicle and the ego vehicle, respectively, while $TTC_{bv,ev}$ denotes the TTC value, as defined in Equation 2, between these two vehicles.

Subareas ② and ③ are distinct due to the absence of background vehicles. To differentiate them from other subareas containing vehicles, the TTC values are set to -1, indicating no vehicular interactions. Therefore, the TTC values for these subareas are defined as

$$S_2 = S_3 = -1. \tag{3}$$

For subareas containing more than one background vehicle, such as subarea ④, the subarea TTC value is set as the minimum TTC value between the ego vehicle and the background vehicles, which is

$$S_4 = \min(TTC_{bv_i,ev}) \tag{4}$$

where i is the index of the background vehicles in this subarea.

At this stage, the TTC values for four distinct subareas have been determined. Representing the scenario directly with these subarea TTC values would lead to a combinatorial explosion in the number of scenario nodes. To mitigate this complexity, the TTC values are discretized into categorical levels. This discretization significantly reduces the complexity of scenario representation. The subarea TTC level L_j is computed by

$$L_j = \begin{cases} -1 & \text{if } S_j = -1 \\ 0 & \text{if } S_j \geq T_{threshold} \\ M - \left\lfloor \dfrac{S_j}{T_{threshold}} \cdot M \right\rfloor & \text{otherwise} \end{cases} \tag{5}$$

where $j \in \{1, 2, 3, 4\}$ represents the index of the subareas. When no background vehicle is present in the subarea, the subarea TTC level is -1. The parameter $T_{threshold}$ is set to a relatively high TTC threshold of 4 seconds. Beyond this threshold, it is assumed that the background vehicle presents minimal threat to the ego vehicle. M is the maximum TTC level, which is set to 4 in this study. Consequently, there are 6 subarea TTC levels: $-1, 0, 1, 2, 3, 4$. The higher the subarea TTC level L_j is, the greater the threat this subarea poses to the ego vehicle.

In addition to focusing on the TTC levels of different subareas, our attention to the scenario also encompasses a comprehensive risk assessment. Recognizing that various subareas pose differing degrees of threat to the ego vehicle, we define the overall risk value of the scenario node N_k as the weighted average of subarea TTC levels, which is

$$risk_{N_k} = \sum_{j=1}^{4} w_j \cdot L_j \tag{6}$$

where k is the index of scenario nodes. The weights associated with the 4 subareas are set to $[0.2, 0.3, 0.3, 0.2]$ empirically.

Finally, we combine the subarea TTC levels and the scenario risk to represent the scenario node. The scenario node is constructed by

$$N_k = [L_1, L_2, L_3, L_4, risk_{N_k}] \tag{7}$$

Here, considering the number of scenario nodes and the diversity of scenarios, we partition the observation area into 4 distinct subareas. Each subarea's TTC is categorized into 6 levels. While alternative approaches for partitioning scenario subareas and classifying TTC levels exist, the proposed scenario node construction methodology remains applicable to these alternative approaches.

SEKG-based Action Risk Prediction Module

The SEKG-based action risk prediction module integrates comprehensive driving knowledge into the Hybrid-Driving framework via scenario prediction and action risk analysis. The construction of the SEKG, illustrated in Figure 3, precedes these processes. A random decision model facilitates decision-making within the ego vehicle to generate more hazardous scenarios. Following the execution of an action, such as the turn-left action by the ego vehicle, the scenario transitions to the next state. The current scenario node, N_{k-1}, and the subsequent scenario node, N_k, are determined according to the methodology described in the simulation and TTC-based scenario node construction module. Directed edges between nodes N_{k-1} and N_k denote both the type of ego action and the transition frequency $f_{\text{Turn-left},N_{k-1}\to N_k}$ from N_{k-1} to N_k under the turn-left action. To complete the SEKG construction, extensive autonomous driving experiments are conducted within a simulation environment, meticulously recording the ego actions and transition frequencies.

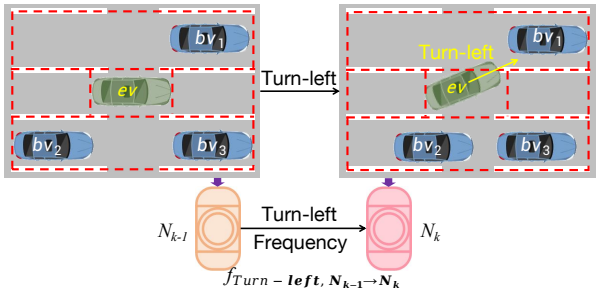


Figure 3: Illustration of the SEKG construction.

Scenario Node Matching: Scenario node matching identifies the most similar node, facilitating SEKG-based scenario prediction and action risk analysis.

To quantify similarity between scenario nodes, we extract the subarea TTC levels and represent them as feature vectors. The similarity between nodes is then assessed using Jensen-Shannon (JS) divergence. To address the potential occurrence of a -1 value within the feature vector, we transform and normalize the feature vector as following

$$F_k = Norm(2 \times N_k[0:3] + 3 \times \mathbf{1}_4) \tag{8}$$

where N_k is the scenario node as defined in Equation 7, $\mathbf{1}_4$ is a 4-dimensional vector consisting entirely of ones, $[1, 1, 1, 1]$, and $Norm(\cdot)$ is the normalization function. Through this transformation, the TTC levels are mapped from $[-1, 0, 1, 2, 3, 4]$ to $[1, 3, 5, 7, 9, 11]$ and then normalized to serve as the feature vector.

For current scenario node N_x and the scenario node N_y in SEKG, their similarity is computed as

$$G(N_x, N_y) = 1 - J(F_x, F_y) \tag{9}$$

where F_x and F_y are the feature vectors of scenario nodes N_x and N_y, respectively. $J(\cdot)$ is the function for calculating the JS divergence.

We select the node N_y with the maximum similarity to N_x in SEKG as the match for N_x. Consequently, scenario prediction and action risk analysis can be conducted based on the SEKG. If N_x is completely equivalent to N_y, then the edge corresponding to N_y is updated accordingly. Otherwise, N_x is incorporated as a new node within the SEKG.

Scenario Prediction and Action Risk Analysis: Upon identifying the scenario node that matches, we proceed with scenario prediction and action risk analysis.

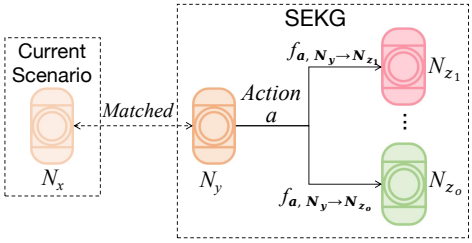


Figure 4: The ego action risk computation.

As illustrated in Figure 4, the current scenario node N_x is matched with node N_y in the SEKG. When the ego vehicle executes an action a, the scenario node may transition from N_y to N_{z_j} with a frequency $f_{a,N_y \to N_{z_j}}$. The associated risk of action a is quantified as

$$risk_a = \frac{\sum_{j=1}^{o} f_{a,N_y \to N_{z_j}} \times risk_{N_{z_j}}}{\sum_{j=1}^{o} f_{a,N_y \to N_{z_j}}} \tag{10}$$

where $f_{a,N_y \to N_{z_j}}$ represents the frequency with which N_y transitions to N_{z_j} under action a, $risk_{N_{z_j}}$ denotes the risk associated with scenario N_{z_j} as defined in Equation 6, and o is the number of scenarios which N_y can transition to under action a.

Hybrid-Driving Action Decision Module

Finally, the scenario information and action risks are synthesized to inform the ultimate driving action decisions, which are executed by the Hybrid-Driving action decision module. Initially, a pre-defined set of driving rules is employed to exclude actions that are evidently unreasonable. Following this, the scenario information, optional actions, corresponding risks, and other relevant factors, are synthesized into comprehensive prompts. These prompts are subsequently processed by LLMs to derive the final action decisions, which are then implemented in simulation.

Rule-based Available Action Filter: The rule-based action filter excludes infeasible and prohibited actions, mitigating the risk of LLMs making erroneous decisions. This filter, detailed in Appendix A.1, applies a comprehensive set of pre-defined driving rules based on scenario-specific vehicle information, significantly reducing the likelihood of incorrect action selection.

LLM Prompts: The LLM prompts consist of three main components: the system prompt, the example prompt, and the scenario information and action risk prompt. The system prompt includes the LLM role and capability settings, task descriptions, chain-of-thoughts, output format specifications, etc. The example prompt includes several historical scenario descriptions alongside corresponding decisions manually refined by human experts. These examples are served for exemplar learning. The scenario information and action risk prompts include current scenario information including road and vehicles, optional ego vehicle actions, and corresponding risks. The specific prompt template is provided in Appendix A.2.

LLM Decision: Based on the provided prompts, the LLM generates decision texts that guide the subsequent actions of the system. Given that the LLM produces outputs in natural language, it is imperative to extract actionable content from these texts. However, the inherent risk of format inconsistencies necessitates the implementation of a correction mechanism. If the output deviates from the expected format, both the erroneous output and the original prompt are reintroduced to the LLM for rectification. Upon obtaining a correctly formatted decision text, the extracted actions are implemented for the ego vehicle.

Experiment Results

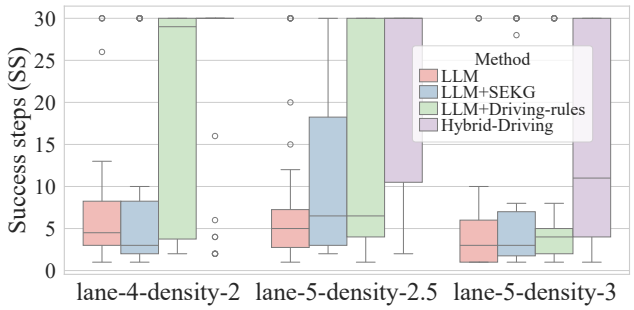
The effectiveness of the proposed Hybrid-Driving framework is evaluated through three primary approaches. First, ablation studies assess the impact of incorporating driving expertise, represented by SEKG and driving rules. This is achieved by comparing the performance of LLM decisions with and without the integration of driving expertise. Second, Hybrid-Driving is tested by applying the SEKG constructed for a 4-lane scenario to 5-lane scenarios, and the results are analyzed to assess the generalization capability. Finally, the Hybrid-Driving framework is benchmarked against state-of-the-art autonomous driving decision-making methods, including recent advancements in LLM and RL, to demonstrate its superiority.

To ensure consistency and comparability, the same experimental setup and evaluation metrics as those employed in DiLu (Wen et al. 2024) are adopted. The experiments are conducted across three distinct scenario settings: lane-4-density-2, lane-5-density-2.5, and lane-5-density-3. In each scenario, SEKG is constructed using a random decision model within the Highway-Env. Approximately 100,000

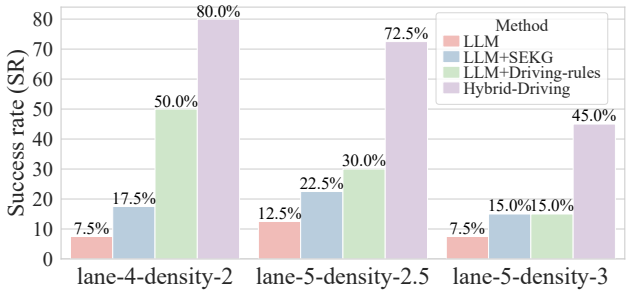
frames of data are collected per scenario setting for the construction of the SEKG. ChatGPT-3.5 is utilized for LLM-based decision-making in Hybrid-Driving. During the inference stage, 40 rounds of decision-making are conducted across the three scenarios, with each round consisting of 30 decisions. The evaluation metrics include Success Steps (SS) and Success Rate (SR). The SS metric quantifies the number of consecutive frames without a collision, with an SS of 30 indicating the successful completion of the driving task. The SR metric represents the proportion of rounds without collisions relative to the total number of experimental rounds.

Validation of Hybrid-Driving Integrating LLM, SEKG, and Driving Rules

To systematically assess the efficacy of the Hybrid-Driving framework's expertise-assisted LLM decisions, augmented by knowledge graphs and driving rules, four distinct configurations are compared: (1) LLM, (2) LLM+SEKG, (3) LLM+Driving-rules, and (4) LLM+SEKG+Driving rules (Hybrid-Driving). The Success Steps (SS) and Success Rate (SR) outcomes for these configurations are depicted in Figure 5(a) and Figure 5(b).



(a)



(b)

Figure 5: Comparison of LLM, LLM+SEKG, LLM+Driving-rules, and Hybrid-Driving under different scenario settings. (a) Success Steps; (b) Success Rate.

The integration of driving expertise, represented through knowledge graphs and driving rules, into LLM autonomous driving decision-making markedly enhances performance. Figure 5(a) presents the box plots for LLM, LLM+SEKG, LLM+Driving-rules, and Hybrid-Driving across three scenario settings. The figure reveals that the median SS for Hybrid-Driving in these scenarios reaches 30, significantly outperforming other methods. Hybrid-Driving's superiority is more pronounced in SR, as illustrated in Figure 5(b). Under consistent experimental conditions, Hybrid-Driving demonstrates remarkable improvements of 72.5%, 60%, and 37.5% across the three scenarios compared to utilizing LLM alone. Approaches incorporating partial driving experience also show considerable enhancements over LLM alone. For instance, LLM+SEKG increases by 10%, 10%, and 7.5%, while the LLM+Driving-rules method improves by 42.5%, 17.5%, and 7.5%, respectively. Experimental results indicate that LLM+Driving-rules outperforms LLM+SEKG, albeit marginally. This is attributed to the comprehensive nature of the developed driving rules, which effectively eliminates actions to prevent common collisions. In contrast, SEKG suffers from inadequate data, limiting its capabilities to provide only basic action risk prediction values. However, Hybrid-Driving harnesses the strengths of both approaches, utilizing SEKG for risk prediction and action selection in scenarios beyond the scope of the driving rules. This combination significantly enhances overall performance.

Methods	lane-4-density-2	lane-5-density-2.5	lane-5-density-3
LLM	1.81	1.67	2.32
LLM+Driving-rules	0.78	1.16	1.79
LLM+SEKG	1.22	1.11	1.42
Hybrid-Driving	**0.44**	**0.72**	**0.92**

Table 1: Mean scenario risks across different methods scenario settings. A higher value indicates a greater level of danger associated with the scenario.

To further validate our framework, we compute the mean risk, as defined in Equation 6, for each frame across all 40 rounds of the decision-making experiment under LLM, LLM+SEKG, LLM+Driving-rules, and Hybrid-Driving. The results, presented in Figure 1, indicate that Hybrid-Driving exhibits the lowest mean scenario risk value during the decision-making process. This finding corroborates the superiority of Hybrid-Driving in ensuring safer autonomous driving decisions. It can be concluded that, compared to LLM alone, Hybrid-Driving effectively guides vehicles towards lower-risk driving scenarios by integrating decision assistance from SEKG and driving rules, thereby mitigating the risk associated with driving scenarios.

In summary, the comparisons of the SS, SR, and mean scenario risk between LLM, LLM+SEKG, and LLM+Driving-rules with Hybrid-Driving further validate the effectiveness of the Hybrid-Driving framework in enhancing the safety and reliability of autonomous driving decisions. These significant improvements underscore the critical importance of integrating comprehensive driving expertise into LLM autonomous driving decisions to achieve superior performance.

Generalization of Hybrid-Driving Framework

The construction of knowledge graphs often requires extensive data tailored to specific scenarios. However, in practical applications, performance deteriorates due to changes in

these scenarios. To address this, we introduce the observation area division and TTC-based scenario node construction mechanisms in the SEKG construction process, which are less affected by variations in lane and traffic density. This approach enables the constructed SEKG to exhibit inherent similarities to different scenario settings and provides generalization capability. To validate this hypothesis, we construct the SEKG using only simulation data from the lane-4-density-2 scenario and conduct generalization tests in the lane-5-density-2.5 and lane-5-density-3 scenarios.

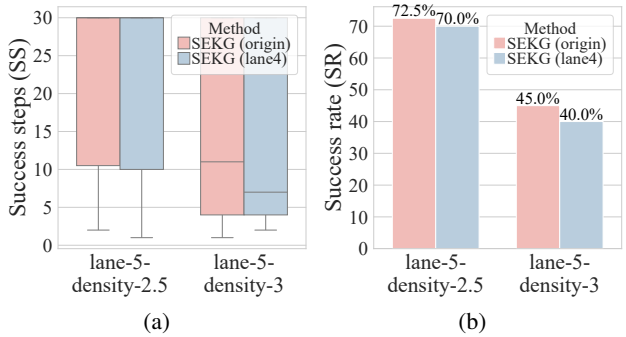


Figure 6: Generalization results of Hybrid-Driving under various scenario settings. (a) Success Steps; (b) Success Rate.

As illustrated in Figure 6, Hybrid-Driving utilizing the lane-4-density-2 SEKG (lane4) exhibits minimal performance degradation in the lane-5-density-2.5 and lane-5-density-3 scenarios compared to Hybrid-Driving using the customized scenario SEKG (origin). Both SEKG (lane4) and SEKG (origin) show similar SS distributions in the lane-5-density-2.5 and lane-5-density-3 scenarios. Additionally, SEKG (lane4) demonstrates slight performance reductions of 2.5% and 5% relative to SEKG (origin), respectively. These results suggest that Hybrid-Driving possesses strong generalization capability, effectively accommodating variations in scenarios.

Comparison with Representative LLM and RL Methods

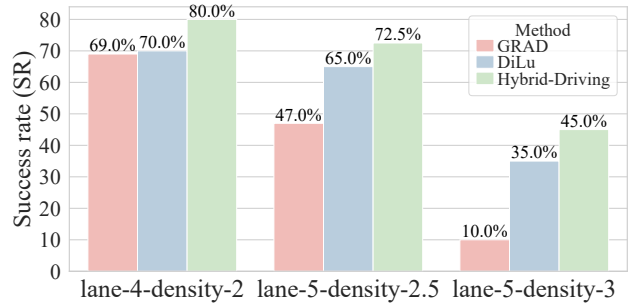


Figure 7: Comparison of Hybrid-Driving, DiLu and GRAD under different scenario settings.

We compare the Hybrid-Driving framework with two latest state-of-the-art approaches: the reinforcement learning method GRAD (Xi and Sukthankar 2022) and the LLM-based autonomous decision-making framework DiLu (Wen et al. 2024). GRAD, specifically designed for autonomous driving, excels in forecasting the future trajectories of surrounding vehicles. DiLu, a significant advancement in LLM-based decision-making, incorporates a memory module with the LLM, leading to enhanced performance.

Due to the lack of publicly available SS data for GRAD and DiLu, we assess the performance of GRAD, DiLu, and Hybrid-Driving in terms of SR. As depicted in Figure 7, the Hybrid-Driving framework achieves SR of 80%, 72.5%, and 50.0% across three distinct scenarios, respectively. These results represent improvements of 11%, 25.5%, and 35% over GRAD in the corresponding scenarios, and 10%, 7.5%, and 10% over DiLu. Compared to GRAD, LLM-driven methods, notably DiLu and Hybrid-Driving, show significant advantages in complex scenarios like lane-5-density-2.5 and lane-5-density-3, highlighting the potential of LLMs for tackling autonomous driving challenges. Additionally, Hybrid-Driving exhibits superior scenario prediction and action risk analysis capabilities compared to DiLu with few-shot memory, thereby enhancing its efficacy in autonomous driving decisions.

Conclusion

The application of LLMs in autonomous driving represents a significant research frontier, leveraging their extensive reasoning capabilities. However, current solutions inadequately address safety and reliability issues due to limitations in handling hallucinations. This study introduces a novel framework, Hybrid-Driving, which integrates LLMs, knowledge graphs, and driving rules to enhance decision-making in autonomous driving systems. We propose SEKG, a method for efficient scenario representation through observation area division and TTC level grading, which facilitates effective scenario prediction and risk assessment of driving actions. Driving rules, derived from human expertise, filter out unreasonable action options. SEKG and driving rules enhance the decision-making capabilities of LLMs, with experimental results demonstrating that this framework substantially improves decision-making success steps and success rates compared to LLMs used in isolation, offering significant advancements over existing LLM and reinforcement learning methods. This research presents a pioneering integration of driving expertise with LLMs, establishing a robust framework that markedly improves the reliability and effectiveness of autonomous driving systems. Additionally, it provides a valuable reference for integrating LLMs with domain-specific knowledge in other fields. Future research will focus on refining SEKG and expanding the driving rule library to further advance LLM applications in autonomous driving.

Acknowledgments

This work was partially sponsored by National Natural Science Foundation of China (Grant No.62306305), SMP-

Zhipu.AI Large Model Cross-Disciplinary Fund, and Youth Innovation Promotion Association CAS.

References

Azamfirei; Razvan; Kudchadkar; R, S.; Fackler; and James. 2023. An analysis of the challenges associated with hallucinations in large models and strategies to mitigate them, with a focus on healthcare applications. *Critical Care*, 27(1): 75.

Bogdoll, D.; Breitenstein, J.; Heidecker, F.; Bieshaar, M.; Sick, B.; Fingscheidt, T.; and Zöllner, J. M. 2021. Description of corner cases in automated driving: goals and challenges. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 1023–1028.

Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; and Amodei, D. 2020. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781713829546.

Calbimonte, J.-P.; Ciortea, A.; Kampik, T.; Mayer, S.; Payne, T. R.; Tamma, V.; and Zimmermann, A. 2023. Autonomy in the age of knowledge graphs: Vision and challenges. *Transactions on Graph Data & Knowledge*, 1(1): 22p.

Chen, L.; Wu, P.; Chitta, K.; Jaeger, B.; Geiger, A.; and Li, H. 2024. End-to-end autonomous driving: Challenges and frontiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–20.

Cui, C.; Ma, Y.; Cao, X.; Ye, W.; and Wang, Z. 2024. Receive, reason, and react: Drive as you say, with large language models in autonomous Vehicles. *IEEE Intelligent Transportation Systems Magazine*, 16(4): 81–94.

Gabashvili, I. S. 2023. The impact and applications of ChatGPT: A systematic review of literature reviews. arXiv:2305.18086.

Gyevnar, B. 2024. Building trustworthy human-centric autonomous systems via explanations. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '24, 2752–2754. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9798400704864.

Huang, Y.; Sansom, J.; Ma, Z.; Gervits, F.; and Chai, J. 2024. DriVLMe: Enhancing LLM-based autonomous driving agents with embodied and social experiences. arXiv:2406.03008.

Leiser, F.; Eckhardt, S.; Leuthe, V.; Knaeble, M.; Mädche, A.; Schwabe, G.; and Sunyaev, A. 2024. HILL: A hallucination identifier for large language models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24. New York, NY, USA: Association for Computing Machinery. ISBN 9798400703300.

Lin; Stephanie; Hilton; Jacob; Evans; and Owain. 2022. TruthfulQA: Measuring how models mimic human falsehoods. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 3214–3252. Dublin, Ireland: Association for Computational Linguistics.

Lin, Z.; Guan, S.; Zhang, W.; Zhang, H.; Li, Y.; and Zhang, H. 2024. Towards trustworthy LLMs: A review on debiasing and dehallucinating in large language models. *Artificial Intelligence Review*.

Mao, J.; Qian, Y.; Zhao, H.; and Wang, Y. 2023. GPT-Driver: Learning to drive with GPT. *arXiv preprint arXiv:2310.01415*.

Shao, H.; Hu, Y.; Wang, L.; Song, G.; Waslander, S. L.; Liu, Y.; and Li, H. 2024. LMDrive: Closed-loop end-to-end driving with large language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15120–15130.

Tian, X.; Gu, J.; Li, B.; Liu, Y.; Hu, C.; Wang, Y.; Zhan, K.; Jia, P.; Lang, X.; and Zhao, H. 2024. DriveVLM: The convergence of autonomous driving and large vision-language models. *arXiv preprint arXiv:2402.12289*.

Wang, P.; Zhu, M.; Lu, H.; Zhong, H.; Chen, X.; Shen, S.; Wang, X.; and Wang, Y. 2023. BEVGPT: Generative pre-trained large model for autonomous driving prediction, decision-making, and planning. *arXiv preprint arXiv:2310.10357*.

Wang, T.-H.; Maalouf, A.; Xiao, W.; Ban, Y.; Amini, A.; Rosman, G.; Karaman, S.; and Rus, D. 2024. Drive anywhere: Generalizable end-to-end autonomous driving with multi-modal foundation models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 6687–6694.

Wen, L.; Fu, D.; Li, X.; Cai, X.; MA, T.; Cai, P.; Dou, M.; Shi, B.; He, L.; and Qiao, Y. 2024. DiLu: A knowledge-driven approach to autonomous driving with large language models. In *The Twelfth International Conference on Learning Representations*.

Xi, Z.; and Sukthankar, G. 2022. A graph representation for autonomous driving. In *The 36th Conference on Neural Information Processing Systems Workshop*, volume 7, 9.

Xu, Z.; Zhang, Y.; Xie, E.; Zhao, Z.; Guo, Y.; Wong, K. K.; Li, Z.; and Zhao, H. 2023. DriveGPT4: Interpretable end-to-end autonomous driving via large language model. *arXiv preprint arXiv:2310.01412*.

Yang, J.; Jin, H.; Tang, R.; Han, X.; Feng, Q.; Jiang, H.; Zhong, S.; Yin, B.; and Hu, X. 2024. Harnessing the power of LLMs in practice: A survey on ChatGPT and beyond. *ACM Trans. Knowl. Discov. Data*, 18(6).

Yuan, J.; Sun, S.; Omeiza, D.; Zhao, B.; Newman, P.; Kunze, L.; and Gadd, M. 2024. RAG-Driver: Generalisable driving rxplanations with tetrieval-augmented in-context learning in Multi-Modal Large Language Model. arXiv:2402.10828.

Zhang; Bo; Tang; and Jie. 2024. Challenges toward AGI and its impact to the Web. *Proceedings of the ACM on Web Conference 2024*, 4(1): 4.