Proceedings of CONF-SEML 2026 Symposium: Importance of Machine Learning Methods and Analysis in Engineering
DOI: 10.54254/2755-2721/2026.AS31938

From Data-Driven to Knowledge-Driven: Prospect of Large Language Models in Cooperative Autonomous Driving Decision-Making

Bowen Zheng

College of Art and Science, The Ohio State University, Columbus, USA
zheng.2491@buckeyemail.osu.edu

Abstract. Large Language Models (LLMs) and Multimodal LLMs (MLLMs) are explored as a cognitive layer to enhance decision-making in autonomous driving systems, addressing the limitations of traditional modular pipelines in long-tail or ambiguous scenarios. This paper conducts a structured and system-oriented review of recent researches on LLM-integrated autonomous driving. The survey categorizes existing methods based on the functional roles that LLMs play in driving pipeline, analyzes representative architecture for system integration and summarizes the major application tasks and evaluation settings. This paper aims to provide an accessible entry point for newcomers by focusing on how language models enhance the decision-making ability of autonomous systems while highlighting emerging design patterns and open challenges. It also discusses current limitations, such as reliance on simulation and safety concerns, and highlight directions for future research. It can be concluded that integrating LLMs into autonomous driving systems establishes a crucial cognitive layer that enhances scene understanding, hierarchical planning, and multi-agent coordination, effectively addressing the limitations of traditional data-driven pipelines in handling long-tail scenarios while improving system interpretability. The most promising LLM-based autonomous driving solutions adopt a hybrid architecture that combines the flexible reasoning capabilities of LLMs with the reliable safety guarantees of conventional planners, though challenges such as the sim-to-real gap, model hallucinations, and real-time performance constraints still require further research to overcome.

Keywords: Autonomous Driving, Large Language Models (LLMs), Cognitive Layer, Decision-Making, Hybrid Architecture

1. Introduction

Autonomous driving systems have typically relied on a modular perception-prediction-planning pipeline to convert sensor data into driving actions. Although massive datasets and deep learning tools have significantly improved those systems, they often struggle when facing "long-tail" or "edge case" scenarios – rare or complex situations that demand high-level reasoning and contextual abstraction. This limitation has gradually driven a shift in research focus from purely data-driven pipelines toward more cognitively capable and knowledge-aware autonomous driving systems [1].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.AS31938

The emerge of Large Language Models (LLMs) and Multimodal LLMs (MLLMs) provides a competitive solution to these challenges. Unlike conventional neural models, LLMs possess inherent strengths in logical reasoning, world knowledge integration, and instruction following. Recent studies have explored their potential across a wide range of tasks, from high-level decision-making to agent-based system orchestration. For instance, representative systems such as Agent-Driver treats driving as a cognitive process boosted by reasoning and memory [2]; DriveLM uses structured language queries to bridge the gap between perception and planning [3]; DiLu demonstrates the effectiveness of LLMs in making complicated and language-driven driving decisions [4].

Despite rapid progress, existing studies on LLM-integrated driving are largely fragmented. Existing researches vary widely in task formation and system-level integration strategies, making it challenging for newcomers to understand the underlying designed principles. Despite the existing broader surveys on multimodal models [1,5], there is a notable lack of a system-oriented summary focused specifically on how LLMs enhance the decision-making core. This paper fills the gap by providing a structured, entry-level survey that systematically categorizes recent LLM-based driving approaches according to their fundamental roles and architectural contributions to autonomous decision-making.

2. LLM integration in autonomous driving: roles, architectures and applications

2.1. Functional roles of LLMs in the autonomous driving pipeline

Large Language Models (LLMs) have emerged as a transformative cognitive layer in autonomous driving, beyond traditional modular or end-to-end architectures. They function as a "reasoning engine" that introduces high-level abstraction and cross-modal coordination rather than replacing established perception or control stacks. Existing research generally identifies several functional roles through which LLMs transform low-level perception outputs into higher-level decision-making capabilities [1].

First, LLMs function as semantic interpreters for scene understanding. By processing multimodal inputs, including camera feeds and LiDAR-based bird's-eye-view (BEV) features, Multimodal LLMs (MLLMs) can translate raw sensory data into structured linguistic descriptions [3]. This process of "Language Grounding" enables the system to identify spatial relationships and hazard priorities, similar to human situational awareness. Unlike rigid numerical feature vectors, these language-based representations offer a coherent, queryable interface that improves how underneath modules interpret the driving environment.

Second, LLMs serve as the strategic core for decision-making. The models utilize Chain-of-Thought (CoT) reasoning to determine driving intentions such as yielding or lane-changing, enabling smooth handling of "long-tail" scenarios, including navigating chaotic construction sites. In this role, the model does more than predict the next action; it explains why an action is necessary according to traffic regulation and contextual nuances, providing a form of interpretability previously absent from autonomous systems [3].

Third, LLMs also act as a facilitator for planning and control. Outputs from LLMs provide a "scaffold" for traditional learning-based controllers to refine into smooth, collision-free motion. This hybrid approach effectively connects the logical flexibility of LLMs with the rigorous safety guarantees of classical control theory [6].

Finally, LLMs emerge as system orchestrators and human-machine interfaces. The vehicle's internal "logic flow" is managed by scheduling perception queries, retrieving traffic regulations and

DOI: 10.54254/2755-2721/2026.AS31938

interacting with digital twins or maps [2]. LLM plays a significant role as a central coordinator that integrate the vehicle's internal states with both human instructions and environmental demands.

In summary, LLMs contribute substantially to a language-mediated cognitive layer. By enhancing abstraction and reasoning, they are fundamentally reshaping the driving stack into a more integrated and interpretable architecture [1].

2.2. Architectural paradigms and system designs

Beyond their functional roles, integrating LLMs into autonomous driving requires well-structured system architectures. These designs mainly vary based on how they construct language representations, execute reasoning, and interface with existing driving modules. Current methods can be categorized into three types: language-based scene reasoning, hierarchical planning assistance, and memory-augmented system orchestration.

2.2.1. Language-based scene understanding and reasoning

First, a common design uses LLMs to convert raw environmental data into interpretable semantic frameworks. In these systems, multimodal inputs such as LiDAR point clouds and bird's-eye-view (BEV) features are processed by encoders or compilers and mapped into a language-based representation space [3]. This allows the LLM to operate with "textual tokens", facilitating higher-level reasoning about objects interactions and spatial constraints.

A key benefit of this approach is its separates sensory processing from high-level cognition. LLMs can leverage their "world knowledge" to identify latent hazards, such as anticipating occluded agents, without requiring complex numerical models for every detail. This design ensures the reliability of high-level reasoning under reasonable perceptual assumptions, even when sensor data is noisy or incomplete.

2.2.2. Hierarchical planning and control assistance

Second, many systems employ LLMs as high-level guides rather than direct controllers. In this setup, the LLM acts as a "strategic consultant" that defines overall driving objectives or rough trajectories. These outputs are then transmitted to traditional planners, which manages motion generation and constraint satisfaction, ensuring smooth movement and collision avoidance while adhering to physical laws [6].

This hybrid architecture combines the reliability of classical control such as safety and stability, with the flexibility of LLMs. By acting as an advisor, the LLM adjusts driving behavior based on social norms or long-term goals, cooperating with traditional control systems to ensure precise vehicle motion. This coordination reduces the safety risks associated with allowing a "black-box" model to directly take control of the vehicle.

2.2.3. Memory, tool-augmentation, and system integration

Third, more advanced architecture enables LLM-centered systems to utilize relevant knowledge and actively interact with other system components by integrating tool-use capabilities and external memory., which stores historical driving data, traffic regulations, and map information [2]. When encountering new scenarios, the system retrieves similar past experiences to support decision-making, analogous to how humans learn from cumulative experience.

DOI: 10.54254/2755-2721/2026.AS31938

Additionally, by integrating tools such as HD maps and traffic rule engines, the LLM functions as a system-level coordinator. Instead of passively receiving data, the LLM proactively requests specific information from various modules. The shift from a passive predictor to an active coordinator enhances the modularity, interpretability, and updatability of autonomous driving systems [1].

2.3. Application benchmarks and research tasks

The combination of LLMs and autonomous driving has explored the fields of diverse application domains. Research has emphasized more flexible, task-oriented intelligence. Instead of focusing on a single performance metric, current studies examine how language-based reasoning can address various requirements of driving, ranging from detailed perception to complex social negotiation with other vehicles.

2.3.1. Interpretability and explanatory scenario analysis

First, a major application lies in how models "read" a scene through both vision and language. In these tasks, Multimodal LLMs (MLLMs) are required to describe complex driving scenarios and respond to specific questions [3,5]. Unlike simple object detection, the model must explain spatial relationships and the reasons behind potential dangers, such as a double-parked car. By enhancing system interpretability, the researches transform the "black-box" perception of old systems into a transparent, human-interpretable narrative, which helps build user trust

2.3.2. Contextual decision – making and strategic planning

Second, LLMs are increasingly applied for high-level planning, acting as "brains" for complex driving moves. Research in this area focuses on setting driving goals by combining traffic regulations with real-time road conditions [6,7]. Since these systems typically collaborate with traditional planners for final execution, the main goal is to verify whether LLM can handle "tricky" scenarios, such as merging into heavy, fast traffic, where traditional rule-based logic is often overly conservative or inflexible.

2.3.3. Multi-agent coordination and social interaction

Third, LLMs are used to facilitate collaboration among multiple vehicles. This represents a new frontier in "Cooperative Driving", where language enables vehicles to negotiate intentions and share a vision. Tasks include using LLMs to infer other drivers' intentions or establish communication protocols for resolving conflicts at unregulated intersections [8]. Although in early stage, this shows how LLMs can act as a "social bridge" in traffic, allowing for much smoother cooperation than traditional data-sharing methods.

2.3.4. Simulation-based validation and testing

Finally, safety being the top priority, most current research is conducted in high-quality simulation environments. Platforms such as CARLA or MetaDrive provide a safe "sandbox" where language-driven reasoning and human-car interactions can be tested without any real-world risk [5,9]. These simulations are essential for verifying whether LLM-based agents can handle rare "edge cases".

DOI: 10.54254/2755-2721/2026.AS31938

While a gap remains between virtual tests and real-world driving, these environments serve as the primary grounds for evaluating the effectiveness of these new driving architectures.

In summary, the application of LLMs in driving reflects a shift toward more human-like cognitive systems. By linking scene understanding, planning and coordination, LLMs do not merely add new features. Instead, they expand the comprehension and capabilities of autonomous vehicles in real-world scenarios.

3. Conclusion

This survey examines how Large Language Models (LLMs) are transforming autonomous driving, facilitating a paradigm shift in the field from "black-box" data-driven models toward cognitively capable, reasoning-based systems. By decomposing the roles, architectures and tasks of LLMs, this study demonstrates that language-based intelligence adds a significant and necessary cognitive layer to the driving stack. It provides a new way for vehicles to coordinate that goes beyond what traditional pipelines can do.

Our review shows that LLMs function optimally as high-level coordinators rather than direct replacements for steering and braking. The most promising approach is a hybrid architecture: combining the flexible reasoning capabilities of LLMs with the reliable safety guarantees of traditional planners and memory-augmented systems. This setup provides a viable solution to "long-tail" scenarios that have long troubled autonomous driving. Additionally, language-mediated interaction enhances scenario interpretability and enables more human-like inter-vehicle collaboration.

However, significant challenges remain unresolved. Most current researches are confined to simulations, resulting in a substantial "sim-to-real" gap that requires addressing. LLM hallucinations and output unpredictability pose a critical safety concern. Furthermore, existing models often fail to meet the millisecond-level latency requirements of real-time vehicle operation.

Moving forward, research should focus on developing lightweight LLMs deployable directly on vehicle hardware and established more robust benchmarks that reflect the complexity of real-world traffic. It is also essential to integrate LLM reasoning with formal safety verification to build systems that are not only intelligent but also genuinely trustworthy. While the field is in its early stages, knowledge-driven LLM integration represents a promising pathway to developing autonomous vehicles capable of true cognition, continuous learning, and effective inter-vehicle collaboration.

References

[1] Wang, Y., Cui, Y., Chen, Z., & Liu, W. (2024). LLM4Drive: A survey of large language models for autonomous driving. IEEE Transactions on Intelligent Transportation Systems.

[2] Mao, J., Qian, L., Xu, H., & Chen, S. (2024). Agent-Driver: Autonomous driving with agent-based large language models. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI).

[3] Shao, Z., Wang, S., Zhang, C., & Li, X. (2023). DriveLM: Driving with multi-modal large language models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[4] Wen, L., Fu, D., Li, X., Cai, X., Ma, T., Cai, P., Dou, M., Shi, B., & He, L. (2023).DiLu: A knowledge-driven approach to autonomous driving with large language models.arXiv preprint arXiv: 2309.16292.

[5] Cui, Y., Chen, Z., Zhang, J., & Liu, W. (2023). Multimodal large language models for autonomous driving: A survey. arXiv preprint arXiv: 2309.16292.

[6] Chen, D., Zhou, Y., Wang, Y., & Li, H. (2023). Talk2Drive: Language-guided autonomous driving via scene-level reasoning. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).

DOI: 10.54254/2755-2721/2026.AS31938

[7] Wang, J., Liu, Y., Chen, R., & Zhao, D. (2023). LanguageMPC: Large language models as decision makers for model predictive control. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).
[8] Mao, J., Qian, L., Xu, H., & Chen, S. (2023). GPT-Driver: Language-based autonomous driving with large language models. arXiv preprint arXiv: 2310.01412.
[9] Li, Y., Wang, J., Zhang, H., & Zhao, D. (2023). Drive like a human: Learning human-like driving behavior with large language models. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA).