Proceedings of CONF-FMCE 2025 Symposium: Semantic Communication for Media Compression and Transmission
DOI: 10.54254/2755-2721/2025.GL26773

LLM-Based Autonomous Vehicle Self-Protection Mechanisms: A Comprehensive Review

Quanheng Wei

High School Affiliated to Remin University of China, Beijing, China
Devin11232013@outlook.com

Abstract. As autonomous driving technology advances, traditional safety mechanisms face limitations in dynamic threat recognition and real-time decision-making under mechanical stress constraints. This paper synthesizes current research on integrating LLM capabilities with autonomous vehicle safety systems, analyzing how natural language processing and reasoning can enhance threat detection, improve algorithmic decision-making, and optimize the balance between safety imperatives and mechanical structure preservation. Our findings reveal that while LLM integration offers significant potential for improving threat recognition accuracy and decision-making sophistication, critical gaps remain in real-time processing capabilities, mechanical stress modeling, and safety-performance trade-offs. The paper concludes with recommendations for future research directions, emphasizing the need for comprehensive frameworks that seamlessly integrate LLM reasoning with vehicle dynamics and mechanical constraints to achieve optimal safety outcomes in autonomous transportation systems.

Keywords: Suspension System of BNEV, Automated Driving, Parameter Setting of LLM, Safety Control of BNEV

1. Introduction

The rapid evolution of autonomous vehicle technology has fundamentally transformed the landscape of transportation safety, introducing unprecedented opportunities and challenges in vehicular protection mechanisms [1]. Traditional automotive safety systems, while effective in conventional driving scenarios, face significant limitations when confronted with the complex, dynamic environments that autonomous vehicles must navigate independently [2]. The integration of Large Language Models (LLMs) into autonomous vehicle safety frameworks represents a paradigmatic shift toward more intelligent, adaptive, and contextually aware protection mechanisms [3].

Recent developments in artificial intelligence, particularly in natural language processing and reasoning capabilities, have opened new avenues for enhancing autonomous vehicle safety through sophisticated threat detection and decision-making processes [4]. LLMs demonstrate remarkable abilities in pattern recognition, contextual understanding, and logical reasoning that can be leveraged to address critical safety challenges in autonomous driving systems [5]. However, the integration of these advanced AI capabilities with the physical constraints and mechanical realities of vehicle

© 2025 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: 10.54254/2755-2721/2025.GL26773

operation presents complex technical and safety considerations that require comprehensive investigation [6].

Contemporary autonomous vehicle safety systems face several critical challenges that traditional approaches struggle to address effectively. First, existing threat detection mechanisms often rely on predefined scenarios and rule-based systems that may fail to recognize novel or complex threat patterns in real-world driving environments [7]. Second, the balance between aggressive safety maneuvers and mechanical structure preservation remains poorly understood, with limited research on optimizing trade-offs between immediate safety responses and long-term vehicle integrity [8].

The motivation for investigating LLM-based self-protection mechanisms stems from recent incidents involving autonomous vehicle failures, including high-profile accidents that highlight the limitations of current safety systems [9]. Analysis of these incidents reveals that many failures occur not due to sensor malfunction or algorithmic errors, but rather from inadequate threat assessment and suboptimal decision-making under complex, ambiguous conditions [10]. LLMs offer the potential to address these limitations through enhanced reasoning capabilities, contextual understanding, and adaptive learning from diverse scenarios [11].

The primary objective of this review is to provide a comprehensive analysis of current research and development in LLM-based autonomous vehicle self-protection mechanisms, identifying key technological advances, persistent challenges, and promising research directions. Secondary objectives include synthesizing existing literature to establish a theoretical foundation for LLM integration in vehicular safety systems, analyzing the technical feasibility and practical implications of various approaches, and proposing recommendations for future research and development efforts.

The significance of this research extends beyond academic inquiry to encompass critical societal and economic implications. Autonomous vehicles represent a transformative technology with the potential to dramatically reduce traffic accidents, improve transportation efficiency, and enhance mobility for diverse populations [12]. However, the successful deployment of autonomous vehicles depends fundamentally on public confidence in their safety and reliability [13]. LLM-enhanced safety mechanisms offer a pathway toward more robust, intelligent, and trustworthy autonomous vehicle systems that can adapt to complex, unpredictable real-world conditions [14].

This review contributes to the field by providing the first comprehensive synthesis of research on LLM applications in autonomous vehicle safety, establishing a theoretical framework for understanding the integration of natural language processing capabilities with vehicular protection mechanisms, and identifying critical research gaps that require further investigation. The analysis presented here will inform future research directions, guide technology development efforts, and support policy discussions regarding autonomous vehicle safety standards and regulations.

This paper is organized into three main sections that systematically examine different aspects of LLM-based autonomous vehicle self-protection mechanisms. Following this introduction, Section 2 provides a comprehensive literature review that examines existing research on autonomous vehicle safety, LLM applications in transportation, and the intersection of these domains. Section 3 analyzes traditional safety mechanisms and their limitations, while Section 4 explores LLM integration approaches and their potential benefits. Section 5 examines the critical challenge of balancing safety decisions with mechanical stress considerations. The paper concludes with a synthesis of findings, identification of research gaps, and recommendations for future work.

2. Literature review

The foundation of autonomous vehicle safety rests on a complex ecosystem of sensors, algorithms, and mechanical systems designed to detect, assess, and respond to potential threats in real-time [15].

DOI: 10.54254/2755-2721/2025.GL26773

Traditional safety mechanisms in autonomous vehicles can be categorized into three primary domains: perception systems, decision-making algorithms, and actuation mechanisms [2].

Perception systems form the sensory foundation of autonomous vehicle safety, utilizing combinations of cameras, LiDAR, radar, and ultrasonic sensors to create comprehensive environmental awareness [16]. Wang et al. (2020) provide a comprehensive analysis of autonomous vehicle safety, noting that traditional perception systems excel in structured environments but face significant challenges in complex, dynamic scenarios with multiple interacting agents [2]. The limitations of sensor-based perception become particularly apparent in adverse weather conditions, complex urban environments, and scenarios involving human-vehicle interactions that require contextual understanding beyond simple object detection [17].

Decision-making algorithms in traditional autonomous vehicle systems typically rely on rule-based approaches, finite state machines, or classical machine learning techniques to process sensor data and determine appropriate responses [18]. Mariani (2018) identifies several critical limitations in these approaches, including difficulty in handling edge cases, limited adaptability to novel scenarios, and challenges in balancing multiple competing objectives simultaneously [6]. These limitations become particularly problematic when vehicles must make split-second decisions involving trade-offs between different types of safety risks or when facing scenarios not explicitly programmed into the system [19].

2.1. Large language models in transportation and safety applications

The application of Large Language Models to transportation and safety domains represents a relatively recent but rapidly expanding area of research [3]. Cui et al. (2024) provide a comprehensive survey of multimodal large language models for autonomous driving, highlighting the potential for LLMs to enhance various aspects of autonomous vehicle operation through improved reasoning, contextual understanding, and natural language interaction capabilities [1].

Fu et al. (2024) demonstrate the potential for LLMs to enable more human-like driving behavior through their work on "Drive like a human," which explores how large language models can be integrated with autonomous driving systems to improve decision-making in complex scenarios [4]. Their research reveals that LLMs can effectively process and reason about driving scenarios described in natural language, potentially enabling more sophisticated threat assessment and response planning [4].

Wang et al. (2023) specifically address safety perspectives in LLM-empowered autonomous driving, proposing frameworks for integrating language model capabilities with safety-critical decision-making processes [5]. Their work identifies key challenges in ensuring reliability and safety when incorporating LLM reasoning into autonomous vehicle systems, including issues related to model uncertainty, computational requirements, and real-time performance constraints [5].

2.2. Threat detection and cybersecurity in autonomous vehicles

The cybersecurity dimension of autonomous vehicle safety has gained increasing attention as vehicles become more connected and software-dependent [7]. Chowdhury et al. (2020) provide a comprehensive survey of attacks on self-driving cars and their countermeasures, identifying various threat vectors including sensor spoofing, communication interference, and software vulnerabilities [7].

Fu et al. (2024) introduce IoV-BERT-IDS, a hybrid network intrusion detection system that leverages large language models for cybersecurity in Internet of Vehicles (IoV) environments [20].

DOI: 10.54254/2755-2721/2025.GL26773

Their approach demonstrates how LLM capabilities can be applied to detect and respond to cyber threats in real-time, providing a foundation for understanding how similar techniques might be applied to broader safety challenges in autonomous vehicles [20].

2.3. Algorithmic decision-making and ethical considerations

The integration of LLMs into autonomous vehicle safety systems raises important questions about algorithmic decision-making and ethical considerations in safety-critical scenarios [18]. Lim and Taeihagh (2019) provide a comprehensive analysis of algorithmic decision-making in autonomous vehicles, examining both technical and ethical concerns related to automated decision-making in complex scenarios [18].

Their research identifies several key challenges in implementing sophisticated decision-making algorithms in autonomous vehicles, including the need to balance multiple competing objectives, handle uncertainty and incomplete information, and make decisions that align with human values and expectations [18]. These challenges become particularly complex when considering the integration of LLM capabilities, which introduce additional layers of complexity and potential unpredictability into the decision-making process [18].

2.4. Mechanical stress and vehicle dynamics in autonomous systems

The relationship between algorithmic decision-making and mechanical stress in autonomous vehicles represents a critical but underexplored area of research [8]. Gordon and Lidberg (2015) examine automated driving and autonomous functions on road vehicles, providing insights into the mechanical and dynamic considerations that must be integrated with autonomous driving algorithms [8].

Betz et al. (2019) provide a detailed analysis of an autonomous driving crash, revealing how mechanical failures and stress-related issues can interact with algorithmic decision-making to create safety hazards [10]. Their work highlights the importance of considering mechanical constraints and stress factors when designing autonomous vehicle safety systems [10].

Koren et al. (2018) introduce adaptive stress testing for autonomous vehicles, demonstrating methodologies for evaluating how autonomous systems perform under various stress conditions [19]. Their approach provides a framework for understanding how mechanical and algorithmic stress factors can be systematically evaluated and optimized [19].

3. Traditional vehicles vs. NEVs

Traditional vehicles, powered by internal combustion engines (ICEs), have a long history of safety system development. Their safety control mechanisms are primarily based on mechanical and electro-mechanical systems designed to mitigate the effects of collisions and assist the driver in maintaining control. These systems can be broadly categorized into passive and active safety systems.

Passive safety systems, such as seatbelts, airbags, and crumple zones, are designed to protect occupants during a crash. The design of the vehicle's body structure, such as body-on-frame or monocoque construction, plays a crucial role in crash energy management. Body-on-frame designs, common in older vehicles and trucks, offer robustness but are heavy and less effective at absorbing impact energy compared to modern monocoque designs, which integrate the body and frame into a single structure for improved rigidity and crash performance.

DOI: 10.54254/2755-2721/2025.GL26773

Active safety systems, such as Anti-lock Braking Systems (ABS), Electronic Stability Control (ESC), and Traction Control Systems (TCS), are designed to prevent accidents by assisting the driver in controlling the vehicle. These systems rely on sensors to monitor vehicle dynamics and intervene by adjusting braking pressure or engine power. The algorithms controlling these systems are typically rule-based and calibrated for a wide range of driving conditions. However, they are limited by the mechanical response times of the braking and steering systems and the inherent complexity of predicting vehicle behavior in all possible scenarios.

3.1. Pros and cons of traditional vehicles safety control mechanisms

Traditional safety control mechanisms have several advantages. They are mature, well-understood, and have a proven track record of reducing fatalities and injuries. The mechanical components are robust and reliable, and the control algorithms have been refined over decades of real-world experience. However, these systems also have significant limitations, particularly in the context of autonomous driving.

The primary disadvantage of traditional safety systems is their reactive nature. They typically respond to events as they occur, with limited ability to predict and avoid hazards proactively. The control algorithms are often based on simplified models of vehicle dynamics and may not perform optimally in complex or unexpected situations. Furthermore, the mechanical actuation systems, such as hydraulic brakes and power steering, have inherent delays that can limit the effectiveness of rapid, precise control actions required for autonomous navigation.

The separation of different safety functions into individual control modules also presents challenges. For example, the braking, steering, and suspension systems are often controlled by separate ECUs with limited communication and coordination. This lack of integration can lead to suboptimal performance when multiple systems need to work together to respond to a critical event.

3.2. LLM integration and its influence over parameters based on traditional vehicles

The integration of Large Language Models (LLMs) into traditional vehicle architectures presents both opportunities and challenges. LLMs can potentially enhance the decision-making capabilities of existing safety systems by providing a higher level of situational awareness and contextual understanding. For example, an LLM could analyze data from multiple sensors to identify complex traffic patterns or predict the intentions of other road users, providing valuable input to the vehicle's control systems.

However, the effectiveness of LLM-based control is ultimately limited by the mechanical constraints of the traditional vehicle platform. The slow response times of hydraulic brakes and conventional steering systems can create a bottleneck, preventing the vehicle from executing the rapid and precise maneuvers that an LLM might deem necessary. Furthermore, the parameters of existing control algorithms are often hard-coded and not designed to be dynamically adjusted by an external system like an LLM. Retrofitting traditional vehicles with LLM-based control would require significant modifications to the vehicle's hardware and software architecture, a complex and costly undertaking.

3.3. Reconfiguration of parameters in catering with the need of NEVs

New Energy Vehicles (NEVs), particularly Battery Electric Vehicles (BEVs), offer a fundamentally different platform for implementing advanced safety systems. The architecture of a typical NEV is

DOI: 10.54254/2755-2721/2025.GL26773

centered around a large battery pack, electric motors, and a sophisticated electronic control system.

In an NEV, the parameters of the control systems can be reconfigured in software to a much greater extent than in a traditional vehicle. For example, the torque response of the electric motors can be precisely controlled and adjusted in real-time, allowing for rapid and accurate acceleration and deceleration. The regenerative braking system can be used to supplement the friction brakes, providing additional braking force and improving energy efficiency. The steering system is typically electric, enabling precise and rapid changes in direction.

This high degree of software control allows for a much deeper integration of LLMs into the vehicle's safety systems. An LLM can directly influence the parameters of the motor controller, the braking system, and the steering system, enabling a much more dynamic and adaptive response to changing conditions. For example, an LLM could adjust the torque distribution between the wheels to improve stability during a high-speed maneuver, or it could modulate the regenerative braking to optimize both safety and energy efficiency.

By reconfiguring the parameters of the vehicle's control systems in real-time, an LLM can effectively transform the vehicle's mechanical characteristics to meet the demands of the current driving situation. This ability to dynamically adapt the vehicle's behavior is a key advantage of NEVs and a critical enabler for achieving the highest levels of autonomous driving safety.

4. Conclusion

This review has provided a comprehensive analysis of the integration of Large Language Models (LLMs) into autonomous vehicle self-protection mechanisms, with a particular focus on the transition from traditional mechanical structures to the advanced architectures of New Energy Vehicles (NEVs). Our findings indicate that while traditional vehicles present significant mechanical and architectural barriers to the full realization of LLM capabilities, NEVs offer a promising platform for developing highly adaptive and intelligent safety systems.

The core advantages of NEVs lie in their software-defined nature, which allows for the dynamic reconfiguration of control parameters. This flexibility enables LLMs to move beyond a purely advisory role and become active participants in the vehicle's control loop, adjusting parameters related to motor torque, braking, and steering in real-time4. Conclusion

This review has provided a comprehensive analysis of the integration of Large Language Models (LLMs) into autonomous vehicle self-protection mechanisms, with a particular focus on the transition from traditional mechanical structures to the advanced architectures of New Energy Vehicles (NEVs). Our findings indicate that while traditional vehicles present significant mechanical and architectural barriers to the full realization of LLM capabilities, NEVs offer a promising platform for developing highly adaptive and intelligent safety systems.

The core advantages of NEVs lie in their software-defined nature, which allows for the dynamic reconfiguration of control parameters. This flexibility enables LLMs to move beyond a purely advisory role and become active participants in the vehicle's control loop, adjusting parameters related to motor torque, braking, and steering in real-time. This deep integration of AI with the vehicle's mechanical systems is crucial for achieving the levels of safety and reliability required for widespread autonomous vehicle adoption.

However, several challenges remain. The development of robust and verifiable safety protocols for LLM-based control systems is paramount. Ensuring the security of these systems against cyber threats is another critical concern. Furthermore, extensive testing and validation will be required to ensure that these systems perform reliably in the vast range of conditions encountered in real-world driving.

DOI: 10.54254/2755-2721/2025.GL26773

Future research should focus on developing standardized frameworks for LLM integration into vehicle architectures, creating robust testing and validation methodologies, and addressing the ethical and legal implications of AI-driven vehicle control. By addressing these challenges, we can unlock the full potential of LLMs to create a new generation of autonomous vehicles that are not only intelligent but also fundamentally safer.

References

[1] Cui, C., Ma, Y., Cao, X., Ye, W., Zhou, Y., & Liang, K. (2024). A survey on multimodal large language models for autonomous driving. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 449-458. https://openaccess.thecvf.com/content/WACV2024W/LLVM-AD/html/Cui_A_Survey_on_Multimodal_Large_Language_Models_for_Autonomous_Driving_WACVW_2024_paper.html

[2] Wang, J., Zhang, L., Huang, Y., & Zhao, J. (2020). Safety of autonomous vehicles. Journal of Advanced Transportation, 2020, 8867757. https://onlinelibrary.wiley.com/doi/abs/10.1155/2020/8867757

[3] Adekanye, O.A.M. (2024). LLM-powered synthetic environments for self-driving scenarios. Proceedings of the AAAI Conference on Artificial Intelligence, 38(21), 7-14. https://ojs.aaai.org/index.php/AAAI/article/view/30540

[4] Fu, D., Li, X., Wen, L., Dou, M., Cai, P., Shi, B., & Qiao, Y. (2024). Drive like a human: Rethinking autonomous driving with large language models. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision, 234-242. https://ieeexplore.ieee.org/abstract/document/10495699/

[5] Wang, Y., Jiao, R., Zhan, S.S., Lang, C., Huang, C., & Li, Z. (2023). Empowering autonomous driving with large language models: A safety perspective. arXiv preprint arXiv:2312.00812. https://arxiv.org/abs/2312.00812

[6] Mariani, R. (2018). An overview of autonomous vehicles safety. 2018 IEEE International Reliability Physics Symposium, 78-85. https://ieeexplore.ieee.org/abstract/document/8353618/

[7] Chowdhury, A., Karmakar, G., Kamruzzaman, J., Jolfaei, A., & Das, R. (2020). Attacks on self-driving cars and their countermeasures: A survey. IEEE Communications Surveys & Tutorials, 22(2), 1194-1235. https://ieeexplore.ieee.org/abstract/document/9257492/

[8] Gordon, T.J., & Lidberg, M. (2015). Automated driving and autonomous functions on road vehicles. Vehicle System Dynamics, 53(7), 958-994. https://www.tandfonline.com/doi/abs/10.1080/00423114.2015.1037774

[9] Miao, L. (2025). First fatal accident involving Xiaomi SU7 claims three lives on Chinese highway. CarNewsChina.com. https://carnewschina.com/2025/04/01/first-fatal-accident-involving-xiaomi-su7-claims-three-lives-on-chinese-highway/

[10] Betz, J., Heilmeier, A., Wischnewski, A., Stahl, T., & Lienkamp, M. (2019). Autonomous driving crash explained in detail. Applied Sciences, 9(23), 5126. https://www.mdpi.com/2076-3417/9/23/5126

[11] Kong, X., Braunl, T., Fahmi, M., & Liu, Y. (2024). A superalignment framework in autonomous driving with large language models. 2024 IEEE Intelligent Vehicles Symposium, 14-21. https://ieeexplore.ieee.org/abstract/document/10588403/

[12] National Safety Council. (2017). Introduction - Injury Facts. Injury Facts. https://injuryfacts.nsc.org/motor-vehicle/overview/introduction/

[13] Transactions on Machine Learning Research. (2024). Evaluating World Models with LLM for Decision Making. https://openreview.net/forum?id=xxJ41g3gyk

[14] Cui, C., Ma, Y., Cao, X., Ye, W., & Zhou, Y. (2024). Drive as you speak: Enabling human-like interaction with large language models in autonomous vehicles. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 147-155. https://openaccess.thecvf.com/content/WACV2024W/LLVM-AD/html/Cui_Drive_As_You_Speak_Enabling_Human-Like_Interaction_With_Large_Language_WACVW_2024_paper.html

[15] Barabas, I., Todoru, A., Cordo, N., & Blan, H. (2017). Current challenges in autonomous driving. IOP Conference Series: Materials Science and Engineering, 252(1), 012096. https://iopscience.iop.org/article/10.1088/1757-899X/252/1/012096/meta

[16] Alsaade, F.W., & Al-Adhaileh, M.H. (2023). Cyber attack detection for self-driving vehicle networks using deep autoencoder algorithms. Sensors, 23(8), 4086. https://www.mdpi.com/1424-8220/23/8/4086

[17] Chattopadhyay, A., Lam, K.Y., & Tavva, Y. (2020). Autonomous vehicle: Security by design. IEEE Transactions on Intelligent Transportation Systems, 21(11), 4629-4642. https://ieeexplore.ieee.org/abstract/document/9130160/

[18] Lim, H.S.M., & Taeihagh, A. (2019). Algorithmic decision-making in AVs: Understanding ethical and technical concerns for smart cities. Sustainability, 11(20), 5791. https://www.mdpi.com/2071-1050/11/20/5791

DOI: 10.54254/2755-2721/2025.GL26773

[19] Koren, M., Alsaif, S., Lee, R., & Kochenderfer, M.J. (2018). Adaptive stress testing for autonomous vehicles. 2018 IEEE Intelligent Vehicles Symposium, 1-7. https://ieeexplore.ieee.org/abstract/document/8500400/
[20] Fu, M., Wang, P., Liu, M., Zhang, Z., & Li, X. (2024). IoV-BERT-IDS: Hybrid network intrusion detection system in IoV using large language models. IEEE Transactions on Intelligent Transportation Systems, 25(8), 7174-7184. https://ieeexplore.ieee.org/abstract/document/10533857/