Security Risks and Governance Framework for Intelligent Agents in Transportation

Long Liang[a], Xiaobo Wu[*]

China Academy of Transportation Sciences, Beijing, China

[a]15117946572@163.com, *718439457@qq.com

Abstract. The integration of Large Language Models (LLMs) and intelligent agents is driving a paradigm shift in transportation from simple intelligence to collaborative autonomy. Agents with perception, reasoning, and closed-loop action capabilities are being deployed across water, land, and air scenarios. However, this rapid adoption introduces complex security challenges. This paper presents a comprehensive study based on real-world deployment cases, offering a quantitative analysis of application scenarios, core technologies, and risk correlations. Our analysis reveals that while LLMs and multimodal perception are dominant, significant security risks persist, particularly in operational safety and algorithmic reliability. We identify five critical risk dimensions: data security, algorithmic robustness, system architecture, business operations, and compliance governance. Finally, we propose a three-dimensional defense framework encompassing "Technical Protection," "Operational Control," and "Industry Governance." This study provides empirical evidence and strategic references for ensuring the safe deployment of intelligent agents in critical transportation infrastructure.

Keywords: Intelligent Agent, Intelligent Transportation, Security Risk, Prevention and Control, Large Language Models

1 Introduction

Intelligent transportation technologies such as large-scale artificial intelligence models, multi-agent systems, and digital twins are reshaping the operational model of the industry. In the past, traffic governance relied mainly on manual investigation and experience-based decision-making. Facing increasingly complex networked transportation systems, traditional methods were insufficient in terms of timeliness and struggled to quickly identify potential safety hazards.

The intelligent agent technology enables autonomous perception, cognitive reasoning, task planning, closed-loop execution, and natural language interaction, representing a technological breakthrough in the field of transportation[1]. This technology is gradually expanding from isolated single-point intelligence to full-scenario collaborative intelligence across road networks, railways, ports, and aviation; it is deeply integrated with large-scale transportation models and business rule engines, promoting the

© The Author(s) 2026
V. Erokhin et al. (eds.), *Proceedings of the 6th International Conference on Internet Finance and Digital Economy (ICIFDE 2026)*, Advances in Economics, Business and Management Research 412,
https://doi.org/10.2991/978-94-6239-830-6_19

shift from experience-driven to data-driven operation models. At the same time, large-scale implementation has brought new features such as data centralization, autonomous decision-making, and cloud-edge collaboration architectures, which have also given rise to new security challenges such as data integrity, algorithm risks, and compliance management.

Currently, in the field of large-scale model intelligent agents and traffic intelligence, a large amount of research results have been accumulated. Gan et al. have summarized the full-chain privacy and security risks of LLM intelligent agents[2], He et al. have confirmed through real cases that autonomous agents can amplify network vulnerabilities and trigger cascading failures[3], and a layered attack surface review has proposed a collaborative protection approach for technology, management, and regulations[4]. However, these studies are mostly limited to general software systems and have not conducted adaptation analysis for critical transportation infrastructure. OWASP and NIST have released AI security governance standards that provide a general risk classification and control framework[5][6], but they lack detailed implementation rules for comprehensive transportation. In related research in the transportation field, it has been confirmed that large-scale models have been widely applied in road networks, ports, and aviation[7][8], and the mainstream technical architecture is "base large model + industry knowledge base + multi-modal perception"; achievements in autonomous driving focus on single-vehicle perception and decision optimization[9], mostly discussing improvements in technical efficiency, while neglecting the security risks arising from large-scale deployment. Multi-agent collaboration research mainly focuses on performance optimization of traffic scheduling algorithms[10][11], rarely exploring security governance challenges such as cross-agent data sharing and ambiguous human-machine rights and responsibilities. In summary, existing research still has significant shortcomings: First, research scenarios are fragmented, mostly focusing on general intelligent agents or single autonomous driving scenarios, lacking a unified risk analysis covering the entire transportation industry; second, risk classification dimensions are disconnected, not synchronously integrating data, algorithms, systems, operations, and compliance to conduct correlated quantitative analysis; third, the protection framework lacks a closed-loop collaborative mechanism, and theoretical solutions lack support from large-scale real engineering cases.

This paper reviews the current application status of intelligent agents in the transportation field, conducts quantitative statistical analysis based on domestic implementation cases, identifies various security risks in the implementation process of intelligent agents, and builds hierarchical and graded response strategies to fill the gaps in existing research, providing empirical evidence and governance references for the stable deployment of intelligent agents for critical transportation infrastructure.

2 Development Status of Intelligent Agents in Transportation

2.1 Concept and Development Background

A transportation agent is an intelligent entity built upon LLMs and multimodal models, integrated with industry expertise and business rules. It operates autonomously with

capabilities in environmental perception, logical inference, and human-computer interaction. Development is driven by: Policy: National plans explicitly supporting large models and digital twins in infrastructure. Technology: Maturation of multimodal perception, vehicle-road-cloud integration, and RAG. Market Demand: Needs for cost reduction, safety enhancement, and precise governance.

2.2 Overall Application Pattern

Transportation agents have achieved comprehensive coverage across water, land, and air transport. Official cases from the Ministry of Transport constitute a significant portion, supplemented by local projects and sector-specific initiatives such as "AI+Expressway," collectively forming a multi-source case library. In terms of business functions, operation and service, safety and regulation, and technological innovation are the predominant application directions, whereas government administration and comprehensive management constitute a smaller but emerging segment (see detailed statistical analysis in Section 2.4).

2.3 Industry Characteristics and Common Challenges

The technology architecture has converged on "foundational large model + traffic vertical knowledge base + multimodal perception + multi-agent collaboration + digital twin". However, challenges persist: Lack of Standards: Absence of common technical systems and safety audit standards. Digital Divide: Weak model generalization across different regions. Immature Interaction: Weak dynamic decision-making in complex situations. Security Lag: Safety governance struggling to keep pace with application technologies.

2.4 Multidimensional Statistical Analysis

To analyze the industry situation, we collected case libraries from the official website of the Ministry of Transport, national special cases of artificial intelligence + expressways, Shanghai's announced and winning smart transportation projects, smart transportation demonstration scenarios in Guangdong Province, and intelligent transportation application scenarios in Gansu Province. These multi-source industry cases collected in this paper constitute the primary case library of this study. From the perspective of the sources of the cases (Figure 1), the Ministry of Transport's official cases account for 58%, while local supplementary projects account for 42%, which is a relatively balanced distribution.

From the perspective of application scenarios, intelligent transportation systems have permeated all aspects of integrated transportation. As shown in Figure 2, the number of implementation cases for the road network is the highest, followed by the port and maritime scenarios, with the combined proportion exceeding 62%. Thus, it can be seen that in traditional core transportation industries such as highways and ports, which involve heavy assets and high safety requirements, intelligent systems have formed a large-scale and in-depth implementation pattern.

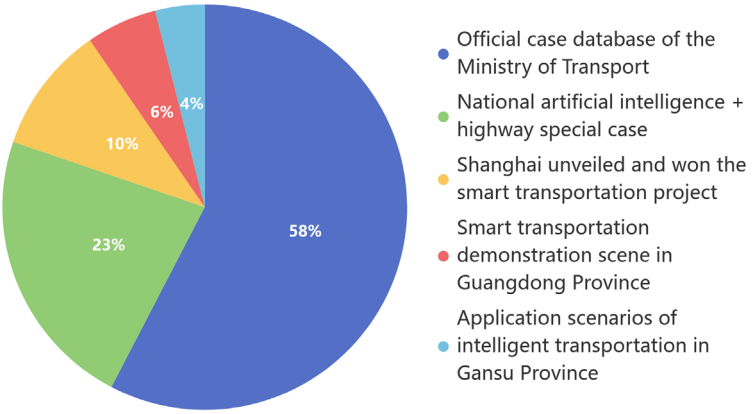


Fig. 1. Source Distribution Map of Case Database.

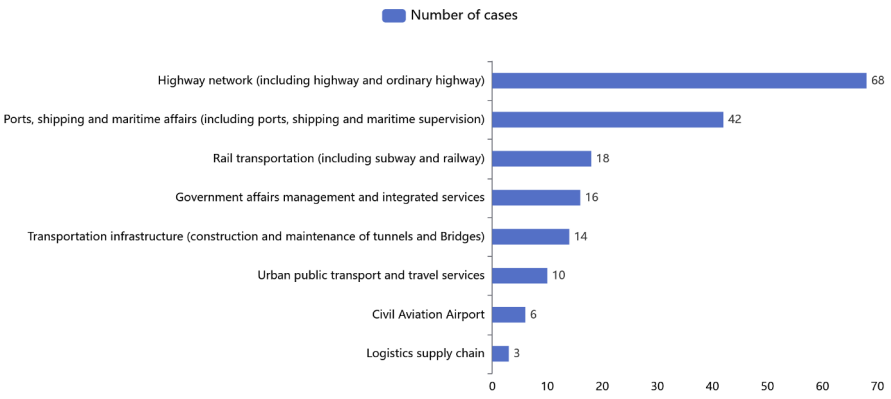


Fig. 2. Distribution of Intelligent Agent Applications across Transportation Fields.

From the perspective of business functions, following the classification framework of the Ministry of Transport, the cases are categorized into four groups: Technology and Innovation, Operation and Service, Safety and Regulation, and Government Administration and Comprehensive Management. As shown in Figure 3, based on statistics from 177 cases, Operation and Service accounts for the largest share (32%), followed closely by Safety and Regulation (30%) and Technology and Innovation (27%), while Government Administration and Comprehensive Management represents 11%. This distribution indicates that agent applications have moved beyond technology validation and become deeply embedded in core operational practices and safety management, whereas governance scenarios are still in their early stages but show considerable potential.

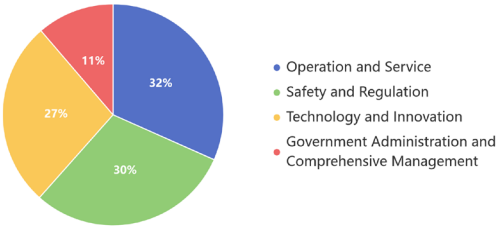


Fig. 3. Distribution of Intelligent Agent Application Cases by Business Function.

The preceding three figures elaborate the overall landscape of transportation intelligent agents from the perspectives of case sources, industrial sectors and business categories, delivering a qualitative overview of prevalent practical scenarios. To supplement quantitative evidence on underlying computational infrastructures and address the insufficient technical quantification in prior literature, this study manually extracts core technical tags from the full sample of 177 cases. The deployment penetration of diverse computer and AI technologies is visualized in Figure 4.

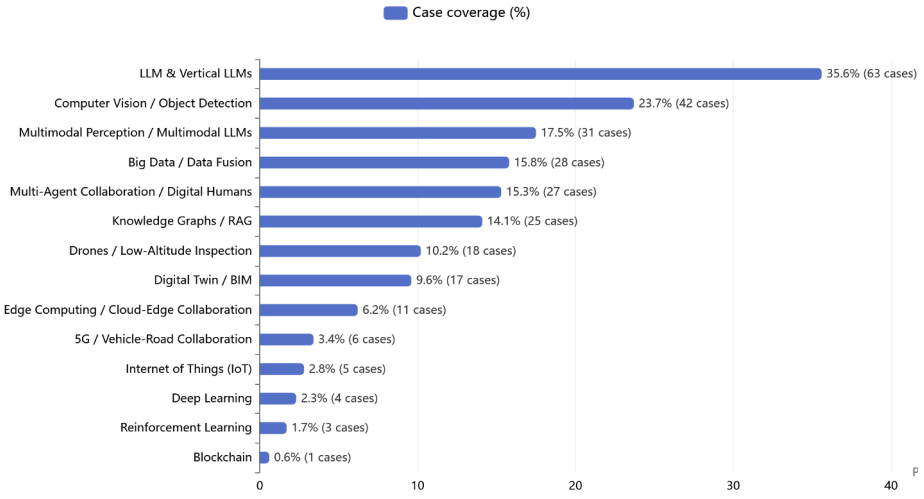


Fig. 4. Distribution of Core Computer and AI Technologies in Transportation Intelligent Agent Cases.

The statistical outcomes in Figure 4 expose stark disparities in the practical adoption of different computational techniques. Built upon such empirical data, Table 1 summarizes the eight primary technical branches and their intrinsic functions, while Figure 4 details their finer-grained deployment penetration (including sub-technologies such as big data, IoT, and 5G beyond the eight branches). Notably, although big data / data fusion exhibits a higher penetration (15.8%) than several primary branches (e.g.,

knowledge graphs / RAG at 14.1%), it is framed as a cross-domain enabling technology rather than an independent technical branch, because it underpins multiple agent capabilities instead of constituting a standalone one. Table 2 maps these technical modules to five risk layers—Data, Algorithmic, System, Operational and Human-Machine Collaboration, and Compliance—grading hazard severity and sorting core threats; these five dimensions are examined in detail in Section 3.

Table 1. Distribution of Core Technologies.

Category	Key Characteristics	Prevalence (n=177)
LLM & Vertical LLMs	Core foundation of the industry; enables semantic understanding and autonomous decision-making.	35.6% (n=63)
Computer Vision / Object Detection	Applied to tasks such as road defect recognition and object detection.	23.7% (n=42)
Multimodal Perception / Multimodal Large Models	Integrates visual, speech, and other multimodal data to enhance comprehensive environmental perception.	17.5% (n=31)
Multi-Agent Collaboration (including multiple Agents)	Facilitates collaborative decision-making among agents, driving evolution from single-point intelligence to system-wide intelligence.	15.3% (n=27)
Knowledge Graphs / RAG (Retrieval-Augmented Generation)	Supports regulatory intelligent retrieval, knowledge reasoning, and augmented generation.	14.1% (n=25)
Drones / Low-Altitude Inspection	Used for autonomous low-altitude flight inspection and monitoring.	10.2% (n=18)
Digital Twin	Constructs digital mirrors for simulation optimization and predictive maintenance.	9.6% (n=17)
Edge Computing / Cloud-Edge Collaboration	Provides low-latency real-time processing and supports cloud-edge-device synergy.	6.2% (n=11)

Table 2. Security Risk Correlation Analysis.

Risk Dimension	Prominence	Key Concerns
Data Security	Medium	Privacy leakage, data tampering, uncontrolled sharing.
Algorithmic Security	High	Misjudgment, black-box decisions, generalization failure in anomalies.
System Security	Medium-High	Edge terminal vulnerabilities, network communication threats.
Operational and Human-Machine Collaboration	High	Human-machine collaboration, emergency handling, liability ambiguity.
Compliance & Governance	Emerging	Lack of standards, data compliance, supply chain risks.

3 Security Risk Analysis

As agents evolve to comprehensive applications, their operation subverts traditional security boundaries, and algorithm vulnerability becomes the core hidden danger of intelligent transportation.

3.1 Data Security Risks

Intelligent agents rely heavily on multi-source sensitive traffic data during perception and decision-making. The whole-cycle collection, transmission and sharing of data bring prominent hidden dangers such as leakage, tampering and uncontrolled circulation, which constitute the basic bottom-line security risks of the whole system.

1. Leakage and Unauthorized Access: Sensitive spatial and industrial data, including UAV patrol imagery, face risks of leakage through internal misuse or external theft without proper desensitization.

2. Tampering and Pollution: Attackers could compromise edge terminals to forge sensor signals or tamper with positioning data, inducing false warnings or incorrect dispatch instructions.

3. Uncontrolled Sharing: Cross-departmental transmission lacks robust traceability, leading to improper data use.

3.2 Algorithm and Model Security Risks

Large language models and multimodal perception models serve as the core cognitive modules of intelligent agents. Defects such as poor generalization, black-box decision-making and vulnerability to adversarial interference will directly cause misjudgment in traffic scenarios, which are high-priority unique risks of intelligent transportation[12].

1. Misjudgment and Generalization Failure: Models trained on specific data may fail during extreme weather or emergencies, generating false alarms.

2. Opacity and Black-Box Decisions: Unobservable decision-making processes hinder post-accident investigations and liability assignment.

3. Adversarial Attacks: Malicious samples or physical interference can deceive visual detection agents[13][14].

4. Iteration Shock: Frequent updates can cause incompatibilities with existing business rules.

3.3 System and Architecture Security Risks

Transportation intelligent agents widely adopt cloud-edge-terminal collaborative architecture supported by digital twins and 5G IoT networks. Exposed edge devices, unencrypted transmission links and insufficient platform redundancy form systemic architecture vulnerabilities that can trigger large-scale cascading failures.

1. Edge Terminal Vulnerabilities: Physically exposed edge terminals with outdated firmware are easy targets for network hijacking[15].

2. Network Communication Threats: Reliance on 5G/IoT links makes systems vulnerable to eavesdropping and command injection.

3. Digital Twin Manipulation: Tampered backend data can misalign digital twins with real-world operations, misleading emergency command.

4. Single Point of Failure: Insufficient redundancy in core nodes can lead to cascading failures.

3.4 Operational and Human-Machine Collaboration Risks

Intelligent agents have been embedded into daily traffic operation, emergency disposal and supervision workflows. The new human-machine collaborative mode breaks traditional operation management logic and brings practical risks including operator capability decline, inefficient emergency response and ambiguous accident liability.

1. Degradation of Human Capability: Long-term reliance on agents can degrade operator skills, hindering effective intervention during failures.

2. Weak Emergency Response: Agents optimized for fixed scenarios struggle with dynamic task planning in mixed-risk emergencies.

3. Ambiguous Liability: Lack of clear human-machine responsibility frameworks complicates liability assignment in accidents.

4. Lack of Coordinated Response: Absence of standardized protocols hinders unified emergency linkage across regions.

3.5 Compliance and Governance Risks

The rapid iteration of intelligent agent technologies outpaces the construction of supporting industrial standards and regulatory systems. Vacant technical specifications, incomplete data compliance constraints and unregulated supply chains create long-term hidden compliance risks for the industry.

1. Lack of Standards: Absence of unified technical specifications and safety access rules creates a lack of baseline constraints[16].

2. Data Compliance Risks: Collection of video and trajectory data carries non-compliance risks if legal requirements are not met.

3. Lagging Supervision: Regulation has not caught up with autonomous decision-making capabilities.

4. Supply Chain Risks: Reliance on open-source frameworks introduces backdoor and service interruption risks.

4 Countermeasures and Prevention Strategies

This paper proposes a three-dimensional progressive defense architecture consisting of technical protection, operational control, and industry governance. As shown in Figure 5, this architecture takes the form of a pyramid: the bottom layer of technical protection serves as the security foundation of the entire system; the middle layer of operational control transforms technical capabilities into institutionalized management practices;

and the top layer of industry governance provides institutional guarantees for the entire system through standards and regulations. The three layers form a closed-loop feedback mechanism of "technical empowerment of operation, operation feeding back to governance, and governance driving technology and operation".

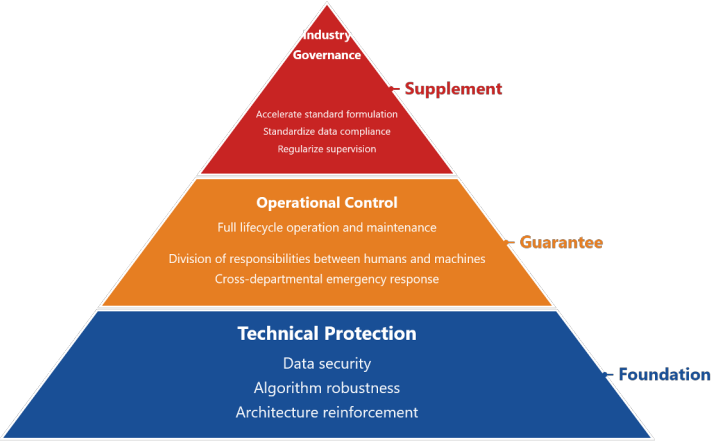


Fig. 5. Three-Tier Pyramid Architecture of Industry Governance.

4.1 Technical Protection Measures

Technical defense is the fundamental barrier against intelligent agent security threats. Corresponding protective technologies are deployed around data full lifecycle, algorithm robustness and cloud-edge integrated architecture to realize active risk interception at the technical layer.

1. Strengthen Full Lifecycle Data Security: Establish full-life-cycle hierarchical data management; desensitize private data and encrypt all transmission and storage links, and introduce federated learning to avoid raw data outflow[17][18]. Establish audit and traceability platforms.

2. Enhance Algorithmic Robustness: Expand training datasets with extreme scenarios and use adversarial training. Develop interpretable AI for traceability. Implement online security testing and manual review for high-risk judgments.

3. Fortify Cloud-Edge Architecture: Regularly patch edge nodes. Build cloud-edge-terminal integrated defense system: use private networks and 5G slicing for command isolation, and deploy firewalls and real-time intrusion detection modules[19]. Ensure redundancy for digital twin platforms.

4.2 Operational and Human-Machine Management

Technical protection requires standardized operation management mechanisms to achieve full-scenario implementation. This section optimizes operation rules from

equipment full-lifecycle maintenance, human-machine responsibility division and cross-regional emergency linkage to eliminate management loopholes in practical business.

1. Improve Lifecycle O&M Management: Implement safety management from design to decommissioning. Conduct regular equipment inspections and algorithm audits. Establish emergency plans for system anomalies.

2. Clarify Human-Machine Responsibilities: Adopt "smart assistance, human-in-command." Reserve manual review for extreme emergencies. Retain daily judgment mechanisms to prevent skill degradation. Define clear liability limits.

3. Build Cross-Departmental Emergency System: Open data interfaces between transportation, public security, and civil aviation. Establish unified information sharing and command mechanisms. Unify incident classification and response protocols.

4.3 Industry Governance and Compliance

Sustainable safe operation of transportation intelligent agents relies on top-level industrial governance constraints. This section proposes governance solutions covering standard formulation, unified data compliance guidelines and normalized supervision to fill the gap of lagging industry regulation.

1. Accelerate Standardization: Accelerate the construction of industry norms, formulate unified standards for technology architecture, safety protection, and performance evaluation to provide a clear supervision baseline for intelligent transportation project construction[20].

2. Standardize Data Compliance: Standardize industry data compliance management, enforce strict adherence to national legal requirements for data collection, storage, and use, and establish unified industry-wide data governance guidelines.

3. Enhance Supervision Mechanisms: Develop long-term normalized supervision mechanisms for autonomous decision-making, and establish full-process supply chain security auditing systems for third-party models and open frameworks.

5 Conclusion

The deployment of intelligent agents in transportation offers transformative potential but introduces multifaceted security risks. Through quantitative analysis of 177 real-world deployment cases, this paper identifies critical vulnerabilities in data, algorithms, systems, operations, and governance, with algorithmic and operational risks being particularly prominent. The proposed three-dimensional defense framework provides a structured approach to mitigating these risks. This study has certain limitations: the case library is primarily domestic, and risk assessment remains largely qualitative. Future research will focus on developing quantitative risk metrics and validating the framework through field deployments, as well as accelerating industry standardization, enhancing the interpretability of traffic large models, and establishing standardized human-machine collaboration emergency protocols to ensure the safe and sustainable evolution of intelligent transportation systems.

Acknowledgements

This work has been supported by Basal Research Fund of Central Public Research Institute of China (Grant No.20250208).

References

1. N. A. Kose and K. Kose. 2026. From concrete to code: a survey of AI-driven transportation infrastructure, security, and human interaction. Sensors 26, 7 (2026), 2219. https://doi.org/10.3390/s26072219.
2. Y. Gan, Y. Yang, Z. Ma, P. He, R. Zeng, Y. Wang, Q. Li, C. Zhou, S. Li, and T. Wang, "Navigating the risks: A survey of security, privacy, and ethics threats in LLM-based agents," ACM Trans. Softw. Eng. Methodol., 2026, arXiv:2411.09523.
3. F. He, T. Zhu, D. Ye, B. Liu, W. Zhou, and P. S. Yu, "The Emerged Security and Privacy of LLM Agent: A Survey with Case Studies," ACM Comput. Surv., vol. 58, no. 6, Dec. 2025, doi:10.1145/3773080.
4. K. Chu, "A Systematic Survey of Security Threats and Defenses in LLM-Based AI Agents: A Layered Attack Surface Framework," arXiv preprint arXiv:2604.23338, 2026.
5. OWASP Foundation, OWASP Top 10 for Agentic Applications 2026, OWASP Ind. Stand., Dec. 2025. [Online]. Available: https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/
6. Nat. Inst. Standards Technol., NIST AI Risk Management Framework, U.S. NIST, 2023. [Online]. Available: https://www.nist.gov/itl/ai-risk-management-framework
7. N. Maksoud, H. AlJassmi, L. Ali, and A. R. Masoud, "Applications of large language models and generative AI in transportation: A systematic review and bibliometric analysis," Results Eng., vol. 26, p. 101699, 2025, doi:10.1016/j.rineng.2025.101699.
8. T. Nie, J. Sun, and W. Ma, "Exploring the Roles of Large Language Models in Reshaping Transportation Systems: A Survey, Framework, and Roadmap," arXiv preprint arXiv:2503.21411, Mar. 2025.
9. Y. Zhu, S. Wang, W. Zhong, N. Shen, Y. Li, S. Wang, Z. Li, C. Wu, Z. He, and L. Li, "A survey on large language model-powered autonomous driving," Engineering, 2025, doi:10.1016/j.eng.2025.07.038.
10. J. Liang, X. Du, Y. Liu, Z. Li, H. Jiang, Z. Li, C. Long, H. Wei, and W. Ketter, "A Survey on Multi-agent Reinforcement Learning for Adaptive Transportation Solutions," SN Comput. Sci., vol. 6, no. 8, p. 955, 2025, doi:10.1007/s42979-025-04475-3.
11. A. Aloui, H. Hachicha, and E. Zagrouba, "Multi-Agent-Based Approaches for Cooperative Traffic Management in C-ITS: Systematic Literature Review," IEEE Trans. Intell. Transp. Syst., 2026, doi:10.1109/TITS.2025.3628897.
12. M. Yu, F. Meng, X. Zhou, et al. 2025. A survey on trustworthy LLM agents: threats and countermeasures. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25). ACM, New York, NY, 6216–6226. https://doi.org/10.1145/3711896.3736561.
13. X. Feng, L. F. Mao, T. T. Xu, et al. 2025. A review on the security of large language models for cooperative vehicle-road-centralized autonomous driving. Journal of Communications 46, 11 (2025), 291–307.
14. T. Y. Zhang, J. F. Liu, Y. K. Guo, et al. 2026. Towards secure and robust vision-language models in autonomous driving: a survey for perception-oriented and decision-oriented attacks. Chinese Journal of Electronics (2026). https://doi.org/10.23919/cje.2025.00.

15. S. Nayak, O. Gungor, and T. Rosing. 2026. Security in collaborative driving: a survey of threats, defenses, and emerging trends. Electronics 15, 11 (2026), 2389. https://doi.org/10.3390/electronics15112389.
16. T. Pavleska, M. Masi, G. P. Sellitto, et al. 2025. Architecture-based governance for secure-by-design cooperative intelligent transport systems. Vehicular Communications 55 (2025), 100967. https://doi.org/10.1016/j.vehcom.2025.100967.
17. U.S. Department of Transportation. 2025. Sector-Specific Adaptation of Intelligent Transportation Systems Cybersecurity Profile. ITS Knowledge Resources System (2025).
18. Y. Mu, H. Chen, and H. Li. 2024. Research progress on security and privacy protection technologies of large language models. Journal of Cybersecurity Science 2, 1 (2024), 40–49. https://doi.org/10.20172/j.issn.2097-3136.240103.
19. C. Chen, Y. Ren, and X. Zhong. 2025. Research on the construction of an integrated and deeply-practical security protection system for intelligent connected vehicles. Information Security and Communication Confidentiality 9 (2025), 43–51.
20. X. Zhang, C. Li, N. Xu, et al. 2025. Security challenges and response mechanisms for trusted large language model agents. Information Communication Technology and Policy 51, 1 (2025), 33–39.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

