OPEN ACCESS

COMMENTARY

The Innovation

Toward the robustness of autonomous vehicles in the AI era

Siheng Chen,[1,8,*] Yiyi Liao,[2,8] Fei Wang,[3,4,8] Gang Wang,[5,8] Liang Wang,[6,8] Yafei Wang,[1,8] and Xichan Zhu[7,8]

[1]School of Artificial Intelligence, Shanghai Jiao Tong University, Shanghai 200240, China

[2]Zhejiang University, Hangzhou 310027, China

[3]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

[4]University of Chinese Academy of Sciences, Beijing 100049, China

[5]Beijing Institute of Technology, Beijing 100081, China

[6]Tsinghua University, Beijing 100080, China

[7]Tongji University, Shanghai 200092, China

[8]These authors contributed equally

*Correspondence: sihengc@sjtu.edu.cn

Received: December 28, 2023; Accepted: December 25, 2024; Published Online: March 3, 2025; https://doi.org/10.1016/j.xinn.2024.100780

Citation: Chen S., Liao Y., Wang F., et al., (2025). Toward the robustness of autonomous vehicles in the AI era. The Innovation **6(3)**, 100780.

INTRODUCTION

In modern transportation, autonomous driving (AD) stands at the forefront of the AI technological revolution, promising to transform the way we commute and interact with urban environments.[1,2] This shift toward automation not only offers increased convenience and reduced human error in driving but also opens avenues for substantial economic and environmental benefits. However, accidents involving self-driving and driver assistance systems underscore the critical challenges in ensuring their robustness. According to the National Highway Traffic Safety Administration, Tesla's Autopilot has been associated with 17 fatalities and 736 crashes since 2019. Such statistics highlight the urgent need for robustness in AD systems, which operate reliably across diverse real-world scenarios, including varying weather conditions, complex traffic patterns, and unforeseen road events.[3] The criticality of robustness in ensuring public trust and the practical viability on a global scale can never be underestimated.[4] Developing AD systems that can handle these real-world challenges is not only a technological imperative but a societal necessity.

Figure 1 illustrates the key elements and challenges in developing robust AD systems. It highlights the diverse data sources needed, core system functions, and the importance of high-fidelity simulations. It also emphasizes integrating control theory and large language models to enhance system reliability and specificity. Empowerment technologies like deep generative models and reinforcement learning are crucial for balancing innovation and safety toward higher autonomy levels.

CHALLENGES IN ACHIEVING ROBUSTNESS IN AD

In the quest for robustness in AD, one paramount challenge is the acquisition of comprehensive real-world data, upon which the effectiveness of autonomous systems relies to a large extent. However, collecting large-scale real-world data is resource intensive. It is estimated that 1 million kilometers of driving data could cost around 1 billion RMB. While virtual simulations and digital twins offer promising alternatives, these methods still face limitations, especially in adapting to complex situations and in the detailed materials texture. Consequently, balancing the need for extensive data against the practical constraints of data collection is a key obstacle in advancing the robustness of AD technologies.

Another formidable challenge in achieving robustness of autonomous vehicles is the black-box nature of deep learning-based models. These AI systems, central to AD, operate on complex neural networks that are intrinsically opaque, obscuring the decision-making process. This lack of transparency is particularly concerning in safety-critical scenarios, such as navigating intersections or responding to pedestrians. The hidden rationale behind its action within its neural networks raises concerns regarding accountability and public trust. To foster broader acceptance and integration of AD technology, it is essential to develop methods that clarify AI decision-making. This involves advancements in explainable AI and establishing standards for interpretability and reliability of autonomous vehicular decisions.

A further challenge in the development of AD systems is their limited awareness of social norms, crucial for seamless traffic integration. As participants in transportation ecosystems, autonomous vehicles must understand and obey unwritten social rules that human drivers intuitively navigate. While current AD systems focus on the technical aspects of navigation and safety, they often overlook the subtleties of human-like social interactions. For instance, these systems may struggle to interpret pedestrians' gestures or human drivers' nuanced behavior, such as yielding or negotiating right of way. The absence of social knowledge can lead to rigid, inappropriate responses in complex driving situations. To ensure smooth coexistence with human traffic participants, it is imperative to imbue AD systems with a deeper understanding of social cues. This entails not just technological advancements but also interdisciplinary research blending AI with social sciences to enhance the social intelligence of AD systems.

An additional AD challenge is the pressure from excessive publicity, which often leads to premature deployment and misuse. The industry's rush toward level 4 (L4) and L5 autonomy has sometimes resulted in overlooking critical safety measures. A poignant example is the 2018 fatal incident involving an Uber Technologies self-driving car, which emphasizes the risks of deploying autonomous technology without thorough readiness for real-world interactions. Such incidents erode public trust in autonomous technology and emphasize the need for a more cautious, measured approach to development and deployment. It is vital to balance innovation with safety, ensuring that the technological leaps do not outpace the comprehensive evaluation.

RECENT ADVANCES AND POTENTIAL SOLUTIONS

Addressing the challenge of data acquisition and utilization in AD, recent advances in digital twin technology and AI-generated content (AIGC) present promising solutions. Digital twins, as virtual replicas of driving environments, maintain virtual-real consistency and enable testing of rare or dangerous driving scenarios, enhancing AD systems and guiding method design with future traffic scenarios incorporating delayed reward strategies. The innovation in vehicle mechanics further bolsters AD algorithm development, making the algorithms more adaptable and responsive to real-world dynamics.

Simultaneously, high-fidelity simulation platforms for scenario reconstruction are crucial to reduce risks with the adaptation to rare scenarios and the material texture edition in 3D reconstructions. AIGC in particular offers rich scenario layouts and content, incorporating dynamic and adversarial objects in 3D reconstructed scenes, ensuring that the simulated environments closely mimic real-world complexities. This combination of digital twins and AIGC mitigates physical data collection limitations and propels the field toward more advanced, reliable, and socially integrated AD solutions.

To address the black box nature of AI models, a promising approach lies in leveraging control theory, especially its robust and nonlinear variants. It provides a structured approach to system design and fault diagnosis, which are critical in enhancing the interpretability and reliability of AI systems in AD.

By integrating knowledge rules and constraints, inspired by control theory principles, AI systems can gain a level of predictability and transparency crucial for safety-critical applications. Such integration not only aids in system-level understanding but also facilitates public trust and accountability. As the field of AD evolves, the synergy between control theory and AI could lead to more robust, interpretable, and transferable models, aligning technological advancement with safety and reliability standards essential for societal acceptance.

Enhancing AD system robustness requires a critical reassessment of guidance and navigation methodologies, ensuring reliable and adaptable vehicle control. By leveraging technologies such as high-definition mapping, real-time data processing, and sensor fusion, autonomous vehicles can achieve superior accuracy in positioning and route planning. This integration enhances the vehicle's ability to navigate complex environments and improves its responsiveness to dynamic changes and unexpected obstacles.

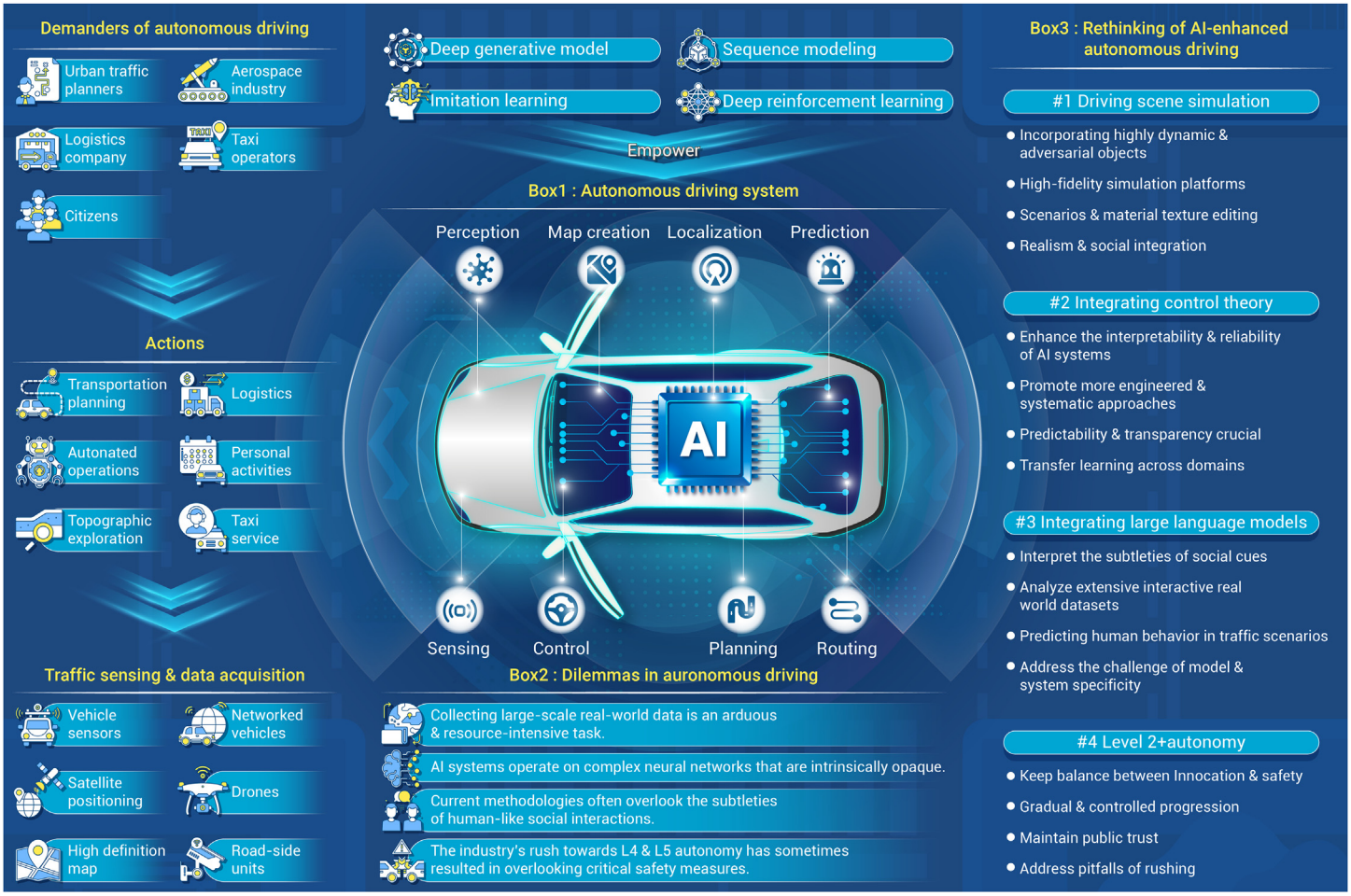

Figure 1. Toward the robustness of autonomous vehicles in the AI era

Addressing the social norms awareness gap in autonomous vehicles, it is an innovative approach to leverage large language models (LLMs) to enhance the vehicles' understanding of human social behaviors and cues.[5] LLMs, adept at processing and interpreting vast amounts of human language data, can provide valuable insights into nuanced human interactions. By integrating these models into AD systems, vehicles can better interpret the subtleties of social cues, such as pedestrians' gestures or human drivers' behavioral patterns.

This integration significantly improves the social awareness and responsiveness of autonomous vehicles. LLMs can analyze vast datasets encompassing real-world human interactions, enabling more accurate prediction of human behavior in traffic scenarios. especially in situations where understanding human intent is key. This approach can significantly enhance the social intelligence of autonomous vehicles, ensuring smoother coexistence with human traffic participants.

To address challenges of excessive publicity and premature deployment in AD, a strategic focus on L2+ autonomy emerges as a potential solution. Currently, most AD systems operate at L1 and L2, where AI assists rather than fully controls, demonstrating success and reliability. L2+ represents a significant advancement, where AI plays a more active role, enhancing the capabilities of the vehicle while still requiring human oversight.

The adoption of L2+ systems signifies a pivotal moment in AD industry. It strikes a balance between innovation and safety, offering advanced features like adaptive cruise control and lane-keeping assistance, This approach aligns with a more cautious and measured development pathway, providing ample opportunity for rigorous testing, refinement, and public acclimatization. This advancement is crucial for maintaining public trust and ensuring that safety remains at the forefront of autonomous vehicle development.

CONCLUSION

The journey toward the robustness of autonomous vehicles in the AI era is marked by complex challenges and innovative solutions. Addressing data acquisition, AI model opacity, social norms awareness, and rapid technological advancement pressure requires a multifaceted approach. Embracing digital twin technology, integrating control theory, leveraging LLMs, and focusing on L2+ autonomy represent significant strides toward achieving robust, reliable, and socially integrated autonomous vehicles. In the AI era, balancing innovation with safety and public trust remains paramount.

REFERENCES

1. Xu, Y., Liu, X., Cao, X. et al. (2021). Artificial intelligence: A powerful paradigm for scientific research. *Innovation* **2**:100179. DOI:https://doi.org/10.1016/j.xinn.2021.100179.
2. Goddard, M.A., Davies, Z.G., Guenat, S. et al. (2021). A global horizon scan of the future impacts of robotics and autonomous systems on urban ecosystems. *Nat. Ecol. Evol.* **5**:219–230. DOI:https://doi.org/10.1038/s41559-020-01358-z.
3. Almalioglu, Y., Turan, M., Trigoni, N. et al. (2022). Deep learning-based robust positioning for all-weather autonomous driving. *Nat. Mach. Intell.* **4**:749–760. DOI:https://doi.org/10.1038/s42256-022-00520-5.
4. Feng, S., Sun, H., Yan, X. et al. (2023). Dense reinforcement learning for safety validation of autonomous vehicles. *Nature* **615**:620–627. DOI:https://doi.org/10.1038/s41586-023-05732-2.
5. Xu, Z., Zhang, Y., Xie, E. et al. (2023). DriveGPT4: Interpretable end-to-end autonomous driving via large language model. Preprint at *arXiv*. DOI:https://doi.org/10.48550/arXiv.2310.01412.

ACKNOWLEDGMENTS

This work was supported by NSFC 62372430, U22B2058, 62173034, and 62171276; Youth Innovation Promotion Association CAS 2023112; and the Science and Technology Commission of Shanghai Municipality under grant 21511100900. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.