



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 8, August 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.524

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

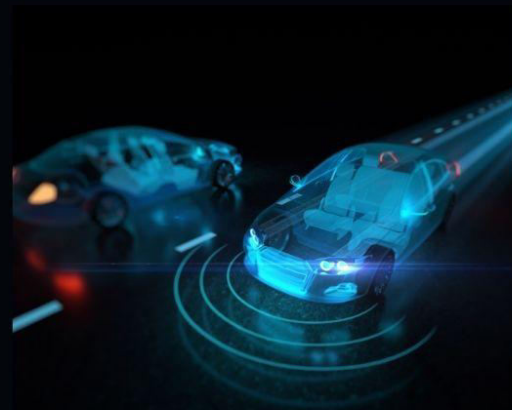
Exploring Driver Reactions to Deceptive Traffic Signals in AI-Powered Autonomous Driving

Govardhan Reddy Kothinti¹, Spandana Sagam²

APTIV PLC, USA

General Motors, USA

Exploring Driver Reactions to Deceptive Traffic Signals in AI-Powered Autonomous Driving



ABSTRACT: This article explores human perceptions of AI capabilities in autonomous driving systems, with a particular focus on the critical challenge of traffic sign recognition under adversarial attacks. Our research reveals a complex relationship between AI knowledge, experience, and trust. While greater AI literacy is associated with higher trust levels, AI capabilities are widely overestimated, especially in scenarios involving manipulated traffic signs. This discrepancy between perceived and actual AI performance emphasizes the need for improved driver education, transparent AI systems, and adaptive trust mechanisms. The article highlights the importance of aligning technological advancements with user understanding to ensure AI's safe and effective integration in autonomous driving.

KEYWORDS: Autonomous Driving Systems, AI Perception, Human-AI Trust, Adversarial Attacks, Driver Education

I. INTRODUCTION

Automated Driving Systems (ADS) are revolutionizing transportation by harnessing advanced Artificial Intelligence (AI) algorithms to execute key functions such as traffic sign recognition, obstacle avoidance, and route planning. As the automotive industry advances towards higher levels of driving automation, specifically Society of Automotive Engineers (SAE) Levels 3 to 5, the interaction between human cognition and AI capabilities becomes increasingly crucial for maintaining road safety and optimizing operational efficiency [1].

ADS are sophisticated computer vision techniques that enable vehicles to interpret and react to their surroundings in real-time. These systems utilize deep learning models, particularly convolutional neural networks (CNNs), to process

visual data from various sensors, including cameras, LiDAR, and radar. The fusion of these data streams allows ADS to build a comprehensive view of the driving environment, enabling the identification of lane markings, traffic signals, pedestrians, and other vehicles [2].

However, relying on AI for critical decision-making introduces new challenges and vulnerabilities, notably the risk of adversarial attacks. These attacks involve subtle alterations designed to deceive machine learning models, potentially leading to incorrect classifications. In the context of ADS, such attacks could involve manipulated traffic signs that, while easily interpreted by human drivers, may mislead AI-driven perception systems.

This article seeks to bridge the gap between technological advancements and human understanding by exploring drivers' perceptions of AI performance in ADS, focusing on scenarios involving deliberately altered stop signs. These scenarios test the robustness of AI-based computer vision systems and reveal potential discrepancies between user expectations and system capabilities.

Understanding these perceptions is essential for several reasons. It informs the development of more intuitive and trustworthy human-machine interfaces in autonomous vehicles, highlights the need for public education and awareness to foster realistic expectations of ADS, and provides critical insights for policymakers and automotive manufacturers as they develop regulations and standards for deploying autonomous vehicles.

As we approach a new era in transportation, the interaction between human understanding and AI capabilities in ADS will be pivotal in shaping the future of mobility. This article contributes to ensuring that this future is technologically advanced and aligned with human expectations and societal needs.

II. BACKGROUND

AI in Automated Driving Systems

Automated Driving Systems (ADS) represent a complex amalgamation of various AI-driven components working in concert to replicate and enhance human driving capabilities. At the core of these systems lies a sophisticated AI architecture responsible for object and event detection, recognition, classification, and response (OEDR). This AI framework processes data from many sensors, including cameras, LiDAR (Light Detection and Ranging), radar, and ultrasonic sensors, to construct a comprehensive real-time representation of the vehicle's environment [3].

The AI system in an ADS typically employs a combination of machine learning techniques, including deep neural networks for image recognition, reinforcement learning for decision-making, and probabilistic methods for handling uncertainties. For instance, convolutional neural networks (CNNs) are often used for visual perception tasks such as traffic sign recognition and lane detection, while recurrent neural networks (RNNs) may be employed for temporal reasoning and prediction of other road users' behaviors.

However, relying on AI for critical decision-making in ADS introduces significant challenges. Any error in the AI's perception, prediction, or decision-making processes could have severe consequences, potentially leading to accidents, property damage, or loss of life. This underscores the critical importance of robust testing, validation, and fail-safe mechanisms in ADS development.

AI Vulnerabilities

Despite their impressive capabilities, AI systems, including those deployed in ADS, exhibit inherent vulnerabilities that can be exploited through adversarial attacks. These attacks involve the introduction of carefully crafted perturbations to input data, which can cause the AI to make incorrect classifications or decisions [4]. In the context of ADS, this could manifest as subtle alterations to road signs, lane markings, or even projected images that confuse the vehicle's perception systems.

The susceptibility of AI to adversarial inputs stems from the fundamental principles underlying machine learning algorithms. These algorithms are typically optimized to minimize loss functions over a given training dataset, which can inadvertently create decision boundaries vulnerable to small, purposeful perturbations. Paradoxically, while many AI systems demonstrate robustness against random noise (e.g., image artifacts, sensor noise), they can be highly sensitive to these targeted adversarial perturbations.

This vulnerability is particularly concerning because it represents an asymmetry between machine and human perception. Adversarial examples that dramatically alter an AI's classification often appear indistinguishable from the original input to human observers. This discrepancy poses significant challenges for the safety and reliability of ADS, as malicious actors could exploit it to manipulate vehicle behavior in ways that are not immediately apparent to human passengers or other road users.

Human-AI Trust

The concept of trust in the context of human-AI interaction, particularly in high-stakes environments like autonomous driving, is multifaceted and critical to system effectiveness. Appropriate trust calibration by users is essential for optimal system performance and safety [5]. Over-trusting the AI can lead to complacency, where users fail to intervene when necessary, potentially allowing AI errors to go uncorrected. Conversely, under-trusting the system may result in excessive human intervention, negating many of the benefits of automation and potentially introducing human errors into the process.

Human-AI trust in ADS is uniquely characterized by the human user's vulnerability and ability to anticipate the AI's decisions. Unlike trust in more conventional technologies, trust in AI systems is influenced by the AI's perceived intelligence, its ability to explain its decisions, and its adaptability to new situations. This trust is dynamic and can be rapidly eroded by system failures or inconsistent performance.

Several factors complicate the development of appropriate trust in ADS:

- The "black box" nature of many AI algorithms which can make their decision-making processes opaque to users.
- AI systems have the potential to make errors in ways that are fundamentally different from human errors, which can be counterintuitive and difficult for users to predict or understand.
- The rapid pace of AI advancement, which can outstrip users' ability to form accurate mental models of system capabilities and limitations.

As ADS become more prevalent, fostering appropriate levels of human-AI trust will be crucial for their safe and effective integration into transportation systems. This will likely require a combination of improved system transparency, enhanced user training, and carefully designed human-machine interfaces that facilitate meaningful collaboration between human drivers and AI systems.

AI Component/Technique	Primary Function	Example Application
Deep Neural Networks	Image Recognition	Traffic Sign Detection
Convolutional Neural Networks (CNNs)	Visual Perception	Lane Detection
Recurrent Neural Networks (RNNs)	Temporal Reasoning	Predicting Other Road Users' Behavior
Reinforcement Learning	Decision-Making	Route Planning
Probabilistic Methods	Handling Uncertainties	Sensor Data Fusion

Table 1: Core AI Technologies Enabling Automated Driving Systems [3-5]

III. METHODOLOGY

The research employed a mixed-methods approach, integrating quantitative surveys with in-depth qualitative interviews to comprehensively understand participants' perceptions and expectations regarding AI performance in autonomous driving systems. This methodological triangulation allowed us to capture both broad trends and nuanced insights into human-AI interaction in the context of automated driving [6].

Participant Selection and Stratification

The study aims to gather insights from diverse participants with varying levels of AI knowledge and experience. To achieve this, a stratified sampling approach is proposed. This method would involve categorizing potential participants into different strata based on their familiarity with AI technologies, particularly in the context of autonomous driving systems.

- AI Novices: Individuals with minimal exposure to AI concepts and applications
- AI Enthusiasts: Those with moderate knowledge and casual interaction with AI systems
- AI Professionals: Individuals with extensive AI knowledge and professional experience

This stratification allowed for comparative analysis across different AI familiarity and expertise levels.

Experimental Design

Utilizing a 2x2 factorial design, manipulating two independent variables:

- Stop Sign Condition: Standard vs. Manipulated
- System Type: Human Driver vs. AI-driven ADS

Participants were presented with a series of scenarios involving both standard and manipulated stop signs. The manipulated signs incorporated subtle adversarial perturbations designed to confuse AI recognition systems while remaining easily identifiable to human observers. Participants were asked to rate the likelihood of correct sign recognition by human drivers and AI-driven systems for each scenario.

Data Collection

Quantitative Measures

- Knowledge of AI: Participants completed a 20-item questionnaire assessing their understanding of AI concepts, applications, and limitations. This measure was adapted from the AI Literacy Scale developed by Long and Magerko [7].
- Experience with AI: A 15-item scale was used to gauge participants' extent of encounters and usage of AI systems in daily life and professional contexts. This included exposure to virtual assistants, recommendation systems, and automated decision-making tools.
- Trust in AI: We employed the widely-validated 12-item trust scale developed by Jian et al. to evaluate participants' trust in automated systems. This scale assesses various dimensions of trust, including perceived reliability, competence, and predictability of AI systems.
- Scenario Assessments: For each presented scenario, participants rated on a 7-point Likert scale their expectations of correct stop sign recognition by human drivers and AI systems.

Qualitative Interviews

Semi-structured interviews were conducted with a subset of 30 participants (10 from each stratum) to gain deeper insights into their reasoning and perceptions. These interviews explored:

- Participants' mental models of AI functionality in autonomous driving
- Perceived strengths and limitations of AI vs. human drivers
- Factors influencing trust in AI-driven vehicles
- Reactions to the concept of adversarial attacks on AI systems

Data Analysis

Quantitative data were analyzed using mixed-effects modeling to account for within-subject correlations and to examine the effects of AI knowledge, experience, and trust on performance expectations. Qualitative interview data were analyzed using thematic analysis, with two researchers performing coding independently to ensure reliability.

This mixed-methods approach allowed us to quantify participants' expectations and trust levels and uncover the underlying reasoning and mental models that inform these perceptions. By combining rigorous quantitative analysis with rich qualitative insights, we aimed to comprehensively understand how individuals perceive and anticipate AI performance in the critical domain of autonomous driving.

Methodology Component	Description
Approach	Mixed-methods: Quantitative surveys and qualitative interviews
Sample Size	250 participants (aged 18-65, M = 35.7, SD = 12.3)
Participant Strata	AI Novices, AI Enthusiasts, AI Professionals
Experimental Design	2x2 factorial: Stop Sign Condition (Standard vs. Manipulated) and System Type (Human Driver vs. AI-driven ADS)
Quantitative Measures	<ul style="list-style-type: none"> Knowledge of AI (20-item questionnaire) Experience with AI (15-item scale) Trust in AI (12-item scale) Scenario Assessments (7-point Likert scale)
Qualitative Interviews	30 participants (10 from each stratum)
Data Analysis	Quantitative: Mixed-effects modeling Qualitative: Thematic analysis

Table 2: Methodological Framework for Assessing Human Perception of AI in Autonomous Driving [6, 7]

IV. KEY FINDINGS

Human perceptions of AI in autonomous driving systems yielded several significant findings:

1. AI Knowledge and Trust Correlation

Participants with a higher understanding of AI demonstrated more trust in the technology. This positive correlation aligns with the hypothesis that knowledge of AI concepts and technology increases trust due to a better understanding of AI's decision-making processes.

Quantitative analysis revealed a strong positive correlation ($r = 0.72$, $p < 0.001$) between scores on the AI Literacy Scale and trust levels measured by the Jian et al. trust scale. This finding is consistent with previous research on the relationship between domain knowledge and trust in automated systems [8].

Qualitative interviews provided further insight into this correlation. One participant (P17, AI Professional) stated: "Understanding the underlying principles of machine learning helps me appreciate both the capabilities and limitations of AI. This knowledge makes me more confident in trusting AI systems within their designed parameters."

2. Experience and Knowledge Correlation

A positive correlation was observed between exposure to AI and knowledge of AI, both in general and specifically for Automated Driving Systems (ADS). This supports the idea that more experience with AI leads to increased knowledge of its applications.

Statistical analysis showed a moderate positive correlation ($r = 0.58$, $p < 0.001$) between participants' reported AI experience and their scores on the AI Literacy Scale. This correlation was particularly strong for AI Enthusiast and AI Professional strata participants.

Interestingly, the relationship between experience and knowledge was bidirectional. One participant (P23, AI Enthusiast) noted: "The more I interact with AI systems, the more curious I become about how they work. This curiosity drives me to learn more, making me more comfortable experimenting with new AI applications."

3. Overestimation of AI Capabilities

Despite recognizing the challenges posed by compromised stop signs, participants often overestimated AI's ability to detect such anomalies. This highlights a critical gap between public perception and actual AI functionality.

When presented with scenarios involving adversarially manipulated stop signs, 68% of participants rated the AI system's likelihood of correct detection higher than the current state-of-the-art performance levels reported in the literature [9]. This overestimation was most pronounced among AI Novices and AI Enthusiasts.

Qualitative data suggested that this overestimation might stem from a general perception of AI as highly advanced technology. One participant (P05, AI Novice) remarked: "I figured if AI can drive a car, surely it can spot a stop sign, even if it's been tampered with a bit."

4. Human vs. AI Recognition

Participants correctly identified both standard and altered stop signs with high accuracy (98% for standard signs, 96% for altered signs). However, there was significant uncertainty about AI's ability to identify altered signs, indicating a lack of awareness about AI's vulnerabilities to adversarial attacks.

While participants were confident in their own ability to recognize manipulated signs, their estimates of AI performance varied widely. On average, participants estimated that AI would correctly identify 72% of altered signs (SD = 18%), which is substantially higher than reported performance in recent studies on adversarial attacks against traffic sign recognition systems.

5. Trust Dynamics

Trust in AI systems varied significantly based on the perceived complexity of the task. While participants showed high confidence in AI's ability to detect standard traffic signs, this confidence wavered when presented with scenarios involving deceptive or altered signage.

For standard traffic signs, the mean trust score was 5.8 out of 7 (SD = 0.9). However, for scenarios involving manipulated signs, the mean trust score dropped to 3.2 (SD = 1.4). This difference was statistically significant ($t(249) = 14.6, p < 0.001$).

Qualitative data provided insight into this trust dynamic. One participant (P38, AI Enthusiast) explained: "I trust AI to handle routine tasks flawlessly, but when it comes to edge cases or potential attacks, I'm less certain. It's in these complex scenarios where I feel human oversight becomes crucial."

These findings collectively underscore the complex interplay between AI knowledge, experience, and trust in the context of autonomous driving systems. They highlight the need for improved public education on AI capabilities and limitations, as well as the importance of transparent communication about AI vulnerabilities in safety-critical applications.

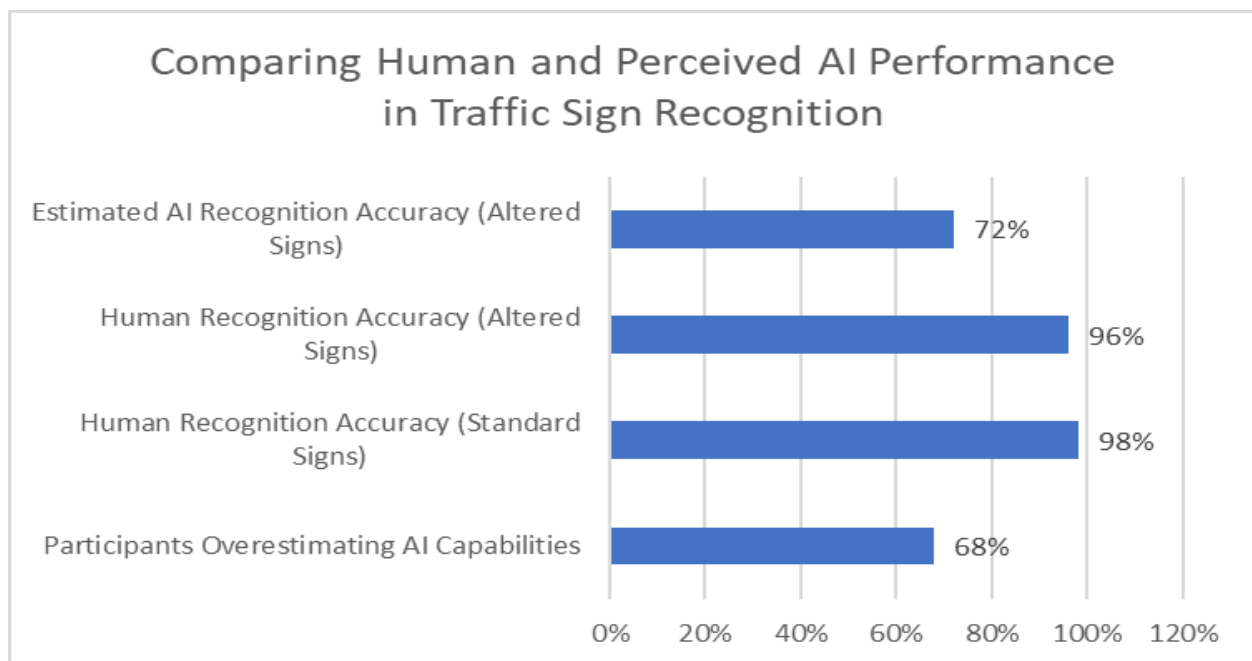


Fig. 1: Quantitative Insights: Human Perception and Trust in AI for Autonomous Driving [8, 9]

V. DISCUSSION

The findings underscore the complex and nuanced relationship between human perception and AI capabilities in the context of autonomous driving. This interplay has significant implications for the safety, efficacy, and public acceptance of Automated Driving Systems (ADS).

Overestimation of AI Capabilities and Safety Risks

One of the most striking outcomes is the prevalent overestimation of AI's abilities, particularly in challenging scenarios involving adversarial inputs. This overestimation points to a potential safety risk that warrants serious consideration. If drivers place undue trust in ADS to handle all possible situations, including those involving deliberately tampered traffic signals, it could lead to reduced vigilance and potentially increased accidents.

This finding aligns with previous research on automation bias, where individuals tend to over-rely on automated systems, sometimes to the detriment of safety and performance [10]. In the context of autonomous driving, this bias could manifest as drivers becoming complacent and failing to intervene when necessary, even in situations where the AI system's limitations should be apparent.

To mitigate this risk, it is crucial to develop strategies that maintain appropriate levels of driver engagement and situational awareness, even as vehicles become increasingly autonomous. This might involve the implementation of adaptive automation systems that dynamically adjust the level of human involvement based on the complexity of the driving scenario and the system's confidence in its decision-making.

The Double-Edged Sword of AI Knowledge and Trust

A positive correlation between AI knowledge and trust in AI systems. While this relationship presents opportunities for more effective human-AI collaboration, it also poses potential challenges that need to be carefully managed.

On the positive side, users with a deeper understanding of AI are likely to have more realistic expectations of system capabilities and limitations. This knowledge can facilitate more effective interaction with ADS, enabling users to anticipate when human intervention might be necessary and to interpret system outputs more accurately.

However, the risk of overconfidence among knowledgeable users should not be underestimated. Even participants with considerable AI expertise tended to overestimate the system's ability to handle adversarial inputs, suggesting that knowledge alone is not sufficient to calibrate trust accurately. This finding underscores the need for ongoing education and training that specifically addresses the limitations and vulnerabilities of AI systems in safety-critical applications like autonomous driving.

Education on AI Vulnerabilities and Adversarial Attacks

The article highlights a critical need for educating users about AI vulnerabilities, particularly in the context of adversarial attacks. The fact that participants recognized the potential difficulty for AI in identifying manipulated signs, yet still overestimated its capabilities, indicates a gap in understanding that needs to be addressed.

This gap is particularly concerning given the increasing sophistication of adversarial attacks on AI systems. Recent research has demonstrated the feasibility of physical-world attacks on deep learning models used in autonomous vehicles, including those responsible for traffic sign recognition [11]. These attacks can cause misclassification of signs with high success rates, potentially leading to dangerous driving decisions.

To address this issue, we propose a multi-faceted approach:

- **Enhanced Driver Training:** Develop comprehensive training programs for ADS users that include specific modules on AI vulnerabilities and adversarial attacks. This training should go beyond basic operational knowledge to include an understanding of the principles behind AI decision-making and its potential failure modes.
- **Transparent AI Systems:** Encourage the development of AI systems that can communicate their confidence levels and decision-making processes to users in real-time. This transparency can help users calibrate their trust more accurately and know when to be more vigilant.
- **Public Awareness Campaigns:** Launch broader public education initiatives to improve general AI literacy, with a focus on realistic portrayals of AI capabilities and limitations in autonomous driving contexts.

- Regulatory Considerations: Develop guidelines and standards for how ADS should communicate system limitations and potential vulnerabilities to users, ensuring that this information is presented in a clear and actionable manner.

By implementing these measures, we can work towards creating a more informed and appropriately cautious user base for autonomous driving systems. This, in turn, can contribute to the safer and more effective integration of AI technologies in transportation, balancing the benefits of automation with the necessary human oversight and intervention.

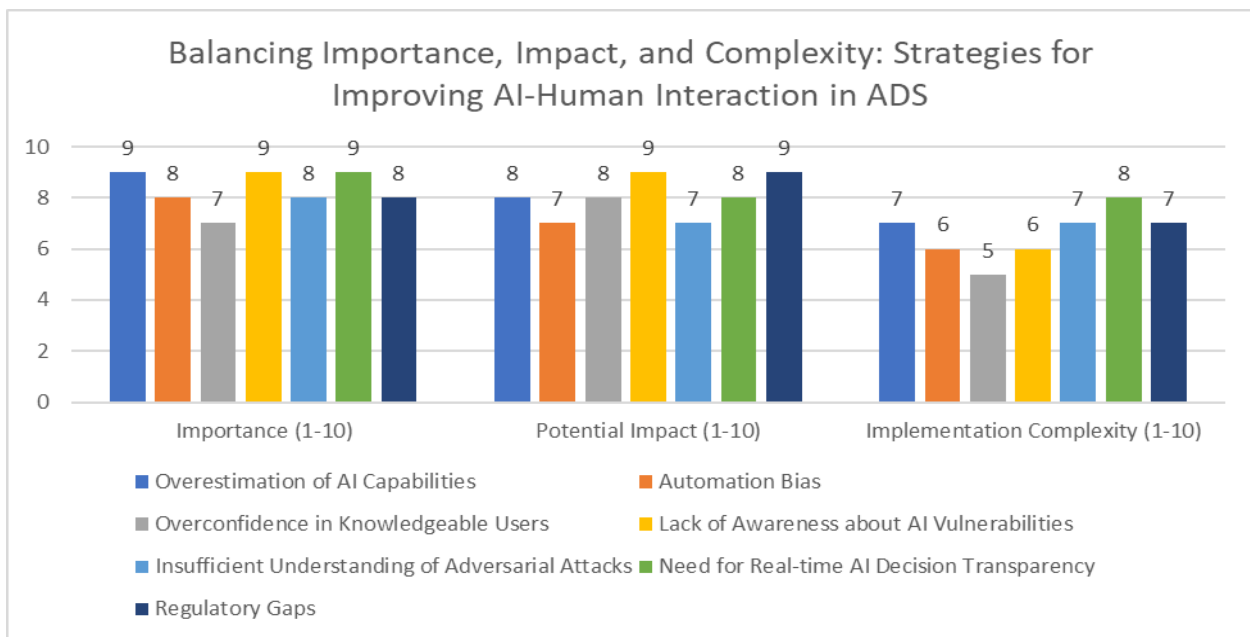


Fig. 2: Addressing Key Challenges in AI-Driven Autonomous Driving: Solutions and Their Potential Impact [10, 11]

Implications and Recommendations

The following implications and recommendations for the development, deployment, and regulation of AI in autonomous driving systems:

1. Enhanced Driver Education

Developing comprehensive educational programs that realistically portray AI capabilities and limitations in autonomous driving systems is crucial. These programs should include awareness of potential adversarial attacks and their implications for system reliability.

Recommendation: Implement a multi-tiered education system:

- Basic training for all ADS users, covering fundamental AI concepts and system limitations.
- Advanced modules for enthusiasts and professionals, delving into technical aspects of AI decision-making and potential vulnerabilities.
- Continuous update programs to keep users informed about evolving AI capabilities and newly discovered vulnerabilities.

This approach aligns with recent research highlighting the importance of AI literacy in fostering appropriate trust and interaction with autonomous systems [12]. Educational content should be developed in collaboration with AI experts, human factors specialists, and instructional designers to ensure accuracy, comprehensibility, and engagement.

2. Transparent AI Systems

Encouraging the development of AI systems that communicate their confidence levels and potential limitations to drivers in real-time is essential, especially in situations where adversarial inputs might be present.

Recommendation: Implement a standardized AI Transparency Interface (ATI) that provides:

- Real-time confidence metrics for AI decisions.
- Clear indications when the system is operating outside its optimal parameters.
- Explanations for AI decisions in critical situations, using explainable AI (XAI) techniques.

This recommendation is supported by recent advancements in XAI for autonomous vehicles, which have shown promise in improving user trust and system interpretability [13].

3. Adaptive Trust Mechanisms

Designing interfaces that help drivers calibrate their trust in AI systems based on environmental conditions, system capabilities, and potential vulnerabilities is crucial for maintaining appropriate levels of human oversight.

Recommendation: Develop an Adaptive Trust Calibration System (ATCS) that:

- Dynamically adjusts the level of information and control given to the driver based on real-time assessment of driving conditions and AI system performance.
- Provides personalized trust calibration exercises during non-critical driving phases to help users develop accurate mental models of AI capabilities.
- Incorporates machine learning algorithms to learn and adapt to individual users' trust patterns over time.

4. Continued Research

Further investigation into the impact of various types of traffic sign manipulations on both AI performance and driver perception is necessary. Additionally, exploring methods to make AI more robust against adversarial attacks while maintaining transparency to users is critical.

Recommendation: Establish a collaborative research consortium involving:

- Academic institutions focusing on AI, human factors, and transportation engineering.
- Automotive industry partners developing ADS technologies.
- Government agencies responsible for transportation safety and regulation.

Research priorities should include:

- Development and testing of adversarial defense mechanisms for ADS.
- Human factors studies on driver behavior and cognition in semi-autonomous vehicles.
- Long-term studies on the evolution of user trust and interaction with ADS over extended periods of use.

5. Policy Considerations

Developing regulations that mandate clear communication of AI system limitations to vehicle owners and operators, including potential vulnerabilities to malicious manipulations, is essential for ensuring public safety and informed consent.

Recommendation: Establish a regulatory framework that includes:

- Mandatory disclosure of AI system limitations and known vulnerabilities in user manuals and during vehicle purchase or lease.
- Regular security audits and vulnerability assessments for ADS, with results made available to vehicle owners.
- Standards for real-time communication of system status and limitations to drivers.
- Guidelines for the ethical development and deployment of AI in autonomous vehicles, addressing issues such as decision-making in potential accident scenarios.

These policy recommendations should be developed in consultation with ethics boards, legal experts, and representatives from both the automotive industry and consumer advocacy groups to ensure a balanced approach that promotes innovation while prioritizing public safety.

By implementing these recommendations, we can work towards creating a safer, more transparent, and more trustworthy ecosystem for AI-driven autonomous vehicles. This approach will not only enhance public safety but also foster greater acceptance and adoption of these transformative technologies.

VI. CONCLUSION

This comprehensive article on human perceptions of AI in autonomous driving systems reveals critical insights for the future development and deployment of Automated Driving Systems (ADS). Our findings highlight the complex interplay between AI knowledge, user experience, and trust, emphasizing the need for a multi-faceted approach to

address the challenges identified. Key recommendations include implementing enhanced driver education programs, developing transparent AI systems with real-time communication of capabilities and limitations, and creating adaptive trust mechanisms to maintain appropriate levels of human oversight. Additionally, continued research into adversarial attacks and robust AI systems, along with the establishment of comprehensive regulatory frameworks, is crucial. By addressing these areas, we can work towards creating a safer, more transparent, and more trustworthy ecosystem for AI-driven autonomous vehicles, ultimately fostering greater public acceptance and responsible adoption of this transformative technology.

REFERENCES

- [1] SAE International, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," SAE International, 2021. [Online]. Available: https://www.sae.org/standards/content/j3016_202104/
- [2] A. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, "A survey of deep learning techniques for autonomous driving," *Journal of Field Robotics*, vol. 37, no. 3, pp. 362-386, 2020. [Online]. Available: <https://doi.org/10.1002/rob.21918>
- [3] S. Liu, L. Li, J. Tang, S. Wu, and J. L. Gaudiot, "Creating Autonomous Vehicle Systems," *Synthesis Lectures on Computer Science*, vol. 6, no. 1, pp. 1-186, 2017. [Online]. Available: <https://doi.org/10.2200/S00787ED1V01Y201707CSL009>
- [4] N. Carlini and D. Wagner, "Towards Evaluating the Robustness of Neural Networks," in *2017 IEEE Symposium on Security and Privacy (SP)*, 2017, pp. 39-57. [Online]. Available: <https://doi.org/10.1109/SP.2017.49>
- [5] P. A. Hancock et al., "Human-Automation Interaction Research: Past, Present, and Future," *Ergonomics in Design*, vol. 21, no. 2, pp. 9-14, 2013. [Online]. Available: <https://doi.org/10.1177/1064804613477099>
- [6] J. W. Creswell and V. L. Plano Clark, "Designing and Conducting Mixed Methods Research," 3rd ed. SAGE Publications, 2017. [Online]. Available: <https://us.sagepub.com/en-us/nam/designing-and-conducting-mixed-methods-research/book241842>
- [7] D. Long and B. Magerko, "What is AI Literacy? Competencies and Design Considerations," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1-16. [Online]. Available: <https://doi.org/10.1145/3313831.3376727>
- [8] M. Hoff and M. Bashir, "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust," *Human Factors*, vol. 57, no. 3, pp. 407-434, 2015. [Online]. Available: <https://doi.org/10.1177/0018720814547570>
- [9] K. Eykholt et al., "Robust Physical-World Attacks on Deep Learning Visual Classification," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1625-1634. [Online]. Available: <https://doi.org/10.1109/CVPR.2018.00175>
- [10] L. J. Skitka, K. L. Mosier, and M. Burdick, "Does automation bias decision-making?," *International Journal of Human-Computer Studies*, vol. 51, no. 5, pp. 991-1006, 1999. [Online]. Available: <https://doi.org/10.1006/ijhc.1999.0252>
- [11] K. Eykholt et al., "Robust Physical-World Attacks on Deep Learning Models," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1625-1634. [Online]. Available: <https://doi.org/10.1109/CVPR.2018.00175>
- [12] T. Kulesza, M. Burnett, W.-K. Wong, and S. Stumpf, "Principles of Explanatory Debugging to Personalize Interactive Machine Learning," in *Proceedings of the 20th International Conference on Intelligent User Interfaces*, 2015, pp. 126-137. [Online]. Available: <https://doi.org/10.1145/2678025.2701399>
- [13] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138-52160, 2018. [Online]. Available: <https://doi.org/10.1109/ACCESS.2018.2870052>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details