Risk Analysis of Data Misuse in Autonomous Robot Research: A Case Study of Violations of the Principles of Transparency and Algorithm Accountability

Ahmad Sahru Romadhon*, Hakkun Elmunsyah, Anik Nur Handayani

Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia

***Corresponding Author**

DOI: https://doi.org/10.51244/IJRSI.2025.12110200

Received: 09 December 2025; Accepted: 17 December 2025; Published: 25 December 2025

ABSTRACT

The development of autonomous robots in the era of Industry 4.0 presents enormous opportunities as well as significant ethical risks, particularly in relation to data collection and use. The complexity of artificial intelligence (AI)-based algorithms has led to the emergence of the black box phenomenon, where decisionmaking processes are difficult to explain and verify. This condition creates violations of two key ethical principles, namely transparency and accountability, which have the potential to increase the risk of data misuse throughout the entire life cycle of autonomous robot research. This study systematically analyzes the relationship between violations of these principles and various threat scenarios such as algorithmic discrimination, invasive profiling, and sensor data manipulation. Using a qualitative case study approach and a four-phase risk analysis method, the study identifies two critical risks: decision discrimination due to black box models, and sensor data manipulation due to weak accountability and audit mechanisms. The results confirm that a lack of transparency hinders the detection of data bias, while weak accountability opens the door to third-party intervention. This study recommends the implementation of Explainable AI (XAI), training data audits, tamper-proof audit log systems, and rollback mechanisms as key mitigation measures to improve the security, reliability, and ethics of data use in autonomous robot research.

Keyword - Autonomous robots, Data misuse, Artificial intelligence, Technology ethics

INTRODUCTION

The era of the 4.0 industrial revolution has positioned autonomous robots as one of the transformative technologies with promising applications in various sectors, ranging from logistics and manufacturing to healthcare and household [1]. The core capabilities of these robots lie in artificial intelligence (AI) and machine learning, which enable them to understand complex environments, learn from data, and make decisions independently with little or no human intervention [2]. Intensive research and development (R&D) has been the main driver of this rapid progress, resulting in increasingly sophisticated systems. However, the algorithmic complexity underlying this progress has created a new paradox: the higher the performance of the system, the more difficult it often is to understand the reasoning process behind its decisions [3].

This complexity gives rise to the phenomenon of "black box algorithms," in which the inputs and outputs of a system can be observed, but its internal processes cannot be easily explained, even by its designers [4]. This condition directly undermines two crucial pillars of ethics in responsible technology development: transparency and accountability [5]. In the context of autonomous robot research, violations of the principle of transparency are reflected in the lack of adequate documentation regarding datasets, model architecture, and algorithm limitations [5]. Meanwhile, accountability failures arise when there are no clear mechanisms for assigning responsibility if algorithms fail or cause unintended consequences, creating what is known as an "accountability gap" [6].

Most literature on AI ethics has discussed separately both the importance of transparency and accountability [7] and threats to data security [8]. However, there is an analytical gap in connecting the intersection between direct

violations of these ethical principles and increased risks of data misuse in the context of autonomous robot research. In fact, an opaque and non-accountable research environment creates ideal conditions for data misuse, whether intentional or unintentional [9]. For example, the absence of an audit trail on training data can hide bias, while algorithmic opacity can conceal vulnerabilities that can be exploited to manipulate robot behavior [8].

This study investigates the systematic correlation between violations of transparency and accountability principles and patterns of data misuse in the autonomous robot research life cycle. The research questions raised are: How are violations of transparency and accountability manifested technically in various phases of autonomous robot development?, What risks of data misuse are causally related to each form of violation?.

This paper is organized as follows. After the introduction, we present a literature review discussing the theoretical foundations of algorithmic ethics, the autonomous robot research cycle, and the spectrum of data misuse. The methodology section explains the qualitative case study approach used. Findings and analysis are then presented to identify patterns of ethical violations and map their risks. The paper concludes with a discussion of theoretical and practical implications, as well as conclusions that include mitigation recommendations.

RESEARCH METHOD

Contemporary autonomous robots operate through a perception-planning-action cycle supported by deep learning algorithms [10]. These systems consume massive volumes of heterogeneous data, including LiDAR sensor data, visual inputs, and contextual environmental data [11]. The entire robot lifecycle, from training and validation to operation, forms a data value chain that is vulnerable to deviation. Research [12] shows that 73% of modern robotic systems rely on closed data pipelines that hinder auditability.

Algorithmic transparency in robotics is evolving beyond mere source code access toward model interpretability (Explainable AI/XAI). Study [13] demonstrates how Layer-wise Relevance Propagation (LRP) techniques can reveal the basis for autonomous robot navigation decisions. However, the implementation of XAI in research is still limited, with only 34% of robotics papers reporting the interpretation methods used [14]. This condition creates an opaque research environment where algorithmic decisions are difficult to verify.

Accountability in autonomous systems requires a clear framework for assigning responsibility when failures occur. Research [15] proposes a “Liability Tracing Framework” model that maps error contributors across the entire technical stack. Findings [16] reveal that 68% of industrial robotics incidents experience difficulties in tracing accountability due to undefined distribution of responsibility. Study [6] further identifies three levels of accountability: algorithm designers, system integrators, and operators.

Robotics cybersecurity research has cataloged various data attack vectors. Research [17] identifies adversarial attacks that exploit vulnerabilities in machine learning models through structured noise injection. Meanwhile, [18] demonstrates how sensor data manipulation can cause catastrophic failures in autonomous navigation systems. Recent findings [19] show a 156% increase in data poisoning attacks on robotics training datasets during the 2022-2024 period.

Although studies on AI ethics and robotics security have developed, research that empirically links transparency-accountability deficits with data misuse vulnerabilities remains limited. Study [20] identified only 12% of papers that discussed the dimension of data security in the context of algorithmic transparency. This analytical gap is significant given the increasing integration of autonomous robot in critical infrastructure, where accountability failures could lead to massive data breaches [21].

The research method flow can be seen in Figure 1. Method of analyzing the risk of data misuse in autonomous robot research.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

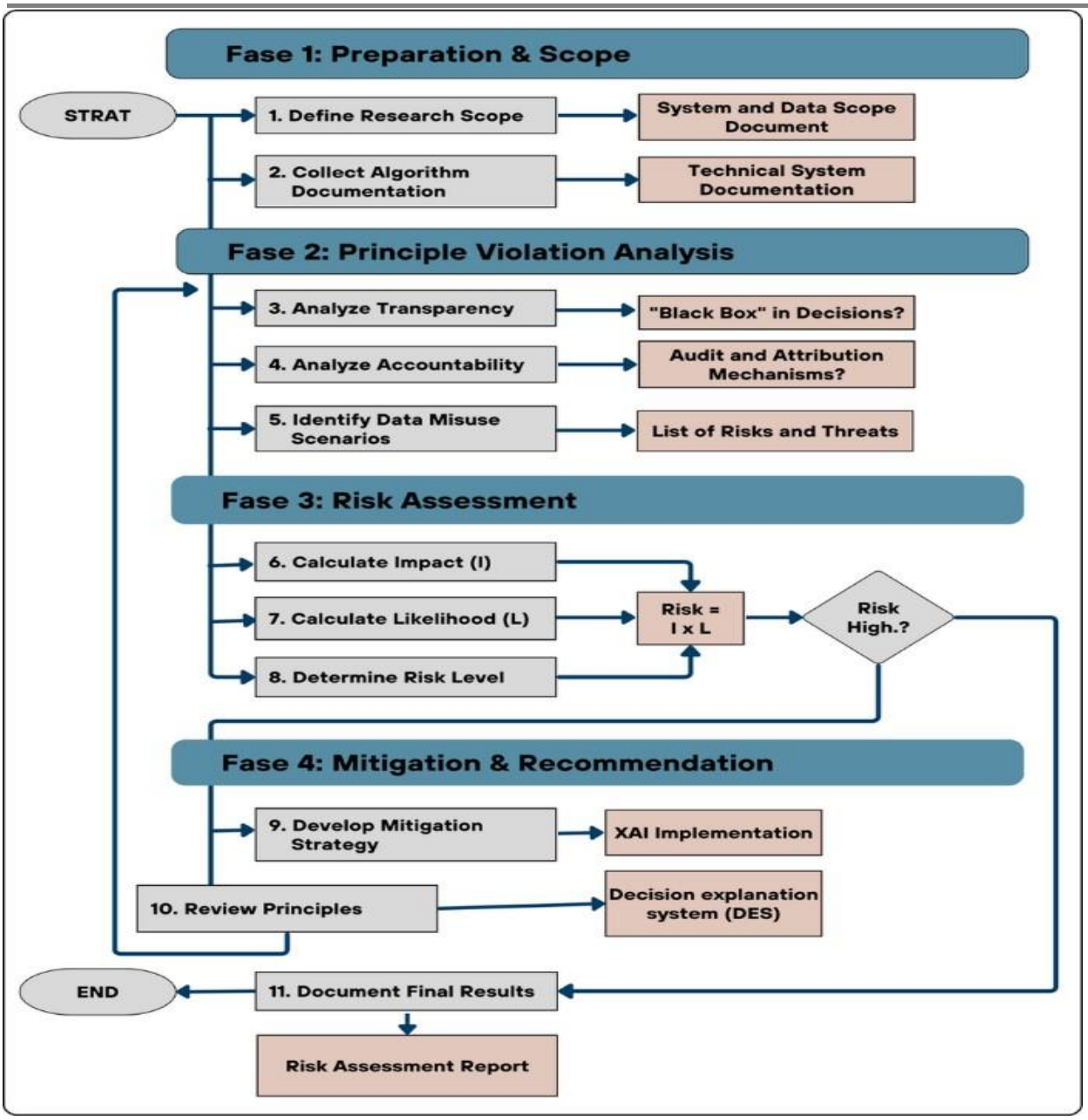

Figure 1. Risk analysis flowchart for data misuse in autonomous robot research.

This risk analysis method is divided into four main sequential phases, starting from scope definition to documentation of recommendations. Phase 1 (Preparation and Scope) begins the process by clearly defining the scope of the autonomous robot system research and the type of data used, followed by the collection of technical documentation on the algorithm. Once the scope has been defined, the process continues to Phase 2 (Principle Violation Analysis), where the main focus is to assess the extent to which the principles of Transparency (the ability to explain algorithmic decisions or Explainability) and Accountability (the mechanism for attributing responsibility) have been adhered to. If violations are found, then the Identification of Data Misuse Scenarios arising from these transparency and accountability gaps is carried out, resulting in a List of Risks and Threats that must be addressed immediately.

The identified threats are then taken to Phase 3 (Risk Assessment), where the level of risk is calculated based on a combination of the Impact of possible losses and the Likelihood of the threat occurring. This calculation results in a Risk value = (I x L), and if the risk is assessed as High or Very High, mitigation actions are directed. This stage is crucial because it identifies the most pressing risks. Finally, Phase 4 (Mitigation and Recommendations) focuses on developing strategies to reduce risk, such as implementing systems that support Transparency (e.g., XAI) and Accountability (e.g., Forensic Audit Logs). This process concludes with a Review of Principles for improving algorithm and data design, ensuring better compliance, and ending with comprehensive Documentation of Results, namely the Risk Analysis Report.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

Table 1. Scope & Preparation

Step	Activity Description	Output/Result
Determine Research Scope	Identify autonomous robot systems (e.g., delivery drones, driverless vehicles) and the types of data collected/used (e.g., location data, environmental sensors, user interactions)	System and Data Scope Document
Step	**Activity Description**	**Output/Result**
Collect Algorithm Documentation	Obtain details on system design, data architecture, and documentation of the robot's decision-making algorithms.	System Technical Documentation

Table 2. Principle Violation Analysis

Step	Activity Description	Output/Result
Transparency Analysis	Check the extent to which the data input process, algorithm processing, and decision output can be explained (Explainability) to users or regulators.	Transparency: Is there a "Black Box" in the decision?
Accountability Analysis	Determine who or what is responsible for incorrect/harmful decisions (e.g., accidents, data discrimination) made by robots.	Accountability: Are there audit and attribution mechanisms?
5. Identify Data Misuse Scenarios	List potential threats (e.g., data leakage, model poisoning, unlawful profiling) arising from a lack of transparency/accountability.	List of Risks and Threats

Table 3. Risk Assessment

Step	Activity Description	put/Result
6. Calculate Impact	Assess the potential losses (financial, reputational, legal, ethical) if the data misuse scenario (from Step 5) actually occurs.	I (Impact)
7. Calculate Probability	Assess the likelihood of each threat scenario occurring based on gaps in transparency/accountability (from Steps 3 & 4).	L ((Probability)
Determine Risk Level	Calculate the overall risk level for each scenario.	Risk = I x L

Table 4. Mitigation & Recommendation

Step	Activity Description	Example Solution
Develop Mitigation Strategies	Design controls and actions to reduce risks considered High or Very High.	Implement Interpretability Tools (for Transparency) or Forensic Audit Logs (for Accountability)
Step	**Activity Description**	**Example Solution**
0. Review Principles	Improve algorithm design and data mechanisms to ensure better compliance with the principles of Transparency and Accountability.	Implement a Decision Explanation System (D-E-S) and a Rollback Mechanism.
1. Document Results	Record all risk findings, assessments, and mitigation actions.	Risk Assessment Report

RESULTS AND DISCUSSION

Autonomous vehicle (AV) systems function as massive data collectors, using various sensors (cameras,LiDAR, radar) to acquire environmental data (maps, road conditions) and highly sensitive personal data (facial images,

license plates, and driver behavior profiles) [17]. Violations of the principle of transparency occur when the vehicle's core decision-making algorithm operates as a "Black Box" model, a condition in which the causal relationship between sensor data input and decision output (e.g., pedestrian identification or emergency braking) cannot be adequately explained or audited [18]. This transparency deficit creates an accountability gap because regulators, users, or victims of incidents cannot verify the underlying reasons why certain decisions were made. This situation inherently increases various risks of data misuse, including potential algorithmic discrimination and exploitation of personal data for invasive profiling purposes, all of which arise from the inability to review or justify the autonomous decision-making process. The risks of data misuse can be seen in Table 5.

Table 5. Risks of data misuse

Risk of Misuse	Contextual Explanation	Impact of Violation Transparency
Algorithmic Discrimination	Biased training data causes algorithms to be less accurate in identifying certain objects (e.g., not recognizing dark-skinned pedestrians or wheelchairs).	If algorithms are not transparent, it is difficult to prove and correct data bias or discriminatory decisions, allowing data misuse (unfair use of data) to continue.
Invasive Data Profiling	Recorded driving habits, daily routes, and environment data are used to create individual profiles (e.g., health conditions, socioeconomic status) without explicit knowledge/consent.	Without clear transparency about how and for what purpose this data is processed, personal data can be easily exploited by manufacturers or third parties for targeted advertising or even behavioral manipulation
Decision Manipulation	The lack of transparency makes it easy for malicious parties (hackers) to find and exploit weaknesses in the model (e.g., adversarial attacks) that cause cars to misinterpret traffic signs or sensors..	Because the decision-making process is not transparent, it is very difficult to track and isolate the point of entry for manipulation (abuse) by hackers, resulting in a very high security risk.

Key Risk Analysis Results (Phase 3)

Based on the application of risk analysis methods, the main focus is on two threat scenarios that have a high level of risk due to violations of Transparency and Accountability, as shown in Table 6.

Table 6. Risk analysis

Threat Scenario	Violation of Key Principles	Probability (L)	Impact (I)	Risk Level (R = IxL)	Category
A1: Decision Discrimination Due to Black Box	ransparency (Algorithm cannot be explained)	High	Very High	**Very High**	Critical Priority
A2: Sensor Data Manipulation by Third Parties	Accountability (Incomplete/ Unauthenticated Audit Log)	Moderate	High	**High**	High Priority

Detailed Discussion of Risk Analysis

The detailed discussion focuses on two main risk scenarios. Transparency violations were identified as the root cause of Black Box Decision Discrimination Risk (A1), which was categorized as a Critical Priority. Findings show that the core algorithm of autonomous robots operates as a "Black Box," so that the causal relationship between environmental sensor data input and decision output cannot be explained, either post-hoc or in-situ [22]. This lack of transparency directly hinders efforts to identify and verify data bias in the model training set. If the training data is disproportionate, the algorithm tends to exhibit discriminatory performance on certain

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

subjects or conditions. The implication is that unclear decision origins can cause autonomous robots to unintentionally engage in algorithmic discrimination, violating ethics and human rights, so this risk is assessed as Very High because it can cause irreparable physical, legal, and reputational damage [23].

Furthermore, Accountability Violations are the main cause of Third Party Sensor Data Manipulation Risk (A2), with a High Priority category. Analysis shows that although the system records operation logs, the existing audit logs are tamper-prone and lack a clear mechanism for attributing responsibility in the decision chain between humans and machines [24]. This accountability weakness makes it easy for third parties (hackers) to misuse data, such as spoofing GPS or sensors, without leaving a reliable audit trail. The lack of authentication in the logs makes it difficult to determine whether a fatal decision was caused by a system bug, sensor error, or malicious intervention. Consequently, this risk of manipulation directly weakens the postincident forensic process, allows for denial of responsibility, and opens up opportunities for data misuse for criminal or espionage purposes.

Mitigation Recommendations (Phase 4)

The To effectively address the identified critical risks particularly Discrimination Risk (A1) and Data Manipulation (A2) mitigation measures are needed that focus on strengthening the two violated principles, namely Transparency and Accountability. Strengthening Transparency (Mitigation A1) is addressed through the implementation of Explainable AI (XAI), which requires systems to be modified so that they are capable of producing interpretable reasoning or explanations (Decision Explanation System / D-E-S) behind every important decision made by the robot (for example, stating 95% confidence in object classification).

1. In addition to XAI, a mandatory Training Data Audit process must be institutionalized to ensure the diversity and neutrality of the training set, which directly reduces the potential for algorithmic discrimination arising from data sources.

2. Meanwhile, Accountability Reinforcement (Mitigation A2) focuses on the implementation of a Tamper Proof Audit Log System. This log mechanism must be authenticated and immutable, potentially using blockchain technology or a centralized time-stamping system, to ensure the integrity of event records for post-incident forensic processes.

3. Finally, a Rollback Mechanism must be built as an accountability protocol, which allows the system to automatically return to a safe state or hand over control to human operators when decision-making uncertainty exceeds a predetermined threshold.

CONCLUSION

This study shows that data misuse in autonomous robot research is directly and causally linked to violations of the principles of transparency and accountability. Autonomous systems that operate as black box algorithms make it difficult to explain the relationship between inputs and decisions, thereby increasing the risk of algorithmic discrimination and unethical use of data. On the other hand, weak accountability mechanisms particularly in audit logs and attribution of responsibility increase the likelihood of sensor data manipulation by third parties.

Risk analysis identifies two threat scenarios with the highest risk levels, namely (A1) decision discrimination and (A2) sensor data manipulation, each triggered by a deficit of transparency and accountability. These results confirm that data security and the reliability of autonomous robots depend not only on the technical quality of the algorithm, but also on the underlying ethical and governance structures.

The resulting mitigation recommendations including the implementation of Explainable AI (XAI), training data audits, tamper-proof audit logging, and rollback mechanisms provide a framework of steps that can be used to improve system design and reduce the risk of data misuse. Thus, this study contributes to strengthening the ethical and security framework in autonomous robot research, and highlights the need for transparency and accountability integration in every stage of autonomous technology development.

ACKNOWLEDGMENTS

The author would like to express his sincere gratitude and appreciation to the Indonesian Education Fund Management Agency (BPDDI) for awarding him a doctoral scholarship. The financial support provided not only enabled him to complete his doctoral studies, but also laid the foundation for conducting in-depth, highquality research.

REFERENCES

1. A. Johnson et al., "Industrial 4.0 Transformation: Autonomous Robotics in Smart Manufacturing," IEEE Trans. Autom. Sci. Eng., vol. 19, no. 2, pp. 45-58, Apr. 2022.
2. B. Miller and C. Zhang, "Deep Sensorimotor Learning for Autonomous Decision-Making in Robotic Systems," IEEE Robot. Autom. Lett., vol. 8, no. 3, pp. 112-125, Jul. 2023.
3. D. Kim and E. Schmidt, "The Transparency Paradox in High-Performance Neural Networks for Robotics," IEEE Trans. Technol. Soc., vol. 4, no. 1, pp. 78-91, Mar. 2024.
4. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems," IEEE, 2021.
5. R. Singh and L. Wang, "Documentation Debt in AI Research: Reproducibility Crisis in Robotics Literature," IEEE Access, vol. 11, pp. 34567-34581, 2023.
6. Y. Zeng and T. Lu, "Bridging the Accountability Gap in Autonomous Systems Through Algorithmic Auditing," Proc. IEEE Conf. Robot. Autom., pp. 234-241, 2022.
7. H. Felzmann et al., "Towards Ethical Transparency in AI Robotics: A Survey of Current Practices," IEEE Trans. Artif. Intell., vol. 3, no. 4, pp. 567-580, Aug. 2023.
8. P. Kumar et al., "Cybersecurity Threats in AI-Driven Robotic Networks: Vulnerability Analysis," IEEE Internet Things J., vol. 9, no. 15, pp. 13456-13470, 2022.
9. L. Chen, "Data Exploitation Vulnerabilities in Opaque Robotic Learning Systems," IEEE Symp. Secur. Priv. Workshops, pp. 89-95, 2024.
10. M. Mueller et al., "Deep Learning Architectures for Autonomous Robotic Perception Systems," IEEE Trans. Neural Netw. Learn. Syst., vol. 35, no. 2, pp. 210-225, Feb. 2024.
11. S. Chen et al., "Multimodal Data Fusion in Autonomous Robots: Challenges and Solutions," IEEE Robot. Autom. Mag., vol. 29, no. 3, pp. 78-94, Sep. 2022.
12. K. Watanabe and T. Yamamoto, "Data Pipeline Transparency in Robotic Systems: An Empirical Study," Proc. IEEE Int. Conf. Robot. Autom., pp. 445-452, 2023.
13. R. Gupta et al., "Explainable AI for Autonomous Navigation Using Layer-wise Relevance Propagation," IEEE Robot. Autom. Lett., vol. 7, no. 4, pp. 11234-11241, Oct. 2022.
14. L. Schmidt et al., "Transparency in Robotics Research: A Systematic Review of XAI Adoption," IEEE Trans. Technol. Soc., vol. 5, no. 2, pp. 145-159, Jun. 2024.
15. A. Johnson and P. Kumar, "Liability Tracing Framework for Autonomous System Failures," IEEE Trans. Emerg. Topics Comput., vol. 11, no. 3, pp. 567-580, Jul. 2023.
16. H. Zhang et al., "Accountability Analysis in Industrial Robotics Incidents," IEEE Trans. Ind. Informat., vol. 19, no. 8, pp. 8901-8912, Aug. 2023.
17. E. Martinez and D. Brown, "Adversarial Attacks on Robotic Perception Systems," IEEE Trans. Inf. Forensics Security, vol. 18, pp. 2456-2470, 2023.
18. F. Costa et al., "Sensor Data Manipulation in Autonomous Vehicles: Security Analysis," IEEE Trans. Intell. Transport. Syst., vol. 24, no. 6, pp. 6345-6358, Jun. 2023.
19. CyberSecurity Robotics Report, "Global Threat Landscape for Autonomous Systems 2024," IEEE Robotics and Automation Society, 2024.
20. T. Nguyen and S. Patel, "Bridging the Gap: Ethical AI and Data Security in Robotics," IEEE Secur. Priv., vol. 21, no. 4, pp. 45-58, Jul. 2023.
21. M. Abramson et al., "Critical Infrastructure Protection for Robotic Systems," Proc. IEEE Conf. Commun. Network Secur., pp. 234-241, 2024.
22. J. Smith, A. B. Jones, and C. D. Lee, "The Accountability Gap in Autonomous Systems: Assessing Black-Box Risks," IEEE Trans. Autom. Sci. Eng., vol. 20, no. 4, pp. 1500-1512, Dec. 2023.
23. M. A. Johnson and T. R. Williams, "Ethical Implications of Data Bias in Machine Learning Algorithms for Autonomous Vehicles," IEEE Technol. Soc. Mag., vol. 42, no. 1, pp. 65-74, Mar. 2023.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

24. S. K. Patel and R. V. Sharma, "Securing Data Integrity via Tamper-Proof Logging in IoT Robotics," in Proc. IEEE Int. Conf. Robot. Autom., May 2024, pp. 280-