



SURFACE VEHICLE INFORMATION REPORT

J3298™

JUL2024

Issued

2024-07

Artificial Intelligence Data for Ground Vehicle Applications

RATIONALE

Given the rapid interest in Artificial Intelligence (AI) and the deployment of AI solutions, it is important to understand the various aspects of data associated with AI for ground mobility, to explicitly state the current issues and note the potential future issues. For the sake of industry standardization of engineering terms, definitions, and applications, the SAE Ground Vehicle Artificial Intelligence Data Task Force has prepared this document for the public, industry, and government to understand the current issues related to data for AI, including the meanings and/or limitations, to mitigate confusion.

INTRODUCTION

In 1956, John McCarthy, American computer scientist and cognitive scientist, coined the term “Artificial Intelligence,” and he defined AI as “the science and engineering of making intelligent machines.” Intelligent machines can be described as computers or computer-controlled systems that perform tasks typically associated with intelligent beings, such as humans. Today, AI is an ever-evolving, fast-moving field that is being applied to various industries. In health care, AI is being used to assist medical professionals in the detection of health-related issues via image analysis. In retail and e-commerce, AI is used by retailers to find patterns in consumer behavior to increase competitive advantage. In manufacturing, AI is being deployed to improve efficiency in the overall workflow.

In the ground vehicle domain, AI is being incorporated into the various facets of mobility, playing a key role in revolutionizing transportation and enhancing overall vehicle capabilities. AI technologies are integrated into various aspects of ground vehicles, ranging from advanced driver assistance systems (ADAS) to automated vehicles (AV). AI enables vehicles to perceive their surroundings through sensors like cameras, lidar, and radar, allowing for real-time analysis and interpretation of complex environments. Machine learning algorithms empower vehicles to make intelligent decisions based on this data, enhancing safety, efficiency, and overall performance. Additionally, AI can contribute to predictive maintenance, optimizing vehicle health and reducing downtime. The continuous evolution of AI in ground vehicles holds the promise of safer, more efficient transportation systems, with the potential to transform how we perceive and interact with automobiles in the future.

Data is at the core of developing the intelligence for the current and perceived future of ground vehicles. Data serves as the foundational resource for training and refining machine learning models that power various AI applications in vehicles. The vast amounts of data collected from sensors and other sources provide crucial insights into the complex dynamics of real-world mobility. This data allows AI models to learn and adapt, improving their ability to accurately perceive and respond to diverse and unpredictable conditions on the road. Moreover, data is instrumental in addressing safety concerns, enabling AI systems to understand rare events and edge cases that are critical for ensuring robust performance in varying situations. Consequently, data diversity is essential to the development of robust and accurate systems. The continuous collection and analysis of data contribute to the ongoing evolution of AI in ground vehicles, fostering innovation, enhancing reliability, and ultimately shaping the future of intelligent transportation systems.

SAE Executive Standards Committee Rules provide that: “This report is published by SAE to advance the state of technical and engineering sciences. The use of this report is entirely voluntary, and its applicability and suitability for any particular use, including any patent infringement arising therefrom, is the sole responsibility of the user.”

SAE reviews each technical report at least every five years at which time it may be revised, reaffirmed, stabilized, or cancelled. SAE invites your written comments and suggestions.

Copyright © 2024 SAE International

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, or used for text and data mining, AI training, or similar technologies, without the prior written permission of SAE.

TO PLACE A DOCUMENT ORDER: Tel: 877-606-7323 (inside USA and Canada)
Tel: +1 724-776-4970 (outside USA)
Fax: 724-776-0790
Email: CustomerService@sae.org
SAE WEB ADDRESS: <http://www.sae.org>

For more information on this standard, visit
https://www.sae.org/standards/content/J3298_202407

Documents
Document Sharing Hub

The impact of data on ground vehicles is profound and multifaceted, influencing various aspects of vehicle operation, safety, and efficiency. All the different types of data could be used for two broad categories of ground vehicle use cases:

- a. To improve the safety, public acceptance, security, privacy, and performance of the ground vehicle itself
- b. To improve the traffic management, cooperative driving, navigation, and environmental impact reduction, etc., that involves the sharing of data and information between multiple traffic participants and infrastructure

Given the importance of data in the development of AI systems, it is important that stakeholders in the ground vehicle domain are knowledgeable of the current state of the industry as relates to data and AI systems. Stakeholders should be aware of some of the common types of datasets that currently exist, the types of licenses that may govern the use of certain datasets, methods and protocols for processing data, and challenges and issues associated with collecting, processing, and retaining data. This document provides information on each of the aforementioned topics.

- General Information Regarding AI and Data for Ground Vehicles

Data for AI development in ground vehicles is generated through a combination of onboard sensors, such as cameras, radar, lidar, and other advanced technologies. These sensors capture information about the vehicle's surroundings, including road conditions, traffic patterns, driver conditions and patterns, pedestrian movement, and potential obstacles. Additionally, vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication systems contribute to the generation of real-time data, enhancing the vehicle's situational awareness.

The collected data is then processed and used to train machine learning models. Machine learning models may be trained via supervised learning, which involves using labeled datasets, where the AI system learns from examples with known outcomes. Unsupervised learning may also be employed to identify patterns and relationships within the data without predefined labels. The continuous feedback loop of data collection, model training, and refinement is crucial for improving the accuracy and adaptability of AI systems.

Furthermore, data generated during actual driving scenarios helps AI models understand and respond to diverse and challenging conditions, contributing to the development of ADAS and automated driving capabilities. This data-driven approach enables vehicles to learn from real-world experiences, increasing safety measures and enhancing overall performance. As technology evolves, the collaborative sharing of anonymized and aggregated data among vehicles and stakeholders also plays a role in advancing AI development in the ground vehicle domain.

- Variances and Discrepancies in AI-Related Data

The ground vehicle domain presents a uniquely challenging landscape for AI, filled with variances and discrepancies in data. Environmental factors like everchanging weather, diverse road conditions, and varying lighting all significantly affect sensor inputs, causing inconsistency in data quality. Additionally, the quality and accuracy of data may be influenced by sensor calibration, maintenance issues, or inherent limitations in the sensing technologies. Furthermore, ground truths, often collected manually, can suffer from subjectivity and human error, introducing further discrepancies. These variations impact both training and performance, as AI models may struggle to generalize from limited data, produce erroneous results from biased datasets, or misinterpret nuanced data patterns. Addressing these challenges requires sophisticated data preprocessing techniques, robust sensor fusion strategies, validation, and careful analysis of ground truth data to mitigate bias and ensure that AI models in ground vehicles are capable of reliable performance across a broad spectrum of real-world conditions.

Many organizations employ standardized internal processes for data management when developing AI systems. However, many researchers, developers, and organizations in the ground vehicle domain are developing AI systems with little to no standard practice across the industry. While the current state of development may be beneficial to a single entity's progress, it could be detrimental the overall progress of AI systems in vehicles. A lack of standard practices, data formats, or recommended processes will make it difficult for vehicles developed by separate organizations to share information between vehicles or with infrastructure. Establishing industry-wide standards can ensure compatibility and interoperability among different AI systems. Developing data governance frameworks that outline the ethical and legal considerations could help address issues related to data privacy, consent, and the responsible use of AI technologies in ground vehicles. Establishing consistent, quantifiable metrics and pushing toward explainability in AI systems in this domain will require actions such as partnerships between the public and private sectors, regulatory oversight, open data initiatives, and continuous evaluation and updating.

TABLE OF CONTENTS

1.	SCOPE.....	4
2.	REFERENCES.....	4
2.1	Applicable Documents	4
2.1.1	SAE Publications.....	4
2.1.2	Other Publications.....	4
2.2	Related Publications	4
2.2.1	SAE Publications.....	4
2.2.2	ANSI Accredited Publications	5
2.2.3	Digital Governance Council Publications	5
2.2.4	DIN Publications.....	5
2.2.5	IEEE Publications.....	5
2.2.6	ISO Publications.....	6
2.2.7	NIST Publications.....	6
3.	TYPES OF STAKEHOLDERS	6
3.1	Original Equipment Manufacturers	6
3.2	Tier 1 Suppliers	6
3.3	Government Agencies.....	6
3.4	Legislators	7
3.5	Infrastructure Owners and Operators	7
3.6	General Public.....	7
3.7	Other AI Committees.....	7
3.8	Academia	7
4.	TYPES OF GROUND VEHICLE DATA	7
4.1	Data Collected by the Vehicle	8
4.1.1	Interior Data	8
4.1.2	Exterior Data	8
4.2	Vehicle-to-Everything (V2X) Data	9
5.	GROUND VEHICLE DATA RESOURCES	9
5.1	Real Datasets.....	10
5.2	Synthetic Data Sets.....	11
6.	PROCESSES AND PROTOCOLS	12
6.1	Data Collection Methods and Issues	12
6.2	Data Processing Methods and Issues	13
6.2.1	Preprocessing Methods	13
6.2.2	Postprocessing Methods.....	15
6.3	Unsupervised Learning	15
6.4	Data Management.....	16
6.4.1	Data Cards	16
6.4.2	Model Cards.....	16
6.5	Privacy and Security	17
7.	SUMMARY AND SUGGESTIONS.....	17
8.	NOTES.....	18
8.1	Revision Indicator.....	18
APPENDIX A		19

1. SCOPE

This SAE Technical Information Report provides preliminary information regarding the current state of data collection, data processing methods, and usage for developing AI and its enabled systems and applications in the ground vehicle domain. This information report is a survey of topics highlighting data's impact on AI solutions and methods that may be used to develop or improve data-related processes.

This report may offer insights that can drive innovation, improve safety, optimize performance, and develop regulatory compliance methods. Solution providers may find this information insightful and realize the potential for collaborative measures in development of their AI-enabled systems. Developers of standards and lawmakers may gain a better understanding of the current state of the industry and find opportunities to develop policies to guide the future of transportation, which in turn directly impacts the public. Other committees and academic affiliates may see links between data for AI in the ground vehicle domain and their domains of interest, which could spawn novel research, development, and standards.

NOTE: While this document is a survey of various topics that pertain to data in AI for ground vehicles, it largely applies to data that is geared toward vehicle navigation and monitoring surroundings that are external to the vehicle.

2. REFERENCES

Section 2 provides other documents that are related to data for AI and is a temporary placeholder for said documents for future discussions.

2.1 Applicable Documents

The following publications form a part of this specification to the extent specified herein. Unless otherwise indicated, the latest issue of SAE publications shall apply.

2.1.1 SAE Publications

Available from SAE International, 400 Commonwealth Drive, Warrendale, PA 15096-0001, Tel: 877-606-7323 (inside USA and Canada) or +1 724-776-4970 (outside USA), www.sae.org.

SAE J2735 V2X Communications Message Set Dictionary

2.1.2 Other Publications

Agarwal, S., Vora, A., Pandey, G., Williams, W., Kourous, H., and McBride, J. (2020). Ford multi-AV seasonal dataset. *The International Journal of Robotics Research*, 39(12), 1367-1376. <https://doi.org/10.1177/0278364920961451>.

Patil, A., Malla, S., Gang, H., and Chen, Y.-T. (2019, May 20-24). *The H3D dataset for full-surround 3D multi-object detection and tracking in crowded urban scenes* [Conference paper]. 2019 IEEE International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada.

2.2 Related Publications

The following publications are provided for information purposes only and are not a required part of this SAE Technical Report.

2.2.1 SAE Publications

Available from SAE International, 400 Commonwealth Drive, Warrendale, PA 15096-0001, Tel: 877-606-7323 (inside USA and Canada) or +1 724-776-4970 (outside USA), www.sae.org.

SAE J3116 Active Safety Pedestrian Test Mannequin Recommendation

SAE J3157 Active Safety Bicyclist Test Targets Recommendation

SAE J3224 V2X Sensor-Sharing for Cooperative and Automated Driving

2.2.2 ANSI Accredited Publications

Copies of these documents are available online at <https://webstore.ansi.org/>.

ANSI/CTA-2090 The Use of Artificial Intelligence in Health Care: Trustworthiness

2.2.3 Digital Governance Council Publications

Available from Digital Governance Council, 1000 Innovation Drive, Suite 500, Ottawa, ON K2K 3E7, <https://dgc-cgn.org/>.

CAN/CIOSC 101:2019 Ethical design and use of automated decision systems

2.2.4 DIN Publications

Copies of these documents are available online at <https://www.din.de/en/>.

DIN SPEC 92001-1:2019 Artificial Intelligence – Life Cycle Processes and Quality Requirements – Part 1: Quality Meta Model

DIN SPEC 92001-2:2019 Artificial Intelligence – Life Cycle Processes and Quality Requirements – Part 2: Robustness

2.2.5 IEEE Publications

Available from IEEE Operations Center, 445 and 501 Hoes Lane, Piscataway, NJ 08854-4141, Tel: 732-981-0060, www.ieee.org.

IEEE P2975.1 Standard for Industrial Artificial Intelligence (AI) Data Attributes

IEEE 1232.1-1997 IEEE Standard for Artificial Intelligence Exchange and Service Tie to All Test Environments (AI-ESTATE): Data and Knowledge Specification

IEEE 2801-2022 IEEE Recommended Practice for the Quality Management of Datasets for Medical Artificial Intelligence

IEEE 2802-2022 IEEE Standard for Performance and Safety Evaluation of Artificial Intelligence Based Medical Devices: Terminology

IEEE 2937-2022 IEEE Standard for Performance Benchmarking for Artificial Intelligence Server Systems

IEEE 2941.1-2022 IEEE Standard for Operator Interfaces of Artificial Intelligence

IEEE 3129-2023 IEEE Standard for Robustness Testing and Evaluation of Artificial Intelligence (AI)-based Image Recognition Service

IEEE 3300-2022 IEEE Standard Adoption of Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) Technical Specification Multimodal Conversion Version 1.2

IEEE 3303-2023 IEEE Standard Adoption of Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) Technical Specification Compression and Understanding of Industrial Data 1.1

IEEE 3304-2023 IEEE Standard for Adoption of Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) Technical Specification Neural Network Watermarking (NNW) V1

IEEE 3806-2023 IEEE Standard for Blockchain-Based Hepatobiliary Disease Data Extraction and Exchange

IEEE 7010-2020 IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being

2.2.6 ISO Publications

Copies of these documents are available online at <https://webstore.ansi.org/>.

ISO/IEC CD TR 5469	Artificial intelligence – Functional safety and AI systems
ISO/IEC DIS 5259-1	Artificial intelligence – Data quality for analytics and machine learning (ML) – Part 1: Overview, terminology, and examples
ISO/IEC DIS 5259-2	Artificial intelligence – Data quality for analytics and machine learning (ML) – Part 2: Data quality measures
ISO/IEC DIS 5259-3	Artificial intelligence – Data quality for analytics and machine learning (ML) – Part 3: Data quality management requirements and guidelines
ISO/IEC TR 24029	Artificial Intelligence (AI) – Assessment of the robustness of neural networks – Part 1: Overview
ISO/IEC TR 24028	Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence
ISO/IEC TS 25058:2024	Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Guidance for quality evaluation of artificial intelligence (AI) systems
ISO/IEC 22989:2022	Information technology – Artificial intelligence – Artificial intelligence concepts and terminology
ISO/IEC 23053	Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
ISO/IEC 38507	Information technology – Governance of IT – Governance implications of the use of artificial intelligence by organizations

2.2.7 NIST Publications

Available from NIST, 100 Bureau Drive, Stop 1070, Gaithersburg, MD 20899-1070, Tel: 301-975-6478, www.nist.gov.

NIST AI 100-1 Artificial Intelligence Risk Management Framework (AI RMF 1.0)

3. TYPES OF STAKEHOLDERS

3.1 Original Equipment Manufacturers

This document may be useful to original equipment manufacturers (OEMs) as they look to understand the current state of data usage and its impact on current and future production of their vehicles. The information provided may help OEMs realize approaches or limitations that could influence design, development, and production of their current and future systems.

3.2 Tier 1 Suppliers

This document may be useful to Tier 1 suppliers who develop hardware and software solutions for ground mobility applications. The information provided may help Tier 1 suppliers realize approaches or limitations that could influence design, development, and production of their current and future systems.

3.3 Government Agencies

This document may be useful for Government Agencies who look to develop policies and provide consistent measures for data usage in the applications of ground mobility (e.g., NIST).