United Nations ECE/TRANS/WP.29/GRVA/2023/17

Economic and Social Council

Distr.: General
10 July 2023

Original: English

Economic Commission for Europe

Inland Transport Committee

World Forum for Harmonization of Vehicle Regulations

Working Party on Automated/Autonomous and Connected Vehicles

Seventeenth session
Geneva, 25-29 September 2023
Item 3 of the provisional agenda
Artificial Intelligence in vehicles

Proposal for a draft resolution with guidance on Artificial Intelligence in the context of road vehicles

Submitted by the expert from the International Organization of Motor Vehicle Manufacturers*

The text reproduced below was prepared by the expert from the International Organization of Motor Vehicle Manufacturers (OICA) and is based on informal document GRVA-16-48. It aims at providing guidance on Artificial Intelligence (AI) in the context of road vehicles.

* In accordance with the programme of work of the Inland Transport Committee for 2023 as outlined in proposed programme budget for 2023 (A/77/6 (part V sect. 20) para 20.6), the World Forum will develop, harmonize and update UN Regulations in order to enhance the performance of vehicles. The present document is submitted in conformity with that mandate.

I. Mandate

1. The Working Party on Automated/Autonomous and Connected Vehicles (GRVA) received inputs regarding Artificial Intelligence in wheeled vehicles, falling in the scope of the World Forum for Harmonization of Vehicle Regulations (WP.29).

2. GRVA consulted the Administrative Committee on the Coordination of Work (WP.29/AC.2) on this matter in Fall 2020. The Committee raised the question of the need to develop a specific Resolution. It decided, for the time being, to request that GRVA continues addressing this item, also with the aim to develop definitions first and then corresponding requirements in the scope of WP.29 activities, if necessary (see GRVA-08-10).

II. Proposal

"Resolution on Artificial Intelligence in the context of road vehicles

Preamble

The [Member States], [Contracting Parties to the 1958 and the 1998 Agreements], participating in the Working Party on Automated/Autonomous and Connected Vehicles,

Having recognized the significant penetration of some Artificial Intelligence (AI) in wheeled vehicles covered in the scope of the agreements administered by the World Forum for Harmonization of Vehicle Regulations (WP.29) [and reviewed the main use cases noted in Annex2],

Having noted that industry currently could use machine-learning algorithms to create frozen software versions and/or to support automated test processes,

Having discussed the technical fundamental aspects of some Machine Learning systems in automotive products, to which the general public refers to as Artificial Intelligence, and discussed corresponding definitions [annexed to this document],

Recalling the adoption of Recommendations on uniform provisions concerning Cyber Security and Software Updates (ECE/TRANS/WP.29/2022/60) and UN Regulation No. 156 (Software Update and Software Update Management System),

Having assessed the importance of proper AI lifecycles for compatibility with existing regulatory regimes and with the safety expectations,

Having acknowledged that the technology is still under development,

Have agreed on the following recommendations using AI-based algorithms within their automotive products:

Software versions

1. This guidance document applies to regulatory requirements for type approval, self-certification, market surveillance, Conformity of Production and Periodic Technical Inspections. Industry should not issue software updates, which will modify or impact an already certified function, without resuming the relevant certification procedure.

2. In the context of the 1958 Agreement and type approval, it means that Industry should not issue software updates, which will modify or impact a type-approved function, without re-engaging with the Approval Authority. Software updates, which do not impact the performance of type-approved characteristics can be deployed without re-engaging with the Approval Authority.

Artificial Intelligence lifecycle

3. It is recommended that a software version should be frozen after having trained an AI-system, which is incorporated in the software. It should be validated and assessed with regards to safety and other relevant requirements. Following that process, the validated software may be deployed in vehicles of a vehicle type.

Training data

3. It is recommended to verify the used data, in terms of data protection and privacy, and vis-à-vis other legal requirements. This resolution is without prejudice to existing market-specific legislations and regulations concerning how personal data is collected and used. Where such regulations exist, they contribute to the overall safety of the AI system through setting personal data management safety standards.

[Annex 1

Simplified definitions in the context of vehicles regulations

The terms below are largely derived from the definitions under review at the International Standard Organization (see ISO/IEC 22989).

- **Agent** is anything which perceives its environment, takes actions autonomously in order to achieve goals, and may improve its performance with learning or may use knowledge.
- **AI lifecycle** consists of the design and development phase of the AI-based system, including but not limited to the collection, selection and processing of data and the choice of the model and the training process, the validation phase, the deployment phase and the monitoring phase. The life cycle ends when the AI-based system is no longer operational.
- **Artificial intelligence (AI)** is a set of methods or automated entities that together build, optimize and apply a model so that the system can, for a given set of predefined tasks, compute predictions, recommendations, or decisions.
- **Bias** is a systematic difference in treatment (including categorization/observation) of certain objects (e.g. natural persons, or groups) in comparison to others.
- **Black box** is a system / software in which the detailed architecture and processing is unknown
- **Black/Grey/White box testing** are tests of systems / software in which architecture and processing is unknown / partially known / known.
- **Connectionist AI (cAI)** systems usually consist of many nodes, called neurons, which are connected with each other in specific patterns, depending on the AI model at hand. Examples of cAI systems are neural networks and support vector machines. In many applications cAI systems are more powerful when compared to sAI systems, e.g. in computer vision. In the majority of cases parameters of cAI systems may not be directly set by the developer. Instead, machine learning algorithms are used together with data to train these systems. The quality of the resulting cAI system is crucially dependent on the quality and quantity of the training data. In contrast to sAI systems cAI systems are in most cases not easily interpretable and not formally verifiable.
- **Data annotation** is the process of attaching a set of descriptive information to data without any change to that data.
- **Data sampling** is a statistical process to select a subset of data intended to present patterns and trends similar to those of the larger dataset being analysed.
- **Dataset** is a collection of data with a shared format and goal-relevant content.
- **Deep learning** is a process whereby neural networks use multiple layers of processing intended to extract progressively higher-level features from data.
- **Explainability** means a property of an AI-based system to express important factors influencing the system's outcome in a way that humans can understand.
- **Fairness / Fairness matrix** is a way of describing bias.
- **Grey box** is a system / software in which the detailed architecture and processing is partially known.
- **Human oversight** is an AI-based system property guaranteeing that built-in operational constraints cannot be overridden by the system itself and are responsive to the human operator, and that the natural persons to whom human oversight is assigned exert ultimate control.

- **Machine learning** is a collection of data-based computational techniques to create an ability to learn without following explicit instructions such that the model's behaviour reflects patterns in data or experience.
- **Machine learning model** is a computer science construct that generates an inference, or prediction, based on input data.
- **Model drift** is a term from the field of machine learning. It refers to the phenomenon that the predictive accuracy of machine learning models can degrade over time. The reasons for this are, for example, that assumptions or variable dependencies that were still valid when the models were created and trained have changed over time. Measures such as retraining or tuning the models can eliminate model drift.
- **Model staleness** is defined as outdated if the trained model does not contain current data and/or does not meet current requirements. Outdated models can affect prediction quality in intelligent software.
- **Online learning** describes incremental training of a new version of the AI-based system during operation onboard production vehicles to achieve defined goals.
- **Predictability** is a property of an AI-based system that enables reliable assumptions by stakeholders about the output.
- **Reinforcement learning** is a discipline of machine learning that permits an agent to learn actions to be taken from patterns in data or experiences, optimizing a quantitative reward function gained along the time.
- **Reliability** is a property of consistent intended behaviour and results.
- **Resilience** is the ability of a system to recover operational condition quickly following an incident.
- **Robustness** is the ability of a system to maintain its level of performance under a wide range of circumstances.
- **Safe-by-design** is a system property enabled by proactive development and lifecycle activities to ensure that risks are brought to an acceptable level through system measures.
- **Semi Supervised learning** is a combination of supervised and unsupervised learning. It uses a small amount of labelled data and a large amount of unlabelled data, which provides the benefits of both unsupervised and supervised learning while avoiding the challenges of finding a large amount of labelled data.
- **Software** is usually created by a process called traditional programming. The programmer manually codes rules using a programming language.
- **Supervised learning** is a type of machine learning that makes use of labelled data during training.
- **Symbolic AI (sAI)** explicitly encodes knowledge using symbolic representations. An example of such a system is a decision tree. Interpreting and formally verifying a sAI system is generally possible and much easier to achieve when compared to connectionist AI systems.
- **Training** is the process to tune the parameters of a machine-learning model.
- **Training data** is a subset of input data samples used to train a machine learning model
- **Transparency of an organization** is the property of an organization that appropriate activities and decisions are documented and communicated to relevant stakeholders in a comprehensive, accessible and understandable manner.
- **Transparency of a system** is property of a system to communicate information to stakeholders.

- **Trustworthiness** is the ability to meet stakeholders' expectations in a verifiable way.
- **Unsupervised learning** is a type of machine learning that makes use of unlabelled data during training.
- **Validation** is done to ensure software usability and capacity to fulfil the customer needs.
- **Validation data** is data used to assess the performance of a final machine learning model
- **Verification** is done to ensure the software is of high quality, well-engineered, robust, and error-free without getting into its usability.
- **White box** is a system / software in which the detailed architecture and processing is known.

Annex 2

Review of use cases in vehicles provided by industry in 2022

	AI Application		Non Safety functions e.g. Infotainment Out of Scope of type approval	Safety functions			
				Driving Function			Non Driving Functions
				Perception	Planning	Actuation	
Conventional Software	**Artifical Intelligence (AI)**	Artificial Intelligence is a set of methods or automated entities that together build, optimize and apply a model so that the system can, for a given set of predefined tasks, compute predictions, recommendations, or decisions	Natural language processing	Out of Scope [Non-AI] Detection of other road users for AEBS, ACC Detection of road infrastructure for LDW, LKAS	Out of Scope Activation of FCW and AEBS based on ego vehicle position and other road users	Not Applicable	Out of Scope Detection of driver's face for ID (under conditions ensuring privacy)
Artificial Intelligence	**Supervised Learning (SL)**	Supervised learning is a type of machine learning that makes use of labelled data during training	Gesture control Voice Recognition	Detection of other road users for AEBS, ACC Detection of passive road infrastructure for LDW, LKAS	Trajectory prediction using drivable path prediction from labelled data (e.g. HD maps)	Not Applicable	Detection of drivers eye gaze / state for DMS Fault detection, Predictive Maintenance
	Unsupervised Learning (UL)	Unsupervised learning is a type of machine learning that makes use of unlabelled data during training		Streamlining data labelling process for less safety critical systems like ISA. Extracting scenarios from real world data to suport validation Generation of synthetic data for supervised learning / distortion of real world data	Trajectory prediction using Kalman filters, KalmanNet or Gaussian Process architectures, or other architectures	Not Applicable	[?]
	Semi Supervised Learning (SSL)	Semi supervised learning is a technique that "learns" from a mix of labelled data and data that is both un-labelled and unstructured. They build on a small set of known exemplars and then use this information to guide unsupervised learning.		Streamlining data labelling process for less safety critical systems like ISA.	Shadow mode' used in development for training control algorithms	Not Applicable	[?]
	Reinforcement Learning (RL)	Reinforcement learning is a type of machine learning utilizing a reward function to optimize a machine learning model by sequential interaction with an environment		Some manufacturers are starting to use RL for perception, could potentionally be used in cooperative perception in the future.	Lane Centering or ACC systems may use RL due to the reduction in cost / data required to train the system	Not Applicable	Predictive Maintenance