

## **Industry input on GRVA-13-04 titled “Outcome of the GRVA workshops on Artificial Intelligence and Vehicle Regulations”**

The input is marked in “tracked change”.

### **I. Mandate**

1. Following the AC.2 decisions of November 2020 and the discussions at the last sessions of GRVA, GRVA requested the secretariat to organize a technical workshop focusing primarily on definitions for Artificial Intelligence, relevant for GRVA activities. The first workshop took place on 18 March 2022. The experts agreed to convene a second workshop on 9 May 2022 to explore the AI use cases and their relevance for GRVA with regards to safety.
2. The experts discussed whether technology neutral performance requirements are sufficient for the purpose of GRVA or if specific provisions would be necessary. The experts developed draft definitions, drafted a table with use cases and their relevance with regards to vehicle regulations and reflected on the potential activities that could be necessary in the framework of the New Assessment Test Method developed by GRVA and its IWG on Validation Method for Automated Driving (VMAD).

### **II. Relevance for GRVA**

3. This short chapter provides two examples aimed at suggesting that GRVA might have to look into Artificial Intelligence in the context of vehicle regulations.

#### **A. Test results reproducibility according to UN GTRs ad UN Regulations**

4. GRVA develops technical requirements and guidance that are technology neutral, unless a specific technology requires appropriate and specific provisions.

5. GRVA discussed (GRVA-12-06) that in the case of functions, which are based on software that is generated by Artificial Intelligence, the outcome associated with this AI for a given situation will not necessarily be predictable.

6. The predictability of test results is an important factor for the type-approval and for the self-certification.

#### **B. Specific features of AI systems used in automotive products**

7. AI systems, used in automotive products, may provide the possibility for offline retraining combined with a thorough validation and Over-the-Air (OTA) updates. This offers a compromise that allows adaptations to model drift and model staleness processes while guaranteeing a certain level of safety and security.

8. GRVA might wish to evaluate whether the provisions regarding software updates (in UN Regulation No. 156 and in the recommendations on uniform provisions concerning cyber security and software updates) adequately address retraining and OTA updates.

### III. List of AI relevant definitions in the context of vehicle regulations

93. The terms below are ~~are taken from~~ ~~inspired by~~ the definitions under review at the International Standard Organization (see ISO/IEC 22989).

~~104.~~ **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.

~~115.~~ **Machine learning** is a ~~collection of~~ ~~data-~~based computational techniques to create an ability to "learn" without an explicitly programmed ~~algorithm~~ ~~outcome~~ such that the model's ~~behaviour~~ ~~behavior~~ reflects the data or experience.

~~126.~~ **Machine learning model** is a ~~mathematical~~ ~~computer science~~ construct that generates an inference, or prediction, based on input data.

~~137.~~ **Deep learning** is ~~an approach to creating rich hierarchical representations through the training of a process whereby~~ neural networks ~~with many hidden~~ ~~use single/multiple layers-~~ ~~of processing intended to extract progressively higher level features from data.~~

~~148.~~ **Supervised learning** is a type of machine learning that makes use of labelled data during training.

~~159.~~ **Unsupervised learning** is a type of machine learning that makes use of ~~unlabelled~~ ~~unlabeled~~ data during training.

~~160.~~ **Reinforcement learning** is a ~~type of machine learning utilizing~~ ~~(collection of) training technique that permits to an agent to learn actions to be taken from experiences, optimizing a quantitative reward function to optimize a machine learning model by sequential interaction with an environment-~~ ~~gained along the time.~~

~~174.~~ **Dataset** is a collection of data with a shared format and goal-relevant content.

~~182.~~ **Data sampling** is a process to select a subset of data ~~samples~~ intended to present patterns and trends similar to ~~that~~ ~~those~~ of the larger dataset being ~~analysed~~ ~~analyzed~~.

~~1943.~~ **Data annotation** is the process of attaching a set of descriptive information to data without any change to that data.

~~2014.~~ **Training** is the process to ~~establish or to improve~~ ~~tune~~ the parameters of a machine-learning model.

~~21.~~ **Training data** is a subset of input data ~~samples used to train~~ a machine learning algorithm, ~~by using training data-~~ ~~model~~

~~22.~~ **Validation data** is data used to assess the performance of a final machine learning model

~~15.~~ **Retraining** is ~~an approach to creating rich hierarchical representations through the training of neural networks with many hidden layers-~~

~~16.~~ **Continuous learning** describes incremental training of an AI system throughout the lifecycle to achieve defined goals governed by pre and post operation risk acceptance criteria and human oversight.

~~17.~~ **Self-learning** describes incremental training of an AI system throughout the lifecycle to achieve defined goals governed by pre and post operation risk acceptance criteria making possible a continuous activation of the new system output with or without human oversight.

~~2348.~~ **Online learning** describes incremental training of a new version of the AI system during operation to achieve defined goals based on post operation acceptance criteria and human oversight without activating the new system output until released.

2419. **Human oversight** is AI system property guaranteeing that built-in operational constraints cannot be overridden by the system itself and is responsive to the human operator, and that the natural persons to whom human oversight is assigned.

2520. **AI lifecycle** consists out of the design and development phase of the AI system, including but not limited to the collection, selection and processing of data and the choice of the model, the validation phase, the deployment phase and the monitoring phase. The life cycle ends when the AI system is no longer operational.

261. **Safe-by-design** is system property enabled by development and lifecycle activities to claim system measures bring risks to an acceptable level.

272. **Trustworthiness** is the ability to meet stakeholders' expectations in a verifiable way.

283. **Bias** is a systematic difference in treatment (including categorization/observation) of certain objects, people, or groups in comparison to others.

294. **Fairness / Fairness matrix** is a way of describing bias.

3025. **Predictability** is a property of an AI system that enables reliable assumptions by stakeholders about the output.

3126. **Reliability** is a property of consistent intended ~~behaviour~~behavior and results.

3227. **Resilience** is the ability of a system to recover operational condition quickly following an incident.

3328. **Robustness** is the ability of a system to maintain its level of performance ~~under any circumstances~~during the whole lifecycle.

3429. **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.

3530. **Transparency of a system** is property of a system to communicate information to stakeholders.

3631. ~~**Explainability**~~ **Explainable**—means a property of an AI system to express important factors influencing the AI system that results in a way that humans can understand.

3732. **Black box** is a systems / software in which the architecture and processing~~detailed functionality~~ ~~is unknown~~Black/Grey/White box [testing] are [tests of] systems / software in which functionality are unknown / partially know / known.

38. **Grey box** is a systems / software in which the architecture and processing~~detailed functionality~~ ~~is partially known~~.

39. **White box** is a systems / software in which the architecture and processing~~detailed functionality~~ ~~is known~~.

40. **Black/Grey/White box testing** are tests of systems / software in which functionalities are unknown / partially know / known.

### III. AI use cases in the automotive sector

Note: The following table was prepared by the experts from CLEPA and OICA

An editable version of this table is available here: <https://unece.org/transport/events/grva-technical-workshop-artificial-intelligence-2nd>

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	<b>Artificial Intelligence (AI)</b> 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	<b>Supervised Learning (SL)</b> 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
	<b>Unsupervised Learning (UL)</b> 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 support 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	[?]
	<b>Semi Supervised Learning (SSL)</b> 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	[?]
	<b>Reinforcement Learning (RL)</b> 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 potentially 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

## IV. Impact of Artificial intelligence on the New Assessment Test Method

Note: The following table was prepared by the experts from CLEPA and OICA.

An editable version of this table is available here: <https://unece.org/transport/events/grva-technical-workshop-artificial-intelligence-2nd>

