Workshop on Artificial Intelligence and Vehicle Regulations

I. Mandate

1. Following the AC.2 decisions of November 2020 and the discussions at the last 4 sessions of GRVA, GRVA requested the secretariat to organize a technical workshop focusing primarily on definitions for Artificial Intelligence, relevant for GRVA activities, and, if possible i.e. if time is available, exploring more in detail the potential role of vehicle regulation(s) and guidance document(s) with regard to AI (see decision 4 of the list of decisions of the 12th GRVA session).

II. Relevance for GRVA

2. 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

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

4. 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 reproducible.

5. The reproducibility 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

6. 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.

7. 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. Proposed items for discussions and related definitions

8. **Artificial intelligence**[i] [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~~is a general concept defined in contrast to the natural intelligence, it is the intelligence exhibited by machines or software.~~]

9. ~~Artificial Intelligence can be subdivided into three categories:~~

[(a) **Artificial Super Intelligence**[i] (ASI), which is exhibiting human intelligence,

---

[1] Taken from the OICA presentation at the Webinar on 21 January 2022, see https://wiki.unece.org/display/trans/Webinar+1x

(b)   ~~**Artificial General Intelligence**[1] (AGI), where the system is able to act like a human in all fields, and~~

~~(c)   **Artificial Narrow Intelligence**[1] (ANI), which focuses on a specific task.]~~

.   **Machine learning**[2] [is a data based computational techniques to create an ability to "learn" ~~(i.e. progressively improve performance on tasks)~~ without an explicitly programmed algorithm such that the model's behaviour reflects the data or experience. ~~ML systems are mainly trained using three methods:]~~

Machine learning Model [is a mathematical construct that generates an inference, or prediction, based on input data

Deep learning is an approach to creating rich hierarchical representations through the training of neural networks with many hidden layers

(a)   **Supervised learning**[3] ~~[aims to develop predictive model based both input and output data.]~~is a type of machine learning that makes use of labelled data during training

(b)   **Unsupervised learning**~~3~~ is a type of machine learning that makes use of unlabelled data during training~~[aims to discover an internal representation from input data only.]~~

(c)   **Reinforced learning**~~3~~ is a type of machine learning utilizing a reward function to optimize a machine learning model by sequential interaction with an environment ~~[is concerned with how a so-called agent should take actions in an environment to maximize some notion of long-term reward]~~

~~.   Artificial Neural Networks3 (ANNs) [are designed to work by classifying information, generally without task-specific programming o r models (as found in usual algorithmics), and are capable of progressively improving their performance and mitigating losses of accuracy.]~~

~~.   A Deep Neural Network3 (DNN) [is an Artificial Neural Network and is part of the class of Machine Learning. It uses a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input.3]~~ Dataset collection of data with a shared format and goal-relevant content

Data sampling is a process to select a subset of data samples intended to present patterns and trends similar to that of the larger dataset being analysed

data annotation is the process of attaching a set of descriptive information to data without any change to that data

Training is the process to establish or to improve the parameters of a machine learning model, based on a machine learning algorithm, by using training data

Retraining is an approach to creating rich hierarchical representations through the training of neural networks with many hidden layers

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

---

[2] Expert from Germany, GRVA-12-03, 12th session of GRVA, January 2022
~~[3] Note by the secretariat, WP.29-175-21, 175th session of WP.29, June 2018~~

Self-Learning[4] describes [incremental training of an AI system troughout the lifecycle to achieve defined goals governed by pre and post operation risk acceptance criterias making posible a continous activation of the new system output with or without human oversight.~~describes the ability of an algorithm to recognize patterns in data and to improve its performance over time automatically~~].

. **Online Learning**[2] describes incremental training of an new version of the AI system during operation to achieve defined goals based on post operation acceptance criterias and human oversight without activating the new system output until released. [~~refers to training the system incrementally on the edge, where data is received in sequential order. Applying it, should not lead to a situation where two vehicles of the same type differ in their retrained AI models.~~]

Human Oversight AI system property guaranteing 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.

. ~~**Black Box effect**[4] [describes that AI systems, depending on the model and its complexity, are not transparent and their decisions not explainable.] (White/Grey Box definition needed?)[5]~~

. The AI **life-cycle** ~~of an AI system~~[2] [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~~]~~.

. ~~**"Closed methodology"**[2] [defines the sequence of the following processes:~~

~~(a) An AI agent may be trained to produce a software.~~

~~(b) Once satisfactory results are reached, the software should be frozen.~~

~~(c) The frozen software should then be validated.~~

~~(d) Once properly validated, such software can be placed in vehicle.]~~

. **"Safe by Design"**[5] system property enabled by development and lifecycle activities to claim system measures –bring risks to an acceptable level[~~refers to the establishment of auditing processes during the development period of an AI system, which include but are not limited to audit on training data, on algorithm models used, on bias etc.~~]

. ~~**Probabilistic nature**[4] [of function based on a software developed using AI (in the ANI context) means: ...]~~

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

Bias is an approach to creating rich hierarchical representations through the training of neural networks with many hidden layers

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

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 any circumstances

Transparency of an organization is the property of an organization that appropriate activities and decisions are communicated to relevant stakeholders in a comprehensive, accessible and understandable manner

Transparency of a System property of a system to communicate information to stakeholders~~is the property of a system that appropriate information about the system is communicated to relevant stakeholders~~

IV. Potential role of vehicle regulation(s) and guidance document(s) with regard to AI

(forthcoming)