



Commission économique pour l'Europe**Comité des transports intérieurs****Forum mondial de l'harmonisation des Règlements
concernant les véhicules****192^e session**

Genève, 5-8 mars 2024

Point 2.3 de l'ordre du jour provisoire

Coordination et organisation des travaux :**Systèmes de transport intelligents et coordination
des activités relatives aux véhicules automatisés****Proposition de projet de résolution visant à fournir
des orientations sur l'intelligence artificielle
dans le contexte des véhicules routiers****Note du secrétariat***

Le texte ci-après est fondé sur le document informel WP.29-191-06, qui a été examiné par le Forum mondial de l'harmonisation des Règlements concernant les véhicules (WP.29) à sa session de novembre 2023. Le présent document est une proposition de projet de résolution visant à fournir des orientations sur l'intelligence artificielle dans le contexte des véhicules routiers. Il comprend des définitions des caractéristiques spécifiques des systèmes basés sur l'intelligence artificielle utilisés dans les produits automobiles, un examen des cas d'utilisation dans les véhicules communiqués par le secteur et un aperçu des interactions entre les principes relatifs à l'intelligence artificielle et la nouvelle méthode d'évaluation et d'essai de la conduite automatisée. Il est établi comme suite à la demande formulée par le WP.29 à sa session de novembre 2023 (ECE/TRANS/WP.29/1175, par. 25). Il est soumis au WP.29 pour examen et adoption éventuelle. Le WP.29 est convenu que ce document pourrait être modifié par un document informel établi par le Groupe de travail des véhicules automatisés/autonomes et connectés (GRVA) à sa session de janvier 2024.

* Conformément au programme de travail du Comité des transports intérieurs pour 2024 tel qu'il figure dans le projet de budget-programme pour 2024 (A/78/6 (Sect. 20), tableau 20.5), le Forum mondial a pour mission d'élaborer, d'harmoniser et de mettre à jour les Règlements ONU en vue d'améliorer les caractéristiques fonctionnelles des véhicules. Le présent document est soumis en vertu de ce mandat.



Préambule

Les Parties contractantes aux Accords de 1958 et de 1998 participant aux travaux du Groupe de travail des véhicules automatisés/autonomes et connectés,

Ayant constaté la place toujours plus importante de l'intelligence artificielle dans les véhicules à roues relevant du champ d'application des accords administrés par le Forum mondial de l'harmonisation des Règlements concernant les véhicules (WP.29),

Ayant relevé que le secteur pourrait actuellement utiliser les outils de l'apprentissage automatique pour appuyer le développement ou l'essai de logiciels avant leur mise en service,

Ayant examiné les aspects techniques fondamentaux de certains systèmes d'apprentissage automatique intégrés dans les produits automobiles, que le grand public connaît sous le nom d'intelligence artificielle, et réfléchi aux définitions correspondantes,

Rappelant l'adoption des Recommandations relatives à des prescriptions uniformes concernant la cybersécurité et les mises à jour logicielles,

Ayant réfléchi à l'importance que revêt le caractère adéquat des cycles de vie de l'intelligence artificielle pour la compatibilité avec les régimes de certification existants,

Conscientes du fait que l'usage de cette technologie pour les applications automobiles est encore en évolution,

Ont arrêté ensemble les recommandations suivantes relatives à l'utilisation d'algorithmes basés sur l'intelligence artificielle dans leurs produits automobiles :

Mises à jour logicielles

1. Le présent document d'orientation s'applique aux prescriptions de certification et à la conformité de la production. Le secteur ne devrait pas publier de mises à jour logicielles qui modifient sensiblement des fonctions déjà certifiées conformément aux recommandations relatives à des prescriptions uniformes concernant la cybersécurité et les mises à jour logicielles sans recommencer la procédure de certification correspondante.
2. Il est recommandé, après avoir entraîné le système d'intelligence artificielle qui sera intégré au logiciel, de faire valider celui-ci par une entité habilitée ou dans le cadre d'un processus de certification et de l'évaluer au regard des prescriptions en matière de sécurité, de sûreté et d'environnement et des autres prescriptions pertinentes. Les systèmes non certifiés faisant appel à l'intelligence artificielle ne doivent pas influencer sur les systèmes certifiés de manière à nuire à la certification. À l'issue de ce processus, le logiciel validé peut être déployé dans des véhicules d'un type donné.

Données à utiliser pour la mise au point de systèmes basés sur l'intelligence artificielle

3. On suppose que les règlements relatifs à la protection des données et de la vie privée ainsi que les autres obligations légales sont respectés. La présente résolution s'entend sans préjudice des dispositions législatives et réglementaires propres à un marché donné qui régissent la manière dont les données personnelles sont collectées et utilisées. Lorsque de telles réglementations existent, elles contribuent à la sécurité globale du système d'intelligence artificielle en fixant des normes de sécurité en matière de gestion des données personnelles.

Annexe 1

Définitions simplifiées applicables dans le contexte des règlements concernant les véhicules – Caractéristiques particulières des systèmes basés sur l'intelligence artificielle utilisés dans les produits automobiles

1. Dans ce qui suit, l'expression « systèmes basés sur l'intelligence artificielle » désigne les systèmes d'intelligence artificielle connexionnistes, tels que les réseaux de neurones, qui sont entraînés à l'aide d'algorithmes d'apprentissage automatique et de données. Ces systèmes présentent des propriétés nouvelles sur le plan qualitatif, qui ouvrent de nouvelles perspectives et posent de nouveaux défis.
2. Les systèmes basés sur l'intelligence artificielle utilisés dans les produits automobiles peuvent permettre un compromis entre les différentes caractéristiques souhaitables du modèle : dérive et obsolescence du modèle, complexité, robustesse, vérifiabilité, prévisibilité et surajustement du modèle, etc. tout en garantissant un certain niveau de sûreté et de sécurité. Les systèmes basés sur l'intelligence artificielle devraient offrir des possibilités de mise à jour.
3. D'autres évaluations périodiques pourraient être nécessaires pour vérifier si les dispositions relatives aux mises à jour logicielles (figurant dans les recommandations relatives à des prescriptions uniformes concernant la cybersécurité et les mises à jour logicielles) traitent de manière adéquate les mises à jour des systèmes basés sur l'intelligence artificielle.
4. Les systèmes basés sur l'intelligence artificielle peuvent contribuer à améliorer la sécurité des véhicules, avec des conséquences positives supplémentaires sur la sécurité routière, par exemple en permettant aux systèmes de conduite automatisés de prévoir des comportements actuellement imprévisibles d'autres usagers de la route (par exemple, la détection d'autres usagers de la route susceptibles d'entrer en collision avec le véhicule).
5. L'utilisation de l'intelligence artificielle et des algorithmes d'apprentissage automatique dans les fonctions homologuées est limitée pour l'instant. S'il existe déjà des processus bien établis pour tester les logiciels conventionnels avant et pendant le déploiement d'un produit automobile, ces processus pourraient être insuffisants pour les logiciels basés sur l'intelligence artificielle. Qu'ils soient créés par l'apprentissage automatique ou non, il convient de tester les logiciels avant leur mise en service afin de s'assurer qu'ils sont conformes à l'ensemble des lois, règlements, recommandations et politiques applicables. Cela vaut également pour le processus de mise à jour. Toutefois, il convient d'évaluer dans quelle mesure les dispositions réglementaires en vigueur peuvent répondre de manière satisfaisante aux besoins spécifiques en matière de test et de mise à jour des logiciels basés sur l'intelligence artificielle et garantir leur bon fonctionnement.
6. Les termes ci-dessous sont en grande partie tirés des définitions de l'Organisation internationale de normalisation (ISO) (voir ISO/IEC 22989), du Ground Vehicle AI Committee de SAE International et de l'Organisation de coopération et de développement économiques (OCDE). La liste des termes n'est pas exhaustive et les définitions fournies sont simplifiées et peuvent donc ne pas être adaptées à des textes réglementaires.
7. Il est d'usage de procéder à des tests approfondis sur les systèmes de boîte blanche/grise/noire afin de garantir le bon fonctionnement du système certifié.

Note : Certaines définitions ci-dessous ont été reprises (ou influencées) par l'OCDE¹, SAE International² ou l'ISO³ :

- **Agent** : tout élément qui perçoit son environnement, agit de manière autonome pour atteindre des objectifs, et peut améliorer son efficacité grâce à l'apprentissage ou utiliser des connaissances.

- **Cycle de vie de l'intelligence artificielle** : phase de conception et de développement du système basé sur l'intelligence artificielle, notamment, la collecte, la sélection et le traitement des données, le choix du modèle et le processus d'entraînement, la phase de validation, la phase de déploiement et la phase de surveillance. Ce cycle prend fin lorsque le système basé sur l'intelligence artificielle n'est plus opérationnel.
- **Intelligence artificielle**³ : ensemble de méthodes ou d'entités automatisées qui, ensemble, construisent, optimisent et appliquent un modèle permettant au système, pour un ensemble donné de tâches prédéfinies, de calculer des prédictions, des recommandations ou des décisions.
- **Système d'intelligence artificielle**¹ : un système automatisé capable d'influencer l'environnement en produisant des résultats (prédictions, recommandations ou décisions) en vue de réaliser un ensemble donné d'objectifs. Il s'appuie sur l'apprentissage automatique et/ou des données fournies par l'homme pour percevoir des environnements réels ou virtuels, convertir ces perceptions en modèles au moyen d'une analyse automatisée (par exemple, avec l'apprentissage automatique) ou manuelle, et utiliser ces modèles pour formuler, par inférence, des options de résultats. Les systèmes d'intelligence artificielle sont conçus pour fonctionner avec différents niveaux d'autonomie.
- **Biais** : différence systématique dans le traitement (y compris la catégorisation/l'observation) de certains objets (par exemple des personnes physiques ou des groupes) par rapport à d'autres.
- **Boîte noire** : système/logiciel dont l'architecture détaillée et le traitement sont inconnus.
- **Tests en boîte noire/grise/blanche** : tests de systèmes ou logiciels dont l'architecture et le traitement sont inconnus/partiellement connus/connus.
- **Systèmes d'intelligence artificielle connexionnistes** : systèmes généralement constitués de nombreux nœuds, appelés neurones, qui sont interconnectés selon des structures particulières, en fonction du modèle d'intelligence artificielle utilisé. Les réseaux neuronaux et les machines à vecteurs de support sont des exemples de systèmes d'intelligence artificielle connexionnistes. Pour de nombreux cas d'utilisation, les systèmes d'intelligence artificielle de type connexionniste s'avèrent plus puissants que ceux de type symbolique, par exemple dans le domaine de la vision par ordinateur. Dans la majorité des cas, les paramètres des systèmes d'intelligence artificielle connexionnistes ne peuvent pas être définis directement par le développeur. Ce sont des algorithmes d'apprentissage automatique couplés à des données qui sont utilisés à la place pour entraîner ces systèmes. La qualité du système d'intelligence artificielle connexionniste qui en résulte dépend essentiellement de la qualité et de la quantité des données d'entraînement. Contrairement aux systèmes d'intelligence artificielle symboliques, les systèmes connexionnistes sont, dans la plupart des cas, difficilement interprétables et non vérifiables formellement.
- **Logiciel classique** : en général, fruit d'un processus appelé programmation traditionnelle. Le programmeur code manuellement les règles en utilisant un langage de programmation.
- **Annotation des données** : processus qui consiste à associer un ensemble d'informations descriptives à des données sans modifier ces dernières.
- **Échantillonnage des données** : processus statistique qui consiste à sélectionner un sous-ensemble de données devant faire ressortir des formes et des tendances semblables à celles mises en évidence grâce à l'analyse de l'ensemble complet de données.
- **Ensemble de données** : jeu de données qui ont le même format et dont le contenu est pertinent pour l'objectif.
- **Apprentissage profond** : processus par lequel les réseaux neuronaux utilisent plusieurs couches de traitement devant permettre d'extraire progressivement à partir des données des caractéristiques de niveau supérieur.

- **Explicabilité** : propriété d'un système basé sur l'intelligence artificielle qui permet d'exprimer les facteurs importants influençant le résultat du système d'une manière compréhensible pour l'être humain.
- **Équité/matrice d'équité** : une des façons de décrire les biais.
- **Boîte grise** : système/logiciel dont les détails de l'architecture et du traitement sont partiellement connus.
- **Contrôle humain** : propriété du système basée sur l'intelligence artificielle qui garantit que les contraintes opérationnelles intégrées ne peuvent pas être ignorées par le système lui-même et qu'elles tiennent compte de l'opérateur humain, et que les personnes physiques auxquelles la supervision humaine est confiée exercent le contrôle en dernier lieu.
- **Apprentissage automatique** : ensemble de méthodes de calcul fondées sur des données qui permettent de créer la capacité d'apprendre sans suivre d'instructions explicites, de sorte que le comportement du modèle fait ressortir des formes dans les données ou l'expérience.
- **Modèle d'apprentissage automatique** : construction informatique qui génère une inférence, ou une prédiction, basée sur des données d'entrée.
- **Modèle** : représentation physique, mathématique ou logique d'un système, d'une entité, d'un phénomène, d'un processus ou de données.
- **Dérive du modèle** : terme utilisé dans le domaine de l'apprentissage automatique. Il désigne le phénomène selon lequel la précision des prédictions des modèles d'apprentissage automatique peut se dégrader au fil du temps. Les raisons en sont, par exemple, que les hypothèses ou les dépendances des variables qui étaient encore valables lorsque les modèles ont été créés et entraînés ont évolué au fil du temps. Des mesures telles que le réentraînement ou l'ajustement des modèles peuvent permettre d'éliminer les dérives.
- **Obsolescence du modèle** : il y a obsolescence du modèle entraîné lorsque celui-ci ne contient pas de données fraîches et/ou ne répond pas aux prescriptions actuelles. Les modèles obsolètes peuvent compromettre la qualité de la prédiction dans les logiciels intelligents.
- **Apprentissage en ligne** : entraînement progressif d'une nouvelle version d'un système basé sur l'intelligence artificielle pendant le fonctionnement de ce système en vue d'atteindre des objectifs précis. Les parties du système soumises à l'apprentissage en ligne doivent être découplées du système de commande du véhicule afin de garantir que la nouvelle version du système basé sur l'intelligence artificielle n'ait pas d'incidence sur le fonctionnement du véhicule tant que sa conformité avec les règles de sécurité applicables n'a pas été vérifiée.
- **Prévisibilité** : propriété d'un système basé sur l'intelligence artificielle qui permet aux parties prenantes de formuler des hypothèses fiables sur les résultats.
- **Apprentissage par renforcement** : discipline de l'apprentissage automatique qui permet à un agent d'apprendre à prendre des décisions à partir de formes dans les données ou les expériences, en maximisant une récompense quantitative au fil du temps.
- **Fiabilité** : propriété qui permet d'obtenir un comportement et des résultats cohérents.
- **Résilience** : capacité d'un système à retrouver rapidement son état de fonctionnement après un incident.
- **Robustesse** : capacité d'un système à maintenir son niveau de performance dans toutes sortes de circonstances. Il s'agit notamment de la capacité d'un système à faire face à des perturbations naturelles et malveillantes dans l'espace d'entrée du système.
- **Sécurité dès la conception** : propriété du système rendue possible grâce à des actions en amont relatives au développement et au cycle de vie, qui vise à garantir que les risques sont ramenés à un niveau acceptable grâce aux mesures prises concernant ce système.

- **Apprentissage semi-supervisé** : apprentissage combinant les apprentissages supervisé et non supervisé. Il requiert une petite quantité de données étiquetées et une grande quantité de données non étiquetées, ce qui permet de bénéficier des avantages à la fois de l'apprentissage non supervisé et de l'apprentissage supervisé tout en évitant les difficultés liées à la recherche d'une grande quantité de données étiquetées.
- **Apprentissage supervisé** : type d'apprentissage automatique qui utilise des données étiquetées pendant l'entraînement.
- **Intelligence artificielle symbolique** : les systèmes d'intelligence artificielle symbolique encodent explicitement les connaissances à l'aide de représentations symboliques. Les arbres de décisions sont un exemple de ce type de système. Il est généralement possible et beaucoup plus facile d'interpréter et de vérifier formellement un système de type symbolique qu'un système de type connexionniste.
- **Entraînement** : processus qui vise à ajuster les paramètres d'un modèle d'apprentissage automatique.
- **Données d'entraînement** : sous-ensemble des échantillons de données d'entrée utilisés pour entraîner un modèle d'apprentissage automatique.
- **Transparence d'une organisation** : propriété d'une organisation selon laquelle les activités et les décisions appropriées sont solidement étayées et communiquées aux parties prenantes concernées de manière exhaustive, accessible et compréhensible.
- **Transparence d'un système** : propriété d'un système qui permet de communiquer des informations aux parties prenantes.
- **Crédibilité** : capacité à répondre aux attentes des parties prenantes de manière vérifiable.
- **Apprentissage non supervisé** : type d'apprentissage automatique qui utilise des données non étiquetées pendant l'entraînement.
- **Validation** : permet de s'assurer de la facilité d'utilisation du logiciel et de sa capacité à répondre aux besoins du client.
- **Données de validation** : données utilisées pour évaluer l'efficacité d'un modèle final d'apprentissage automatique.
- **Vérification** : opération effectuée dans le but de s'assurer que le logiciel est de bonne qualité, bien conçu, robuste et exempt d'erreurs, sans répercussions sur sa facilité d'utilisation.
- **Boîte blanche** : système/logiciel dont l'architecture détaillée et le traitement sont connus.

Annexe 2

Examen des cas d'utilisation dans les véhicules communiqués par le secteur en 2022

Type d'intelligence artificielle AI	Type d'apprentissage automatique	Fonctions non liées à la sécurité Hors du champ d'application de l'homologation de type	Fonctions liées à la sécurité			
			Fonction de conduite			Fonctions non liées à la conduite
			Perception	Planification	Commande	
Logiciel classique	Aucun		Hors du champ d'application (dans le cadre du présent document)	Hors du champ d'application (dans le cadre du présent document)	Hors du champ d'application (dans le cadre du présent document)	Hors du champ d'application (dans le cadre du présent document)
Intelligence artificielle symbolique	Aucun ou tout type d'apprentissage automatique	Par exemple infodivertissement, traitement du langage naturel	Par exemple détection des autres usagers de la route pour AEBS, régulateurs ACC Détection des infrastructures routières passives pour systèmes LDW et ALKS	Par exemple activation des systèmes FCW et AEBS en fonction de la position du véhicule soumis à l'essai et des autres usagers de la route	Sans objet actuellement	Par exemple détection du visage du conducteur pour l'identification (dans des conditions garantissant le respect de la vie privée), éthylotest anti-démarrage
Système d'intelligence artificielle connexionniste + Apprentissage automatique	Apprentissage supervisé	Commande gestuelle Reconnaissance vocale	Détection des autres usagers de la route pour AEBS, régulateurs ACC Détection des infrastructures routières passives pour systèmes LDW et ALKS	Prédiction de trajectoire à partir de données étiquetées (par exemple, cartes HD)	Sans objet actuellement	Détection du regard/de l'état du conducteur pour le système de surveillance du conducteur. Détection des défaillances, Maintenance prédictive
	Apprentissage non supervisé		Rationalisation du processus d'étiquetage des données pour les systèmes moins critiques pour la sécurité, tels que les ISA	Prédiction de trajectoire à l'aide de filtres Kalman ou d'architectures KalmanNet, gaussiennes ou autres	Sans objet actuellement	Détection des défaillances (détection non supervisée d'anomalie)

Type d'intelligence artificielle AI	Type d'apprentissage automatique	Fonctions non liées à la sécurité Hors du champ d'application de l'homologation de type	Fonctions liées à la sécurité			
			Fonction de conduite			Fonctions non liées à la conduite
			Perception	Planification	Commande	
Système d' intelligence artificielle connexionniste + Apprentissage automatique			Extraction de scénarios à partir de données du monde réel pour appuyer la validation Génération de données de synthèse pour l'apprentissage supervisé/distorsion des données du monde réel			
	Apprentissage semi-supervisé		Rationalisation du processus d'étiquetage des données pour les systèmes moins critiques pour la sécurité, tels que les ISA	La technique dite « shadow mode » est utilisée dans le cadre du développement d'algorithmes de contrôle de l'entraînement	Sans objet actuellement	
	Apprentissage par renforcement		Certains fabricants commencent à utiliser l'apprentissage par renforcement pour la perception ; celui-ci pourrait être utilisé dans le cadre de la perception coopérative à l'avenir	L'apprentissage par renforcement peut être utilisé pour les systèmes de centrage de voie ou ACC compte tenu de la réduction du coût et des données nécessaires à l'entraînement du système	Sans objet actuellement	Maintenance prédictive

Nota : Les cellules grisées indiquent les éléments qui ne relèvent pas du champ d'application du présent document.

Annexe 3

Incidence de l'intelligence artificielle sur les nouvelles méthodes d'évaluation et d'essai

Pour un aperçu complet et schématique de l'interaction des différents piliers, scénarios et prescriptions de sécurité des systèmes très complexes tels que ceux utilisés dans la conduite automatisée, il convient de consulter le document de référence de la CEE intitulé « Nouvelle méthode d'évaluation et d'essai de la conduite automatisée » dans sa version actuelle. Les systèmes basés sur l'intelligence artificielle peuvent également être conçus de la même manière.

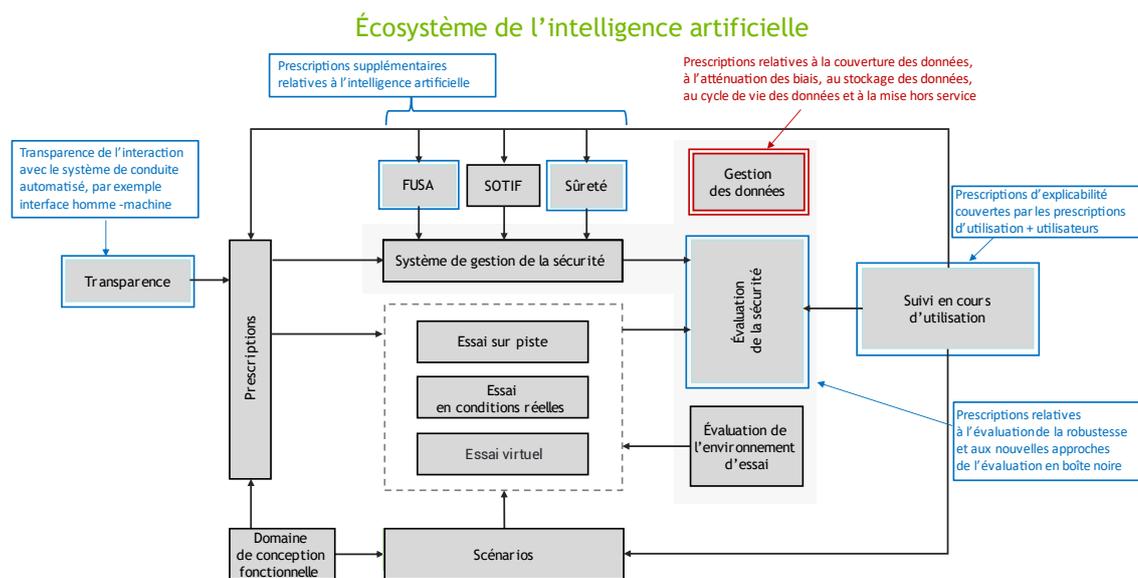


Figure 1 : Exemple de schéma

Note du secrétariat:

La figure ci-dessus est une représentation de la nouvelle méthode d'évaluation et d'essai de la conduite automatisée, en gris, avec deux cases supplémentaires (« Transparence » et « Gestion des données »). Le texte en couleur explique comment la NATM peut s'appliquer aux systèmes basés sur l'intelligence artificielle et aider à évaluer les principes clefs applicables à l'intelligence artificielle tels que la transparence, l'extensibilité, la robustesse, etc.