Projet PRISSMA - Plateforme de Recherche et d'Investissement pour la Sûreté et la Sécurité de la Mobilité Autonome (2021-2024)

Le développement des systèmes de transport routier automatisés est étroitement lié à la capacité à vérifier la sécurité du système global opéré sur un parcours ou une zone prédéfini. Cette validation des systèmes peut inclure des briques spécifiques à l'intelligence artificielle (IA), intégrées dans la notion de système de systèmes.

Le contenu de cette fiche est inspiré des travaux produits dans le cadre du projet français PRISSMA.

Le projet PRISSMA, pilier d'un Grand Défi IA

Créé en 2018, le Fonds pour l'innovation et l'industrie finance des projets d'innovation de rupture, afin de garantir la souveraineté scientifique et technologique, et le développement économique de secteurs industriels stratégiques français. Depuis sa création, cinq grands défis portant sur l'intelligence artificielle (IA) ont été retenus. Le projet PRISSMA s'inscrit dans le second défi, intitulé « Sécuriser, certifier et fiabiliser les systèmes fondés sur l'intelligence artificielle », qui s'articule autour de trois piliers :

- Le premier pilier, organisé autour du projet confiance.ai, se concentre sur la conception et l'industrialisation de systèmes à base d'IA de confiance ;
- Le second pilier, structuré autour du projet PRISSMA, porte sur l'évaluation des systèmes à base d'IA afin d'en garantir le fonctionnement ;
- Le troisième pilier travaille à la définition d'un environnement normatif pour la certification des systèmes à base d'IA.

Objectifs du projet PRISSMA

L'objectif de PRISSMA est de délivrer une méthode de validation des systèmes de mobilité intégrant des briques IA, avec en particulier une démonstration de sécurité.

Plus précisément, le projet visait à :

- Identifier et recenser les objectifs de sécurité et de sûreté pour les systèmes de mobilité autonome à base d'IA et développer les processus complets de validation de la fiabilité en vue d'une mise en exploitation commerciale des services de mobilité autonome ;
- Assurer la disponibilité de concepts partagés permettant de répondre à la complexité des systèmes de mobilité autonome à base d'IA, pouvant être utilisés au niveau international ;
- Participer à la mise en œuvre de prérequis permettant à la France de se positionner au niveau européen pour accueillir un des centres de tests (Testing Facilities) pour la mobilité autonome qui seront développés dans les prochaines années.

Principaux enjeux de l'intégration de systèmes d'IA dans les systèmes de mobilité

Les systèmes d'IA génèrent des problématiques nouvelles et spécifiques qui sont résumées ici :

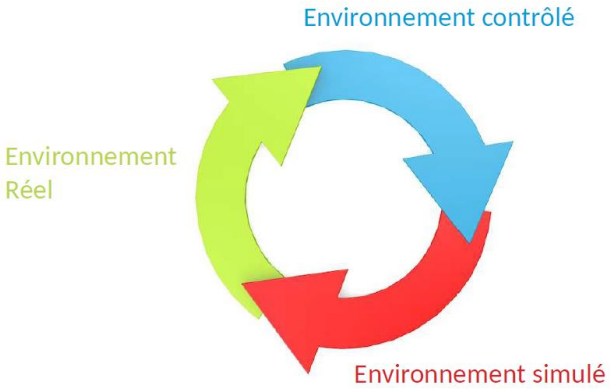
- Leur perception de l'environnement est différente de celle de l'humain ;
- Leurs réponses à des problématiques données sont probabilistes et liées à la quantité et à la qualité des données utilisées pour l'apprentissage (risques liés au surapprentissage, à l'interpolation…);
- Leur fonctionnement est souvent opaque, alors qu'il est important pour un système de mobilité de comprendre les éléments pris en compte pour la production d'un résultat, notamment pour l'analyse a posteriori en cas d'accident ;
- Leur utilisation est susceptible de générer de nouvelles fragilités (communication, cybersécurité…).

Méthodologie proposée par le projet PRISSMA pour la validation des briques d'IA intégrés dans le véhicules

Les principaux enjeux de la méthodologie proposée sont les suivants :

- Un caractère générique, applicable à tous les systèmes ;
- Une adaptabilité aux différentes fonctions des systèmes considérés et au type de système déployés ;
- Une neutralité vis-à-vis des algorithmes d'IA intégrés ;
- Une capacité à tenir compte des retours d'expérience et des mises à jour des systèmes considérés.

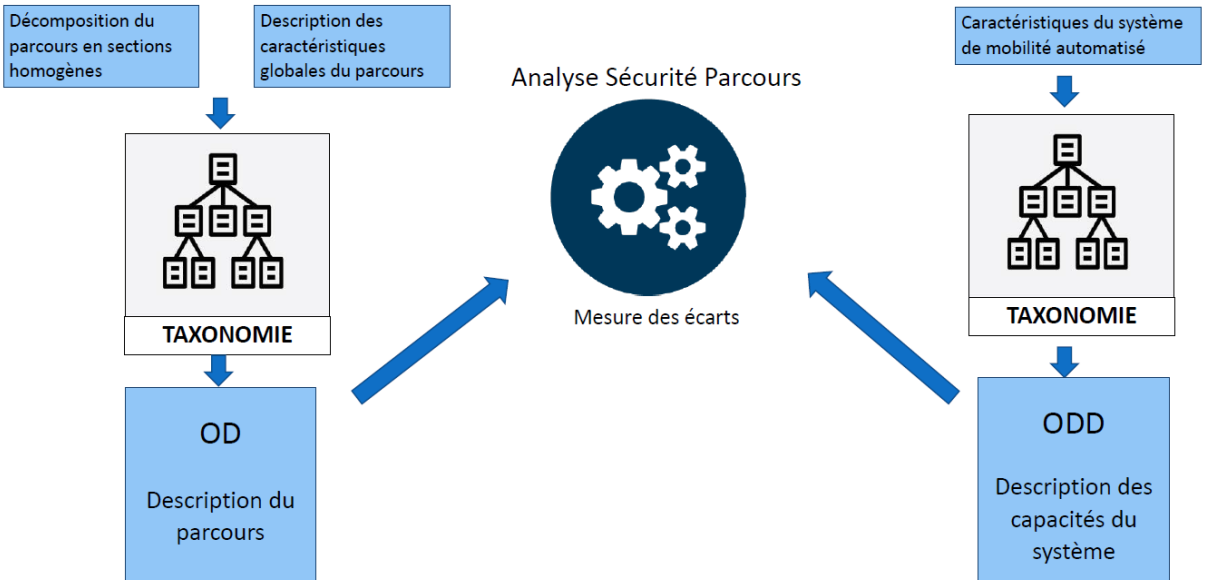
Cette méthodologie se décompose en plusieurs étapes :



1. <u>Analyse de l'environnement réel :</u>

Le fonctionnement en situation réelle est l'objectif des systèmes d'IA intégrés dans les systèmes de mobilité autonome. L'analyse de l'environnement réel s'appuie sur plusieurs éléments :

- L'analyse du parcours, décrit au moyen d'une terminologie dédiée. Cette analyse vise notamment à identifier les caractéristiques importantes du parcours pour un système d'IA puis à décomposer le parcours en différentes sections unitaires ;
- La caractérisation du système de mobilité utilisé pour réaliser le parcours. Cette caractérisation comprend notamment la liste des systèmes d'IA intégrés dans le système de mobilité, ainsi que leur fonction et les contraintes spécifiques qui leur sont associées ;
- Le recensement des principaux risques de communication et de cybersécurité.

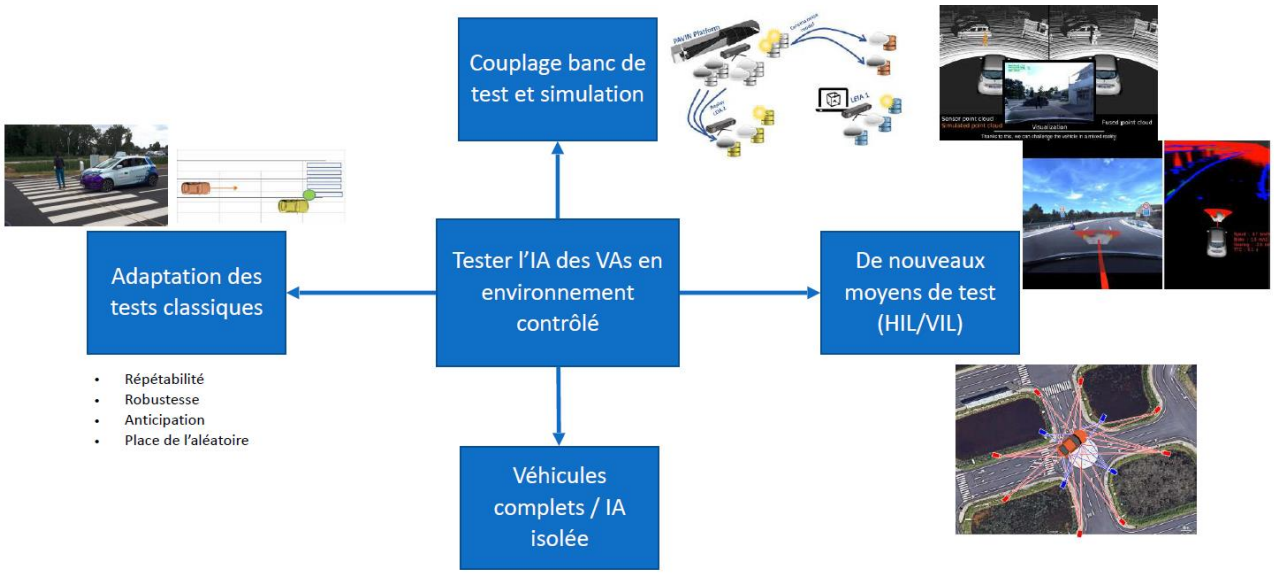


2. Définition d'environnements contrôlés :

La seconde étape de la méthodologie proposée par PRISSMA est la définition d'environnements contrôlés. Les éléments de cette seconde étape sont les suivants :

- Contrôle des caractéristiques spécifiques des systèmes d'IA (à l'échelle composant, sous-système, système de mobilité) :
  - Répétabilité et précision : capacité du système d'IA à apporter, dans les mêmes conditions, une réponse similaire ;
  - Robustesse et résilience : capacité du système d'IA à maintenir sa conformité à des exigences de performance et/ou de sécurité en présence de données d'entrée extérieures à son domaine d'emploi (par exemple en raison d'un défaut sur un capteur ou cyberattaque) ;
- Définition de configurations de référence pour la simulation de complexités croissantes :
  - Véhicule seul ;
  - Véhicule en interaction avec quelques acteurs ;
  - Véhicule en interaction avec l'ensemble des acteurs potentiels (notamment l'infrastructure) ;
  - Roulage en configurations nominales et dégradées.

L'intégration de systèmes d'IA à des systèmes de mobilité demande de nouveaux tests :



3. Passage en environnement simulé :

A partir des résultats des deux étapes précédentes, il est donc possible de construire des environnements de test simulé qui amplifient les possibilités de tests des systèmes d'IA intégrés dans les systèmes de mobilité :

- Construction et mise-à-jour de modèles virtuels d'environnements réels, dits « jumeaux numériques » ;
- Recensement des modèles numériques intégrés aux systèmes, notamment les systèmes d'IA ;
- Analyse du comportement du système de mobilité (capteurs, prise de décision) en fonction de l'environnement simulé et du comportement simulé des acteurs en interaction avec le véhicule ;
- Exploration de l'ensemble des scénarios possibles à partir des modèles validés et identification de scénarios critiques représentatifs (ex. situations accidentogènes, quasi-accident…).

4. Retour en environnement réel :

Le retour en environnement réel constitue le niveau de validation le plus élevé. Le système de mobilité intégrant des systèmes d'IA est soumis à des tests supervisés dans l'environnement correspondant à sa mission et intégrant différentes configurations :

- Interactions réelles avec les autres usagers ;
- Reproduction de scénarios critiques représentatifs.

La supervision permet d'observer les comportements du système de mobilité et donc de valider son bon fonctionnement notamment vis-à-vis de scénarios critiques.

Synthèse des enseignements du projet PRISSMA

L'objectif de PRISSMA était de délivrer une méthode de validation des systèmes de mobilité intégrant des briques IA. Les principales conclusions concernant l'impact de la présence de systèmes d'IA sur les processus de validation sont les suivantes :

- La nécessité de l'introduction de nouveaux critères de validation permettant de répondre aux enjeux des systèmes d'IA :
  - Qualité et gestion des jeux de données utilisés pour les tests, afin notamment de contrôler les biais liés au surapprentissage et donc de permettre une utilisation fiable du système d'IA ;
  - Intégration des notions de répétabilité des résultats et de robustesse.
- La complémentarité entre audit et tests (en environnements simulé, contrôlé et réel) pour évaluer les risques associés aux systèmes d'IA ;
- L'enregistrement automatique des événements lors du fonctionnement du système, la détection et la remontée des situations non-prévues, la recherche de l'amélioration continue et la non régression ;
- La mise en place d'interfaces homme-machine appropriées et le transfert suffisant d'informations à l'utilisateur afin de répondre à l'enjeu d'opacité du fonctionnement des systèmes d'IA ;
- Une formulation probabiliste de la validation (en moyenne / avec forte probabilité, intégration des incertitudes...).