PRISSMA Project
Plateforme de Recherche et d'Investissement pour la Sûreté
et la Sécurité de la Mobilité Autonome
04/2021 - 04/2024

[L3.2] HOMOLOGATION TESTS PROTOCOLS & STRATEGY

Global author: R. REGNIER

Main authors (UTAC PART): A. PIPERNO, J. BOURET, M. DESSAINT, C. SERBOUH

Main authors (CEREMA/LNE PART): C. SEGONNE, P. DUTHON, A. KALOUGUINE, M. BOUDALI

Main author (INRIA/TRANSPOLIS PART): P. KOCH

Main authors (VALEO/IGN PART): R. DENIS, F. WILLIAMS, Y. TAHIROU, A. SEBA, P. BOUQUET

Main author (SPHEREA): C. GAVA

Keywords: AI based vehicles homologation, repeatability tests, robustness tests, AI overfitting tests, anticipation tests, closed road tests, test bench, Simulation, evaluation, AI systems, validation methods, verification process, evaluation protocol, scenario generation, metrics, KPI, performance indicators, simulation tools, sensor models, traffic simulation, graphical and physic engines, simulation environments, proof-of-concept

Abstract. This document describes the intermediate state of the implementation of proofs-of-concept (POC) that aim at demonstrating the use of simulation tests during the homologation and certification processes of autonomous vehicles. Several POC are currently being developed within the PRISSMA project and their particular ongoing work is presented separately.

Résumé. Ce document décrit l'état intermédiaire de la mise en œuvre des preuves de concept (POC) qui visent à démontrer l'utilisation des tests de simulation lors des processus d'homologation et de certification des véhicules autonomes. Plusieurs POC sont actuellement en cours de développement dans le cadre du projet PRISSMA et leurs travaux particuliers en cours sont présentés séparément.

SUMMARY

Chapter 1: UTAC POC

Context (EN)

AI based vehicles could have some safety weak points regarding for example repeatability, robustness, anticipation and overfitting for official known tests. So UTAC PRISSMA WP3 team has built first answers and proposals to adapt or to create new homologation tests scenarios / protocols / testing tools / evaluation metrics for the first WP3 POC tests in UTAC coming soon in February & March 2023. The second WP3 POC tests planed beginning of 2024 will bring confirmation and fine-tuning & modification of them.

Our Inputs are deliverables of PRISSMA WP1 (particularly L1.4), WP2 & WP4 & WP6 (particularly scenarios for virtual/physical/open-road tests), WP8 (regulation/standards first works). We also preliminary made a review of available vehicles with intelligent & predictive ADAS functionalities, and made a bibliography/state of the art of research works & papers related to « tests for AI & AI for tests » and to AI evaluation tools & metrics in the critical industries (planes, trains...).

UTAC WP3 first proposal is to test three vehicles (VW Golf 8 with predictive ACC, ZOE NEXYAD « MotorONE » research prototype with AI based anticipation driving, VALEO Drive4you delivery robot). There will be three categories of new tests (repeatability & robustness, anticipation, overfitting), with some existing or new scenarios (standing pedestrian, hidden crossing pedestrian, strong curve/intersection…) & existing or adapted testing tools (new various pedestrian dummies…), with new evaluation metrics (measure of performance of Automated emergency breaking but also of anticipation and no-use of emergency maneuvers like AEB,..).

Contexte (FR)

Les véhicules à base d'IA pourraient avoir des faiblesses et des risques sécuritaires relativement aux points faibles de l'IA : répétabilité, robustesse aux limites, anticipation, surapprentissage des essais officiels d'homologation. C'est pourquoi l'équipe UTAC du WP3 de PRISSMA a construit des premières réponses et propositions pour adapter les essais d'homologation véhicules ou les compléter par de nouveaux scenarios / protocoles / outils / métriques d'évaluation d'essais véhicules pour les premiers essais qui arriveront en février et mars 2023 pour le POC UTAC. Le second POC UTAC WP3 est planifié début 2024 et permettra de confirmer & améliorer ces réponses et solutions.

Nos données d'entrée sont les livrables du WP1 de PRISSMA (en particulier L1.4), des WP2, WP4, WP6 (en particulier les scénarios pour les essais virtuels/physiques/sur routes ouvertes), et du WP8 (en particulier les premiers travaux pour réglementer et standardiser l'IA dans l'automobile). Nous avons également fait une revue des véhicules intelligents disponibles pour les essais et les POC, et une bibliographie/état de l'art des travaux et articles de recherche relatifs à l'évaluation de l'IA, outils et métriques, en particulier pour les industries critiques (avions, trains...).

La proposition résultante est de tester 3 véhicules (VW Golf 8 et son ACC intelligent prédictif, la ZOE prototype de la start-up française NEXYAD et sa conduite anticipative à base d'IA, et le robot livreur Drive4you de VALEO) , avec 3 catégories de nouveaux essais (répétabilité et robustesse, anticipation, sur-apprentissage) , avec des scénarios existants ou nouveaux (piéton immobile, piéton caché qui traverse, courbe/intersection sévère, ..), des outils existants ou adaptés d'essais (mannequins varies de piéton,…) et de Nouvelles métriques (mesure de la performance de l'arrêt d'urgence véhicule (AEB) mais aussi de son anticipation pour éviter ces manœuvres d'urgence).

1. Adaptation of tests for approval and inputs of others PRISSMA WPs

The objective of WP3 UTAC is to prepare the adaptation of approval tests for AI-based vehicles.

At the beginning of the project, in the second half of 2021, we prepared this WP3 POC outlining what tests and methods could be used to approve vehicles with AI. The question was also what kind of AI could happen and in which systems and functions?

We therefore exchanged with the main experts and leaders of PRISSMA WPs: Rémi Régnier and Guillaume Avrin for LNE (WP1, 2, 3), Emmanuel Arbaretier for APSYS (WP6, 7), Dominique Gruyer for UGE (WP 2), Bertrand Leroy for VEDECOM (WP1) and Paul Guillemard for CEREMA (WP4).

The following are the main trends:

AI will arrive gradually in all vehicle functions, first perception, then route planning, trajectory, and control, Driver Monitoring, IHM, maneuvers like automated minimum risk maneuvers (MRM).

The first pre-regulatory work of the GRVA automotive regulation group concludes that AI is necessary for automated vehicles because human driving behavior and best practices are not precise/quantitative requirements, not programmable for an automate, but can be learned by AI system. It could arrive on premium automated vehicles in 3 years."

Experts do not see on-board « live » learning in vehicles in the short/medium term, as this would lead to changes in vehicle behavior that are impossible to validate. OEM process is to validate and freeze a software for a certain time, generally one or three years.

An IA system is not deterministic, does not meet a specification; It is a black box that can only be validated statistically:

- On potentially dangerous scenarios
- In relation to requirements/criteria/metrics which remain to be defined (data and learning, development, outcomes and safety).

There is therefore a need for new metrics and new scenarios;

A catalogue of critical scenarios will be known/learned by AI! Moreover, it will not offer rare scenarios for validations.

The AI only masters what it has learned (Operational Design Domain (ODD)), so we will need tests of robustness (edge case), very numerous & expensive therefore if possible virtual.

To be able to perform these virtual tests, the OEM models of sensors, fusion, vehicle decision, vehicle control, actuators commands…will be required and also a huge computing capacity. Hence, show the current projects and attempts to communalize the means of simulation by subcontracting them, and opportunity for the regulations to require that the OEM model be provided for type approval and also the data used for the AI training. On the other hand, at minimum the type approval Technical Service could provide the OEM secret randomized scenarios (corner case defined with the OEM at the limit of its ODD), for OEM testing them in SIL-HIL-VIL.

The LNE clearly sees the approval of components made by LNE and the AI-based vehicle approval made by UTAC.

A predictive model will be needed if no OEM model is available, for three objectives

- for many simulations for safety virtual verifications,
- for some approval physicals tests of robustness verification (identify the edge cases to be tested) and verification of the correlation of tests/simulation,
- for explanatory-interpretability (understand-explain the black box).

So a lot of testing will be needed to build a simplified predictive model by predictive modelling.

For the approval, an audit will also be necessary (of the database and learning, validations,..).

Both AI will be approved and each AI-based vehicle (according to AI act).
One example of this is the Cyber Security and SW/OTA double approval process.

The big problem with AI right now is perception, very hard to work out. It is also very difficult to specify an ODD in perception (examples: objects, sunset truck, pedestrian morphology, weather characteristics). Therefore, it is very difficult to make a specification of perception and to validate the perception function (OEM needs & type approval).

Therefore, it is very difficult to assess the reliability of a perception subsystem (allocation of requirements for reliability, which is necessary for the safety of operation).

IA will arrive in a few years into the vehicles decision systems, because on-screen learning over thousands of kilometers of filmed driving becomes possible.

For Predictive and explanatory models, OEMs needs it for validations and are in much better position than the Technical Approval Service to have or build them.

The Technical Service of approval must however be competent (as in Safety or cyber), to be able to decide type approval, and also to offer these skills to small OEMs (via the projects and programs French or European as TEF).

These first principles, dating from 2021, were then supplemented by the work and deliverables of the PRISSMA WPs in 2022, which must also be taken into account when defining the approval tests.

- 1.1 Adaptation of tests according to vehicle ODD and first WP8 inputs

According to the requirements of all autonomous vehicle regulations (ALKS, ADS, draft of the "Arrêté français autonomous urban shuttles"), the OEM will have to declare to the customers and to the type approval authority its ODD (Operational Design Domain).

For example the speed : an OEM will declare its autonomous driving functionality is safe and operational for speeds of not more than a certain value of speed, for example 30 km/h.

This constant of the declaration of ODD is therefore an important input for the approval tests of automated vehicles: these limits are the limits on which the AI based vehicle will be tested, verified and approved.

As seen in the chapters below, while remaining within the budget and time constraints of WP3 PRISSMA, we have tried to find vehicles with different driving systems with AI and different ODD, as varied as possible, in order to solidly build our proposals to adapt type approval tests for all kind of AI-based vehicles.

- 1.2 Adaptation of tests according to OEM homologation safety audit and first WP6 inputs

PRISSMA WP6 aims to construct and adapt the safety audit of the vehicle type approval.

There is consensus on PRISSMA that the approval process of an AI based vehicle should begin with this functional safety audit, which will provide first inputs and themes and priorities for the approval tests & verifications.

We summarize this by the **diagram below**, which was one of the conclusions of the WP3 + WP4 meeting of 16/09/2022:

WP6 audit (vehicle safety weak points / validations, ODD & vehicle limits)

⇩

WP2 virtual approval tests for dangerous/complex scenarios

WP3 physical approval tests for ODD limits scenarios & critical scenarios

⇩

WP4 physical approval tests on real open roads for real tests & verifications

- 1.3 Adaptation of tests according to needs/complementarity with virtual tests/open road tests (WP1 & WP2 & WP4 inputs).

In October 2022, the work of WP1 and WP2 proposed a first distribution of the scenarios to be test, virtually or physically, depending on several dimensions:

- The first dimension is the hazardous, feasible, or expensive nature of physical testing, which simulation enables to avoid.
- The second dimension is the digital model availability for virtual testing: will the manufacturer provide executable software models?
- The third dimension is the time available for approval, as virtual tests sometimes require more time of preparation than closed track tests.

- 1.4 Inputs from WP1 for methods and metrics to evaluate IA repeatability, robustness, overfitting.

WP1 and in particular its deliverable L1.4 of October 2022 aim to provide an overview of the state of the art and recommendations on methods and metrics to evaluate systems based on AI.

The pages 63-65 of the PRISSMA deliverable L1.4, **reproduced below**, asks three questions for the validation tests:

1. How trustworthy are the uncertainty estimates of our model under perturbations?
2. How robust are the predictions of our model under perturbations?
3. How do uncertainty and accuracy of different methods co-vary under perturbations?

Concretely, we previously described corruptions and perturbations proposed, and ideally would like the model predictions to become more uncertain with increased shift, as far as shift degrades accuracy. This is usually called “covariate shift. Hereafter, we start by selecting a subset of perturbations, following state of the art results, allowing model evaluation and validation with reduced cost. Next, we explain decision process.

1. Data perturbations

(a) Data-set shift: We propose the following shift for autonomous driving system:

- Time of day / Lighting
- Geographical location (City vs suburban)
- Changing conditions (Weather / Construction)

They may be simulated using domain adaptation technique [35] that has emerged as a new learning technique to address the lack of massive amounts of labeled data by using labeled data in one or more relevant source domains to execute new tasks in a target domain. In our context, we propose the following validation condition.

(b) Adversarial perturbations

(c) General corruptions

(d) OOD samples

2. Robustness validation: In general, there are two different approaches one can take to evaluate the robustness of a neural network: attempt to prove a lower bound, or construct attacks that demonstrate an upper bound. The former approach, while sound, is substantially more difficult to implement in practice, and all attempts have required approximations.

On the other hand, attacks used in the latter approach are not sufficiently strong and fail often; the upper bound may not be useful. Moreover, as seen before, there exist different types of adversarial attacks and defenses for machine learning algorithms, which makes assessing the robustness of an algorithm a laborious task. Thus, there is an intrinsic bias in these adversarial attacks and defenses to make to further complicate matters.
For instance, an evaluation process must avoid a model dependence behavior, insufficient evaluation, a perturbation dependent result. This requires a model agnostic adversarial robustness assessment. In [36], authors have recently observed that dual synchronized attacks based on $L0$ and $L\infty$ distance-norms allow a good robustness assessment on several neural network architectures. Moreover, their results suggest that $L1$ and $L2$ metrics alone are not sufficient to avoid spurious adversarial samples and it is better to combine dual norms (1 and ∞) to construct an upper bound on the robustness of the model.

3. Uncertainty validation: naturally, we expect the accuracy of a model to degrade as it predicts on increasingly shifted data, and ideally, this reduction in accuracy would coincide with increased forecaster entropy. A model that was well calibrated on the training and validation distributions would ideally remain so on shifted data. On the completely OOD data, one would expect the predictive distributions to be of high entropy. Essentially, we would like the predictions to indicate that a model "knows what it does not know" due to the inputs straying away from the training data distribution.

First Recommendation of 'perturbation in black box' robustness tests

To evaluate the robustness of a system with IA, this deliverable L1.4 recommends on page 64 of "adversarial attacks in black box", in order to see what the reaction of the system is and then gradually to adjust the attack to the system

These attacks (or perturbations) with misleading data/configurations and at the limit of the system ODD are possible in the machine learning phase of an AI but also in the operational phase; For very famous example, AI based sensor vision were attacked by road signalization panels with little black rectangles.

The work of WP5 (cybersecurity) aims to protect the database for AI learning because to know this database is very helpful to attack it in operational phase.

Attacks (or disturbances) of corruption are also recommended, that is with data/configurations for use unavoidable & normal but misleading because at AI limits, like weather limits (fog, snow, cold, vibrations or movements decreasing image quality). The AI based system can be weak on these limit conditions because it made very learning on them.

But, according to the meeting with the AI expert and leader of WP1, Rémi Regnier, on 7/9/2022, these attacks are easy and relevant in virtual tests but difficult and expensive in physical tests because a step-by-step process is necessary to identify the system limits.

Second Recommendation of robustness tests:

This type of testing seems to be very suitable for closed-track testing by UTAC, according to the same meeting with the AI expert and leader of WP1, Rémi Regnier.

Third Recommendation for uncertainty testing (repeatable/stable or chaotic system)

Again, these tests seem suitable for closed track UTAC tests, according to the meeting with AI expert and WP1 leader Rémi Regnier.

These tests assess the uncertainties of the system due to the different dispersions/ margin of error of its components (sensors, position & RTK ...) and the propagation of uncertainty in the neuronal networks of the AI system.

The system will be validated repeatable and stable if for very close inputs, there are very close results, otherwise it will be labeled chaotic.

2.Review of AI-based vehicles/driving functions available for POC and tests

2.1 Meetings/discussions with the experts of UTAC and PRISSMA, with NEXYAD, VALEO and MILLA OEMs, and resulting choice for vehicles & planning of UTAC POC and tests

Throughout the second half of 2021, meetings were held with UTAC and PRISSMA experts. The exchanges with the UTAC experts made it possible to have the inputs and first visions of new intelligent ADAS functions and of regulatory and consumerist works (Euro NCAP).

The discussions with PRISSMA experts provided a technological vision of progress of AI and of AI possible applications for automotive industry, they also provide options and ideas of solutions to evaluate and type approve AI based vehicles.

UTAC Expert Vision: New Intelligent Speed Control Functions

As it is often the case, Euro NCAP is the precursor and incentive for new driving intelligent functions that will improve safety. These new functions, called Speed Limit Information Functions (SLIF) and Speed Limit Control Functions (SLC), do arrive in the future Euro NCAP safety assessments, which are still unofficial and are being discussed in the Euro NCAP WGs (in which UTAC participates).

The Euro NCAP will gradually introduce bonus points in its vehicle evaluations if such functions of driving can manage (with an alert to the driver or with an automated speed reduction) the situations in the **figure below**, called features:

Systems that are able to properly identify road features where a speed, lower than the legal speed limit, is more appropriate and/or advised or the vehicle should come to a stop can attract points based on the number of road features. These road features are listed in the table below and example traffic signs of a limited number of countries are specified in Appendix I.

ROAD FEATURES	Points	Required Action
Curves*	2	Show and start reducing to appropriate speed
Roundabouts*	2	Show and start reducing to appropriate speed
Junctions*	1	Show and start reducing to appropriate speed
Traffic Lights	2	Warning only
Stop Signs	1	Warning only
Yield Signs	1	Warning only
No Entry	1	Warning only
TOTAL	**10**	

* Only eligible for scoring when linked to ISA and/or i-ACC

Euro NCAP is well aware of nature/numbers/root causes of road accidents in the main European countries and is convinced on the well-known fact (and widely shared by the French authorities in charge of road safety, DSR and ONISER) that excessive speed is the main cause of road accidents.

The Working Group 'France rating level3 Euro NCAP', led by UTAC and attended by French manufacturers (Stellantis, Renault, Valeo) has the same vision and confirms (**see figure below**) that the best automated vehicle is the one that avoids emergency maneuvers through the use of intelligence and anticipation, as the human driver know to drive:

At the LNE Forum for Evaluation of AI on 24 November 2021, during the round table UTAC confirms this vision with the following example: « an intelligent vehicle should not have to choose between crashing an old-women crossing or a baby running on the road, it should be able to anticipate and to avoid this critical situation ».

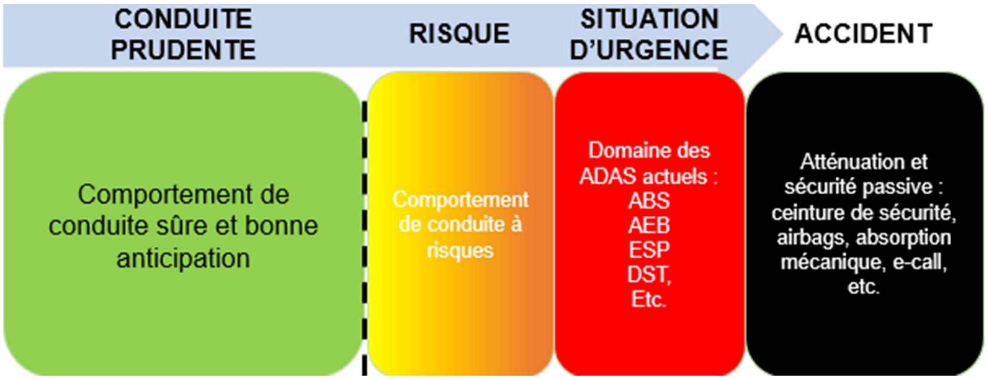
The proactive ACC and the 'safe speed' developed by the French start-up NEXYAD

Among all the manufacturers contacted for the UTAC tests of the POC of PRISSMA WP3, NEXYAD is clearly the most skilled about anticipation functions, working for more than 10 years on AI based intelligent automated driving functions.

NEXYAD has gained experience since the 2000s through 12 collaborative research programs with road safety and infrastructure experts from 19 countries, and NEXYAD has developed a new driving functionality that estimates road risk and therefore adapted and relevant safe speed. This relevant speed can be lower than authorized speed limit!

NEXYAD calculates in real time the level of risk of the situation, according to the context (infrastructure, traffic, presence of vulnerable persons...), according to speed, configuration of the road, signaling, visibility, proximity... Also, according to more than 5000 rules built by IA from a database of road accidents built during its experiment on 12 collaborative research programs with road safety and infrastructure experts during 10 years and in 19 countries.

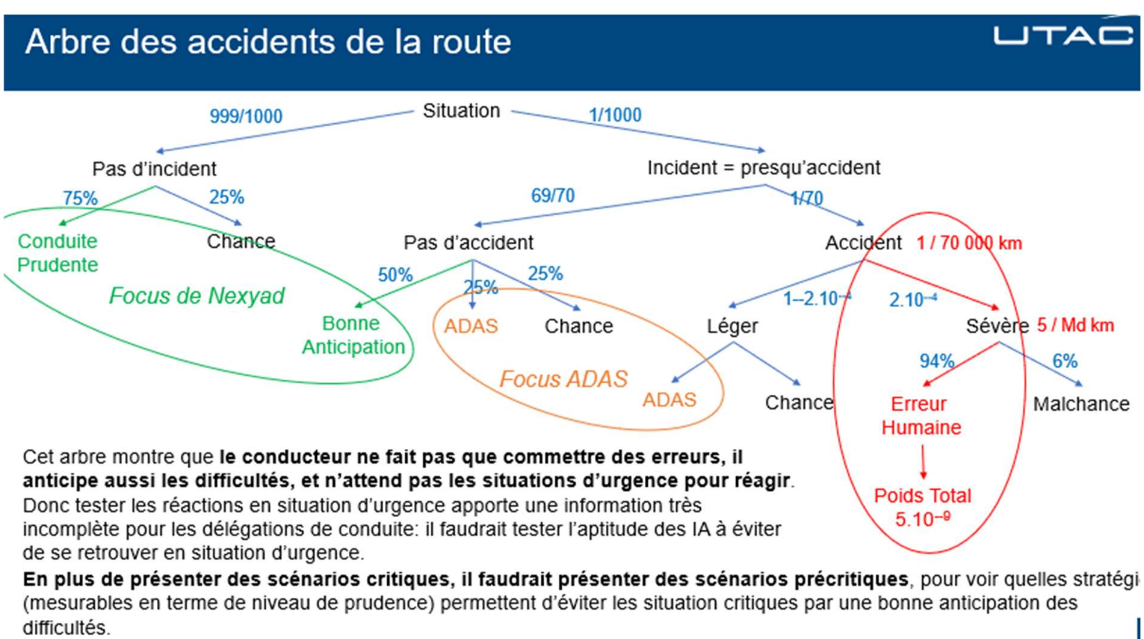
The level of risk calculated by NEXYAD, called safety score, is illustrated with the **figure below**, and the commonsense principle that « to drive safe with a low risk, you have to stay in green situation and not be close to red high risk situation »:



NEXYAD also agrees with the principle and consensus (Euro NCAP, French authorities DSR and ONISER), that speed is the main cause of road accidents.

NEXYAD has studied accidents extensively (over 10 years and in 19 countries) and estimates that the prudent driving and anticipation, at the very beginning of the chain, represent 99.9% of the behaviors observed. Risk behaviors that do not have consequences are absent from the statistics, but they do exist and sometimes lead to emergencies, which are rare but focus all the attention of manufacturers and OEMs. Fatal accidents are even rarer (around five deaths per billion km in the OECD).

Thus, NEXYAD is very advanced in accident analysis and is almost the only one to have a numerical analysis (in probability) not of accidents but of the "near-accidents". On the **figure below**, the road accidents and near-accidents tree, in which NEXYAD estimates that for one accident there have been 69 near-accidents, which is potential accidents that have been avoided through good driver reactions. NEXYAD used these 69 potential accidents to build its anticipative system, which have to anticipate them, like a good and prudent human driver:



Two new innovative and intelligent proactive functions developed by NEXYAD:

Thus NEXYAD has developed (and patented) two new functions of intelligent driving, and is in discussions with many French, German and Japanese manufacturers to market them: these driving functions use the estimation of the risk of NEXYAD and the consequently relevant safe speed to minimize the risk and to stay in the green zone of driving risks (previous figure).

The risk is estimated according to the road map (arrival on a steep curve, a tight crossroads...) and also what the vehicle sensors see (vehicle poorly parked, crowded crossroads, low field of vision

NEXYAD's 2 innovative and intelligent proactive driving functions are:

- **A safety assistant (named "safety coach") who alerts** the driver when his driving behavior is no longer prudent (risk too high) in relation to the driving context (accident reduction estimated by NEXYAD of at least 25%).
- **An intelligent and proactive ACC** that automatically regulates the vehicle speed according to the driving context (up to 75% accident reduction according to NEXYAD)
- NB the difference in the result between 25% reduction of accident in alert mode compared to 75% of the intelligent ACC mode is explained by the fact that the driver may not immediately and always take into account the warnings and not slow down.

These two new functions are being implemented on a prototype vehicle, the NEXYAD DREAMOTOR1, **see photos below**, which is therefore one of the most advanced prototypes in the world (On PRISSMA there is no French actor among vehicle manufacturers and it's difficult to know their skills and developments on these very upstream and very competitive subjects). NEXYAD is part of the French industrial research and development group of the Regions Normandy/Ile de France, called NEXTMOVE (previously MOVEO), which supported and facilitated these innovative projects.

This prototype vehicle was rapidly tested in September 2022 by the team PRISSMA of UTAC, on a one-hour circuit, allowing identifying interesting scenarios for PRISSMA tests in 2023; This prototype should be operational in March or April 2023, which could be just in time for the PRISSMA 2023 UTAC POC & tests.

We have regular meetings with NEXYAD about this planning of WP3 POC & tests, and this will be clarified in February: either NEXYAD prototype is ready for WP3 POC 2023 tests, or we have to find another interesting vehicle. To rent an up-to-date level 2 commercializes vehicle like a Tesla or a Mercedes is also a good option.

Exchanges with MILLA

In mid-2021, exchanges took place with the French manufacturer MILLA, which presents the interest of having developed a delivery robot, **picture below,** which is the ideal use case for POC PRISSMA. However, these exchanges have not been very extensive and have no way to achieve the integration of robot-delivery to WP3 tests of UTAC. MILLA is involved in many projects French and European research, and on our side The PRISSMA WP3 test budget is not unlimited and we had to aim 2 to 3 vehicles Maximum for our trials.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

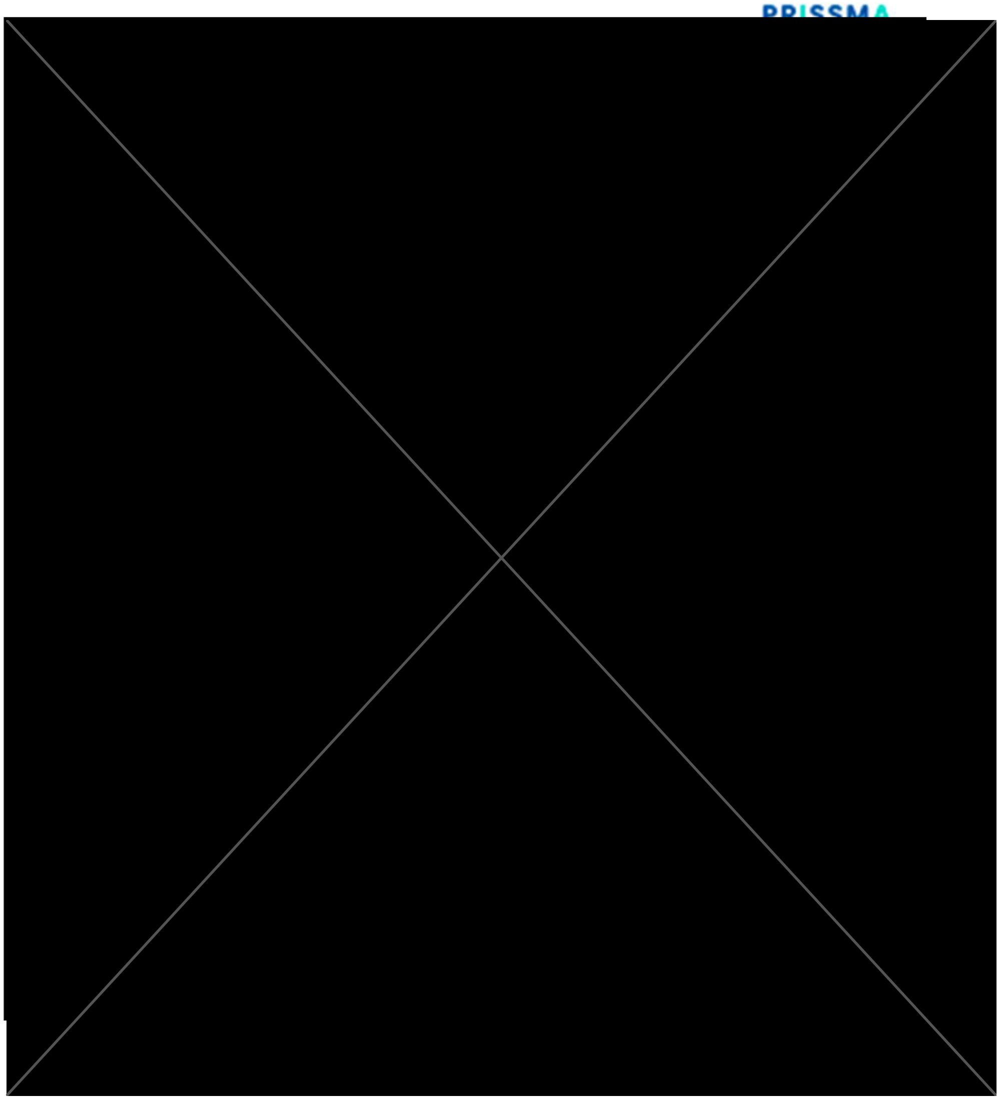
[redacted]

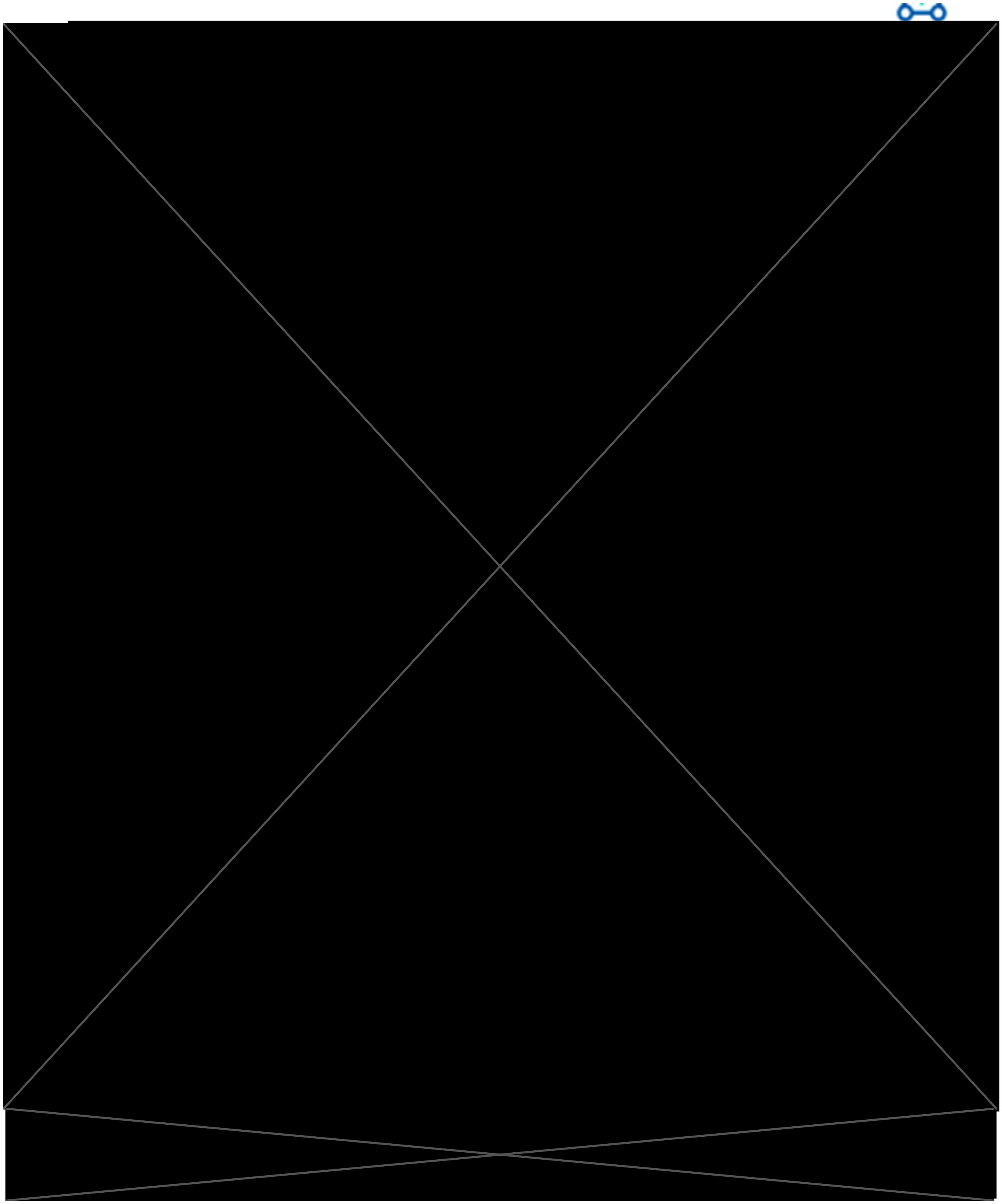
[redacted]

[redacted]

[redacted]

[redacted]





[redacted]

Other vehicles available and testable for UTAC PRISSMA WP3 tests & POC:

According to that the objective of WP3 part UTAC is to prepare for the adaptation of the approval test to AI based vehicles, our needs are ideally to make these tests on different vehicles with a maximum of AI on board, not only like today on camera sensors.

Rent an ultra and autonomous level 3 vehicle like **the New Mercedes Class S**, see **picture below,** recently approved for German motorways is therefore an interesting opportunity.

However, the vehicle has not yet been approved for the French motorways and no cooperation exists with the constructor to enable the driving function (ALKS) in France and to have access to the results internal to the vehicles and its computers and functions. Another point is that today the Knowledge of technology content, AI and performance of this new vehicle is very limited. Therefore, this opportunity seems a little premature for the first part in 2023 of the UTAC WP3 tests and will have to be reconsidered for the second part of tests & POC in early 2024.

The opportunity of **two VEDECOM EasyMile autonomous shuttles** has also been studied. These two shuttles (**see figure below)** were experimented on open roads more than 1 year on a circuit linking a bus stop and the VEDECOM site in Versailles-Satory.

This approach has not been taken, because it is from the older generation of shuttles, with few AI available and which have only been validated with the manufacturer only on a few predefined and fixed paths. Therefore, it should also be necessary to benefit from a wide collaboration with the manufacturer before being able to make any test.

We also came to the same negative conclusion for the opportunities of testing:

- the ‘old’ **shuttle ARMA NAVYA from the UGE**, proposed by UGE for UTAC tests

- **the autonomous Renault ZOE of INRIA** in Grenoble, which is technically very interesting but not available because already used for WP3 TRANSPOLIS POC in Lyon.

Other intelligent and predictive ADAS are on the market, particularly among German manufacturers, and we have studied them, and quickly tested them (in next chapter 2.2 for the VW GOLF8 predictive ACC).

Most VW vehicles (golf 8, Arteon 2017, ID3, New Polo) have effectively a predictive ACC, capable to read (with on-board cameras) agglomeration entrance/exit speed limitation signs and (with road HD maps) strong curves (of roads, roundabouts,…), and also capable to automatically adapt its speed through the ACC function.

This is a basic predictive driving feature, without AI and machine learning nor driving risk evaluation as NEXYAD proposes.

Here below is a good summary of this functionality of the Golf 8, found in Volkswagen documentation:

> **ACC with predictive speed detection**
>
> The latest generation of Automatic Cruise Control (ACC) works in advance on the Golf. The system calculates the Golf's position using route data and the navigation system's GPS, enabling it to anticipate and reduce speed before bends, roundabouts, junctions, speed limits and urban areas. At the same time, 'ACC' uses the front camera's recognition of traffic signs and regulates speed as soon as a speed limit is detected. The most advanced version of 'ACC' also incorporates a traffic jam assistant.

As well as **here below** on SEAT vehicles (which is a Volkswagen Group brand):

> **Adaptive Cruise Control (ACC)**
>
> Until now, the ACC system has adapted the car's speed to that of the vehicles in front, thanks to front radar. With the New SEAT Leon, this system equips the vehicle with new predictive elements that allow the driver to adapt his or her driving speed to the road and to the GPS data provided by the navigation system, which also allows the driver to correct the speed according to the road layout, bends, roundabouts, junctions, speed limits or work zones. In addition, using information from the front camera and sign recognition, the system can adjust the vehicle's speed when the limits change.

The Audi brand also offers regulation of Predictive Speed (on A4, Q3, Q7...), **see here below**:

> Predictive control 1) uses map information from the navigation system to adopt anticipatory driving. For traffic sign identification, the system also takes into account the information received from the sign detection identified by camera link. The system brakes autonomously in the event of a speed limit or before changes in the road layout (bends, intersections or roundabouts), then accelerates your vehicle again to reach the memorized speed).

The users of these predictive ACCs testify on the internet to the effectiveness of this function and are satisfied with the flexibility of the speed regulator which automatically detects and adapts the vehicle speed before entering on a roundabout or on a strong curve.

However, there are also many dissatisfied people who say that they no longer use this function (by disabling it) because it regularly generates false alarms (false positives) and sharp slowdowns or even sharp brakes when there is no risk, just because they read speed limitations signs from others close roads or from incorrect roadmaps data:

See example of dissatisfaction here below:

Note that this speed control, which anticipates turns, intersections, areas with limited speed seems also very interesting for the environment and the consumption/emission of the vehicle, announcing about 10% off consumption/emission reduction, **see here below:**

That is also why many trucks have now this function: MERCEDES, DAF Trucks...

The predictive ACC function is also proposed in the after-sales (second assembles), but in this case it is rather the community of users that indicates the zones where it is necessary to slow down.

The French or Japanese manufacturers do not offer any predictive ACC, but announcement could be imminent from STELLANTIS and NEXYAD.

2.2 VW GOLF8 predictive ACC preliminary tests and identification of interesting scenarios

These preliminary tests confirmed that VW Golf 8 and its predictive ACC would be interesting to test in the UTAC POC in order to estimate its performance and define new tests related to repeatability, robustness, anticipation & overfitting verification of intelligent functions.

How predictive ACC (called Travel Assist) works:
This system combines two driver assistance functions, Adaptive Cruise Control (ACC) for longitudinal assist and Lane Assist for lateral assist.

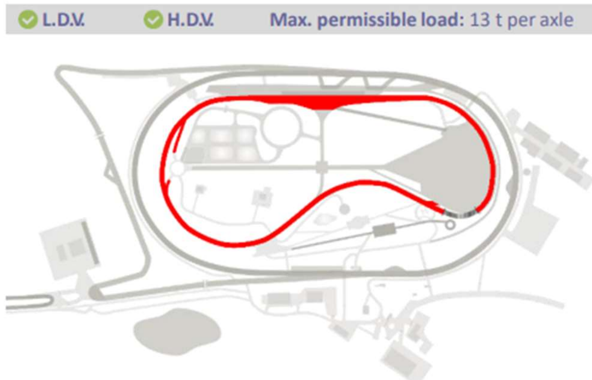
A button on the multifunction steering wheel, which therefore triggers longitudinal speed assist and lateral position assist, activates this function. For safety reasons, the driver must keep his hands on the steering wheel for the guidance to be effective.
In addition to this longitudinal speed guidance, an anticipation function can be added. The system calculates the position of the Golf based on GPS and route data from the navigation system

and must adapt the speed in advance to the approach of bends, roundabouts, crossings, speed limit zones etc. At the same time, it uses the traffic sign recognition system via the front camera and must adapt the speed as soon as a limitation is detected.

Test areas at UTAC:

UTAC tested this function on different test tracks simulating different environments:

TEQMO Highway:

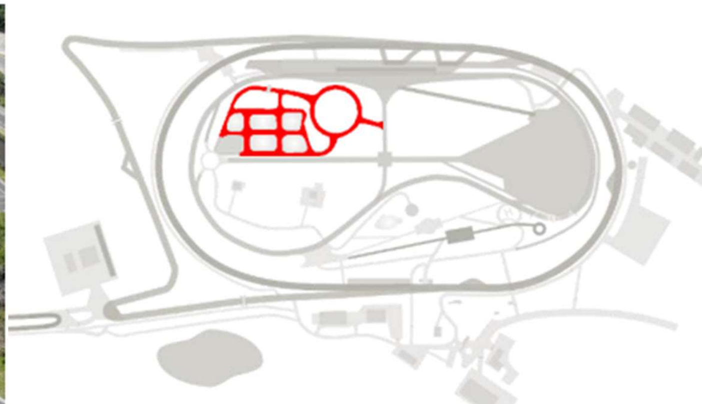


« Highway » circuit

Features
- Length: 2.2 km
- Width: 3.5 m
- 2 and 3 lanes with multiple entries & exits
- Tunnel and tollbooth
- Road marking & signaling

Services
- Platooning
- Adaptive cruise control
- Lane keeping
- Road sign recognition

TEQMO City:

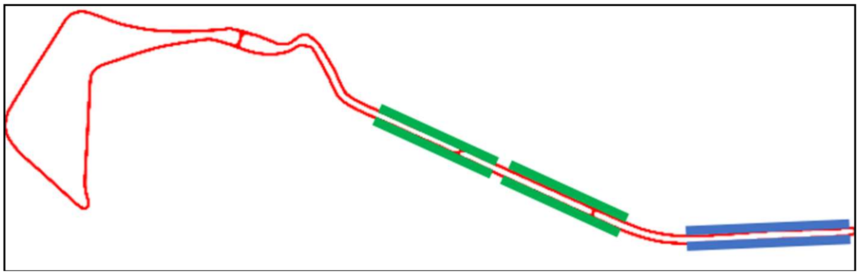


« City » zone

Features
- Area : 38,500 m²
- 2 and 3 lanes 3.5 m wide
- Lighting & signaling
- 3 crossroads incl. 1 with roundabout and 1 with traffic lights, 7 « T » intersections
- 1 roundabout 50 m radius, 1 level crossing

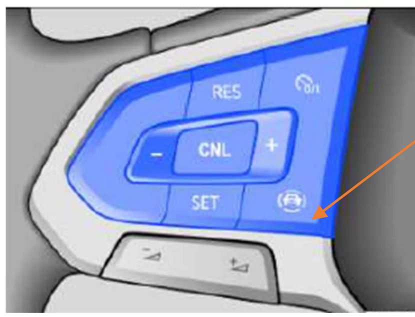
Services
- Autonomous Shuttle
- Various projects
- Bus stations

Road Circuit :



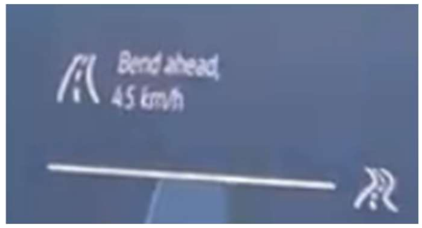
Tests done:

Different runs were carried out on the 3 types of tracks. The "Travel Assist" function is activated by pressing the steering wheel button:



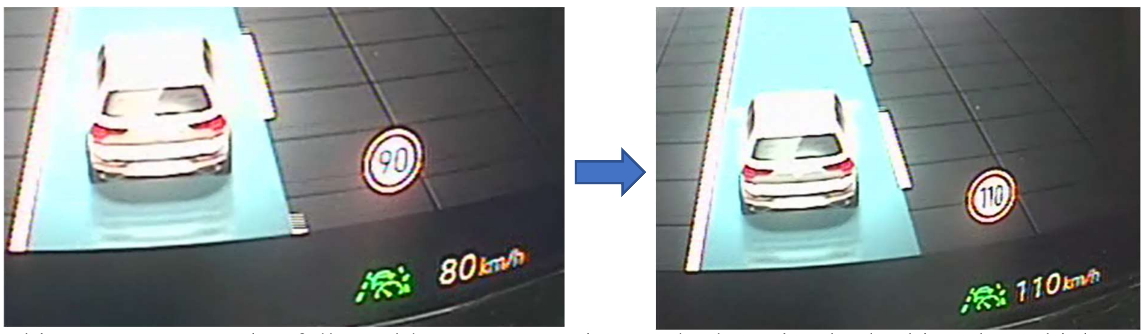
A reasonable speed instruction is given to the system at the start depending on the environment. The vehicle was equipped with a VVBOX to take two synchronized video views (Dashboard and road), the runs have been recorded and stored on a SharePoint.

Runs of several minutes were carried out on the 3 types of tracks with activation of "Travel assist". During driving, the so-called "anticipative" feature could be observed in different places, approaching a bend or a dangerous curve by this type of message on the Dashboard:



« Bend ahead, 45 km/h »

This message was followed by an automatic speed adaptation by braking the vehicle, at the speed recommended by the message.
The adaptive function by reading the speed limit signs could also be observed, for example when passing a traffic sign 110kph:

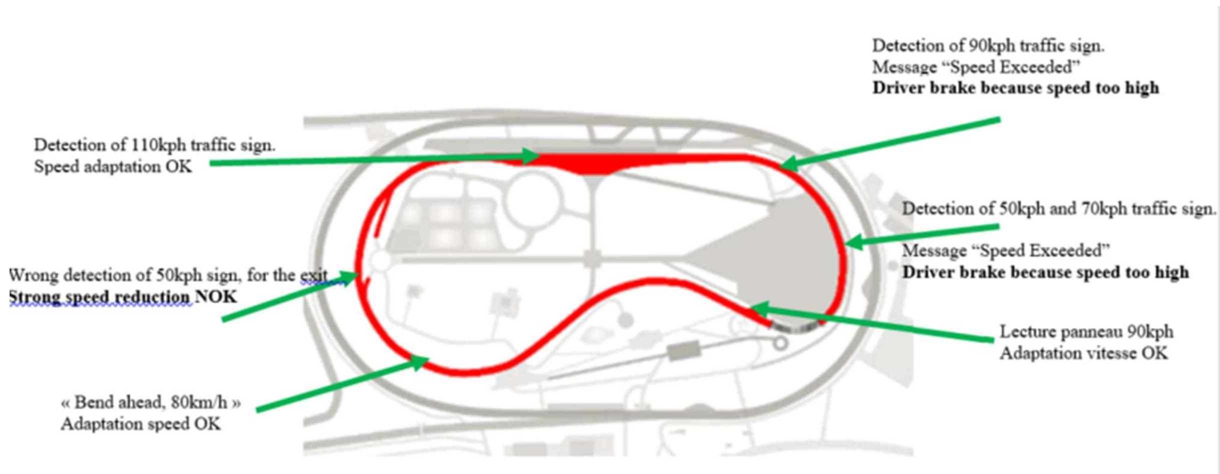


This message was also followed by an automatic speed adaptation by braking the vehicle, at the speed read on the traffic sign.

Observations:

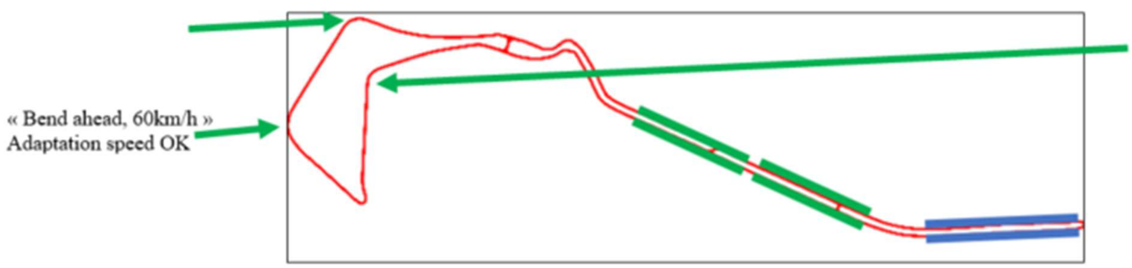
- TEQMO City:
  No anticipation was observed on this track.

- TEQMO Highway:
  Interesting track with several adaptive reactions of the system:



- Road Circuit :

  Some speed adaptations in curves:



Appearance of an additional pictogram when driving on the road circuit, dangerous turn sign before almost every bend.

Conclusion:

The anticipation system proposed in this vehicle showed some interesting reactions on the UTAC tracks.
The tracks being quite specific, the system can have trouble anticipating correctly, a driving in real conditions on open road would be interesting.
The bad anticipation of the Travel Assist system forces the driver to react and press the brake pedal, so the Travel Assist is deactivated. A using of the system on the highway on a long-distance journey should be more relevant and require less driver input. Otherwise, the user could choose not to activate.

3. Bibliographic studies and state of the art for « tests for AI and AI for tests »

Driverless vehicles are one of the technologies that has shown substantial progress in recent years. Driver Assistance Systems (ADAS) like lane-keeping, adaptive cruise control, collision avoidance is becoming more and more efficient and their deployment will reduce accidents and reduce the travel time [1]. This progress is due to the artificial intelligence (AI) embedded in the AVs which represents one of its most important components [2].

It has been shown that various AI approaches provide promising solutions for the development of AVs, which provides a fused process between the data acquired by the vehicle and the results of the implemented AI model as shown in Figure 1.

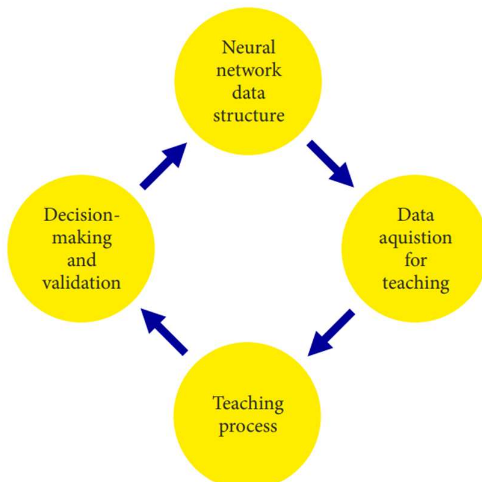


Figure 1 Fusion of artificial intelligence with VA [23]

These approaches have been applied in various applications such as perception [3], motion planning [4], decision making [5] and safety validation [6], [7], [8].

3.1 Use of AI in Avs

- **AI for perception**

  Perception is an action similar to human vision. AVs use sensors to analyze and monitor the environment. The authors of [9] have grouped existing perception methods into two categories:

  - **Mediated perception**

    It develops detailed maps of the AV's environment by analyzing the distances of vehicles from surrounding objects (vehicles, pedestrians, trees, road markings, etc.). In this approach the AV uses AI algorithms like CNN (Convolutional Neural Network) for object detection, the simplest example is the detection of traffic lights which in some cases is more efficient than human, the efficiency of the DNNs (deep neural networks) used reaches 99.46% [10]. In addition to that classification methods have been used, for example [11] proposed a machine learning method for perceptual classification that reached a competitive recognition performance of 99.54% on the reference data of German traffic signs.

- **Direct perception**
  This method does not create a map of the vehicle or a detailed trajectory plan but controls directly the output of the steering angle and Speed of the vehicle with the DNNs proposed by [12]. PilotNet, which is a framework for CNN, was proposed by [13] [14] to train AVs to steer on the road with camera images as input and direction parameters as output.

➢ **AI for localization and mapping**
Most autonomous vehicle manufacturers such as Uber, Google...etc. use mapping methods that involve driving specific roads beforehand and collecting detailed sensor data, such as 3D images and high precision GPS data. The authors of [15] proposed to fuse data from GPS, inertial odometer, and cameras to estimate the trajectory of a vehicle. A two CNN system was used by [16] one for short range object detection (2-25 m) and the other for detection (15-55 m) given the low resolution of the input images then the outputs of the two CNNs were combined to estimate the final range projection.

➢ **AI for decision making**
After receiving the different data, the AVs use them to make decisions and act with the environment with different applications like automatic parking or planning. Researchers in [17] proposed a two-stage random forest-based classifier to assist autonomous systems; this system was validated in the Audi Autonomous Driving Cup, a university level competition. For path planning, an algorithm for clustering obstacle trajectories and optimizing continuous contingency trajectories for VAs has been proposed [18].

Artificial intelligence has propelled the development of autonomous vehicles, but as it is a non-deterministic technology [20] that gives non-repeatable results, another challenge is added to the development of an AV which is the testing of AVs. The traditional homologation methods that aim to ensure the safety of vehicles on the market are no longer applicable to AVs. Especially since a recent study shows that AVs have to travel "hundreds of millions of kilometers and, in some scenarios, hundreds of billions of kilometers to create enough data to clearly demonstrate their safety" [19], so it becomes necessary to design new solutions for testing AVs.

3.2 Challenges of testing AVs

The deployment of an autonomous vehicle system faces several challenges that are summarized by [20]:

➢ **driver out of the loop**
In a fully automated system, the driver does not have to intervene, so the entire responsibility is on the autonomous vehicle [21]. Therefore, there are uncontrollable situations where the human factor cannot establish a safe state so the systems must be designed to a higher Automotive Safety Integrity Level (ASIL) [22].

➢ **complex requirements**
In addition to the exclusion of the human factor, the system must handle the different situations it may face, such as different weather changes, improbable incidents such as non-visible signs, etc., which increases the requirement for AV design.

➢ **non-deterministic algorithms,**

Many machine learning-based algorithms have been implemented on AVs, these models are probabilistic and non-deterministic so they are not repeatable and sometimes provide results that are only correct with a certain probability so testing and validating such systems is a challenge for two reasons:

- The difficulty of exercising a particular boundary case because boundary cases only act if the system receives very specific data as input.
- The difficulty to verify if the test results are correct or not because we have different behaviors due to the non-determinism.

- **inductive learning algorithms**[1]

  There are different approaches to machine learning such as supervised and unsupervised, whatever approach is adopted leads to inductive learning in which the examples used for learning are used to build a model. This makes it difficult to validate these models because the data used for testing must be different from the data used for training in order to detect overfitting.

- **fail-operational systems**

  Failed operational systems are used in fields such as aerospace but it is still difficult to design them. First, it is necessary to ensure redundancy, i.e., if a system fails, it is necessary to ensure the existence of other systems to take over.

As the testing and validation of an AV embedding artificial intelligence faces several challenges it becomes necessary to build robust test systems respecting certain criteria.

3.3 Features of tool for testing AVs

- **Safety**

  It is dangerous to test AVs on public roads, so it is necessary to redirect to simulations or track tests.
- **Efficiency**

  Since the biggest challenge when it comes to testing AVs is to perform tests that produce repeatable results, it is necessary to target the tests in order to be able to detect the maximum number of failures.
- **Coverage**

  It consists of finding all high-probability failure scenarios, or areas of the ODD where performance is below the acceptable level.
- **Black-box interaction**

  AVs are considered as black boxes that preserve the integrity and confidentiality of the implemented system. Therefore, the tests are oriented towards the evaluation of a black box system.
- **Adaptability**

  As vehicle behavior is constantly changing, basing tests on fixed scenarios does not guarantee the proper functioning of the vehicle, so it is necessary to design tests that are adaptive to vehicle behavior.

[1] Inductive learning consists of giving examples of a function in the form of data (x) and the output of the function (f(x)). The objective of inductive learning is to learn the function for new data (x).

- **Unbiasedness**
  Unbiasedness implies that risk estimates are not systematically distorted by past performance or test environment artifacts that deviate from actual operating conditions.
- **Prioritized results**
  To obtain usable information, security testing must classify and prioritize failures in order to target the tests to be performed.

3.4 Existing intelligence testing approaches

There are different approaches to test an autonomous vehicle. The authors of [19] have grouped them into two categories:

3.4.1 Scenario-based testing

When a vehicle is deployed in public roads, it can face different and sometimes not usual situations, so it must be able to face them by exploiting well the data received from its environment to take the right decisions to interact with the environment. Therefore, it is necessary to test a vehicle in specific and unusual scenarios as shown in figure 2.

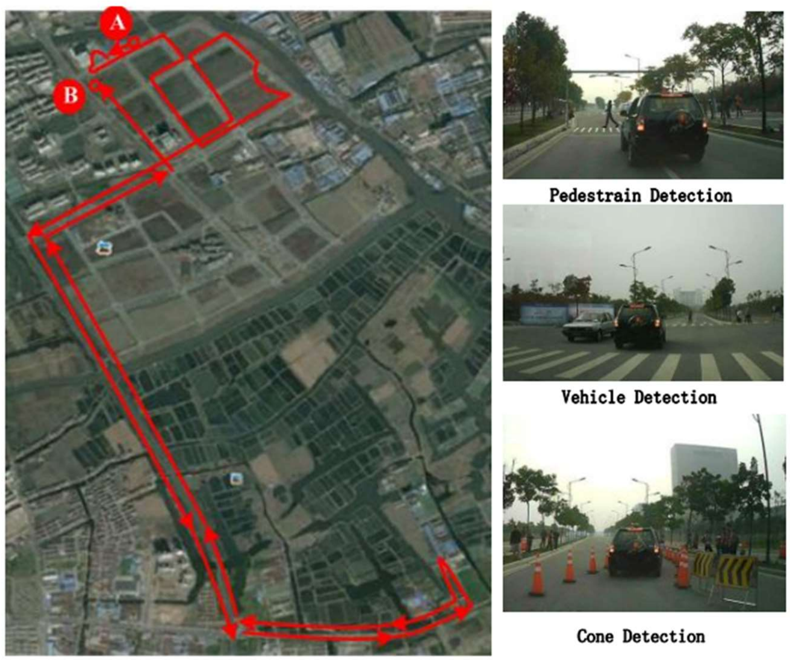


Figure 2: A scenario in which test vehicles must pass through several intersections, a tunnel, and a work zone where pedestrians and vehicles may appear [19].

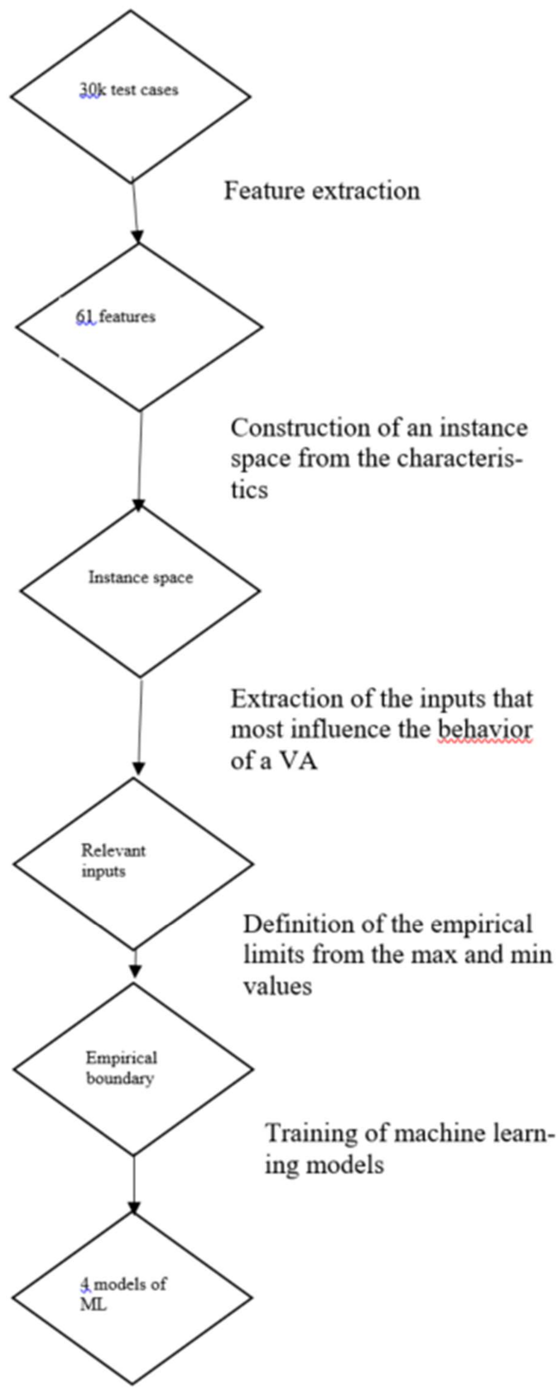
To execute a scenario, it is necessary to consider several parameters such as the position of the vehicle, the weather, the state of the road ... etc., according to the input parameters a vehicle reacts, so that each group of characteristics correspond to outputs that will allow the interpretation of the behavior of a vehicle.

Since AVs are equipped with artificial intelligence like DNNs, these networks must have an appropriate number of parameters to be able to learn from huge databases without losing their own previous experience [24]. This makes a vehicle learn from many input features which increases the complexity for testing with scenarios as it becomes difficult to execute a scenario from many inputs, so the authors of [25] proposed a method to select features that act most on the behavior of an autonomous vehicle. They based their research by asking several questions to which they tried to find answers, as summarized in Table 1.

Question	Answers
How to identify the key characteristics of an effective test case?	Use of ISA (Instance Space Analysis) which builds a two-dimensional instance space based on the input characteristics of the test cases and the test result. The generated instance space provides visual insights into the impact of various features (test input) on the effectiveness of test cases (test result).
Can we predict the outputs of the test case by introducing the key features?	Use of a machine learning model, i.e., implementing a predictive model that from a combination of input parameters will predict the output. This will allow targeting test scenarios in simulation or real life.
Has the system been sufficiently tested?	Usually, the code coverage metric is used for software testing, this metric indicates the rate of code executed during the test, the higher the rate, the fewer bugs there are, but it is not possible to use it in simulation or in real life for testing AVs based on artificial intelligence. For that a new metric called Instance Space Coverage is used

Proposed system

To implement the ISA system proposed by [25] the researchers ran 30k test scenarios from which they extracted 61 features; these features were used to construct an instance space. This instance space was used to extract the features that most influence the behavior of the AV and then empirical bounds were determined from the max and min values. Four machine learning models (random forest RF, k-Nearest Neighbors kNN, Decision Tree DT, Multilayer Perceptron MLP) were used to predict the behavior of the vehicle from the relevant inputs.


30k test cases
Feature extraction
61 features
Construction of an instance space from the characteris-tics
Instance space
Extraction of the inputs that most influence the behavior of a VA
Relevant inputs
Definition of the empirical limits from the max and min values
Empirical boundary
Training of machine learn-ing models
4 models of ML

The test scenarios are distinct and impossible to list. In general, they involve only simple maneuvers, a small number of vehicles and short driving times. For this reason, the researchers have proposed a tool called SAFE TEST that combines two technologies, namely NADE [26] naturalistic and adversarial environment, and an augmented reality test environment [27].

➢ **Augmented reality test environment (called VIL in PRISSMA)**

This solution was developed to address the lack of vehicle interaction available to researchers using test tracks such as Mcity. Currently, these facilities simply offer empty roads to perform tests. To add real vehicles to a test scenario, companies would have to spend thousands of dollars and hours on coordination and control. The team developed an augmented reality environment that allowed them to add virtual traffic to Mcity in the background that the test vehicle considers "real." The simulated vehicles are easily controlled, so specific test scenarios can be repeated perfectly every time.

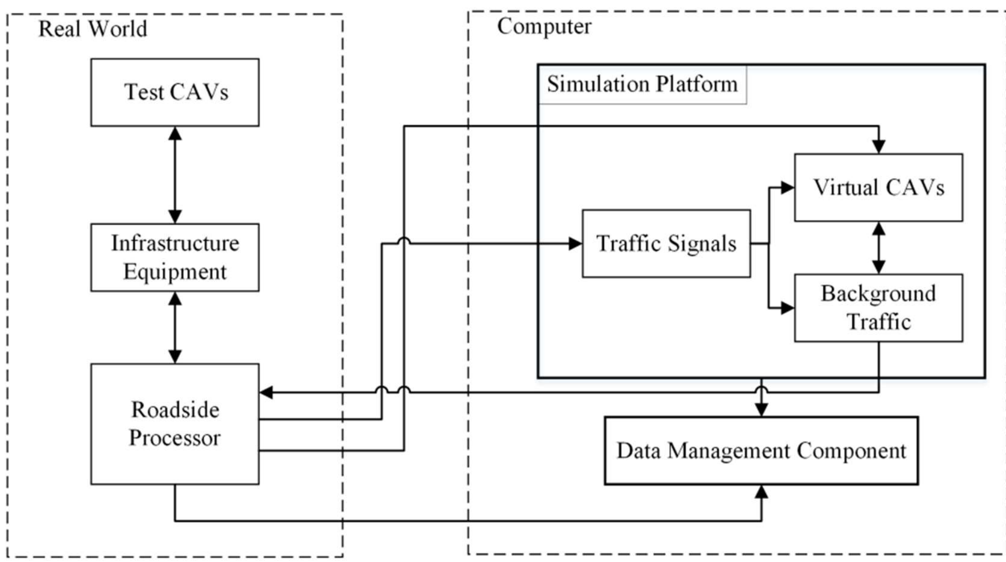


Figure 3 Overall Design of augmented reality environment [27]

The proposed system consists of three elements as shown in Figure 3

- **Simulation platform**

The simulation platform receives data from the CAV and then uses the GPS data to update the position of the vehicle in the simulation environment and build the test scenarios.

- **Autonomous and connected vehicles**

A Lincoln MKZ Hybrid was used for the test of the augmented reality environment. This vehicle is fully connected and automated and is equipped with various sensors as shown in the Figure 4.

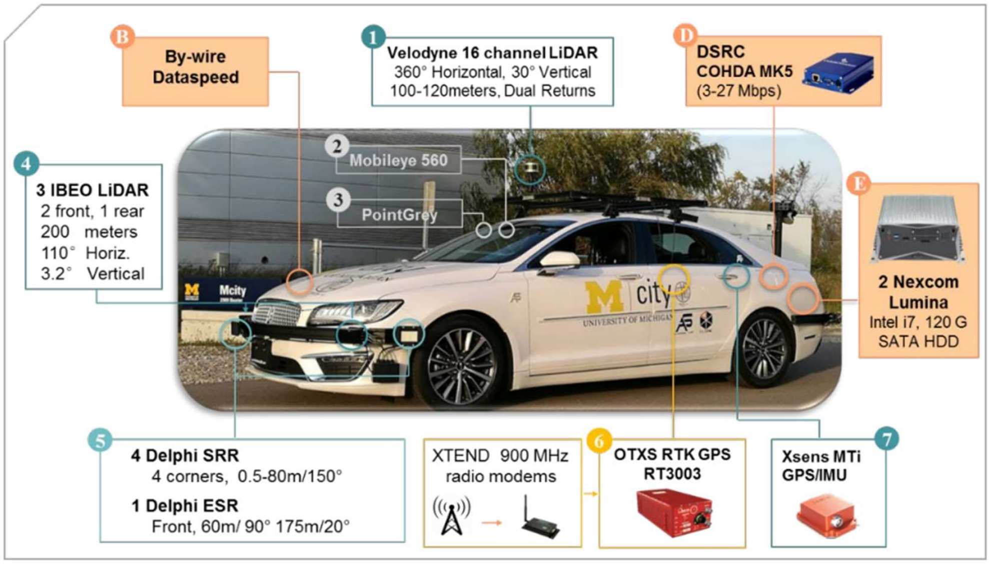


Figure 4: Vehicle used for the simulation [27]

- **Communication network**

The communication network transmits data between the simulation platform and test CAVs

In order to test the proposed system three scenarios have been set up: Level crossing, red light crossing, traffic light priority.

- **NADE (Naturalistic and Adversarial Driving Environment) Equivalent to VIL**

This research addresses the lack of safety critical scenarios that a test vehicle might experience on the road. These rare situations, such as a vehicle merging in front of you as shown in figure 5 or sudden braking, are important for an AV to respond to ensure public confidence. NADE inserts background vehicles that perform these contradictory and rare maneuvers at a much higher rate. Simultaneously, the environment uses natural driving data from the University of Michigan to ensure impartiality.

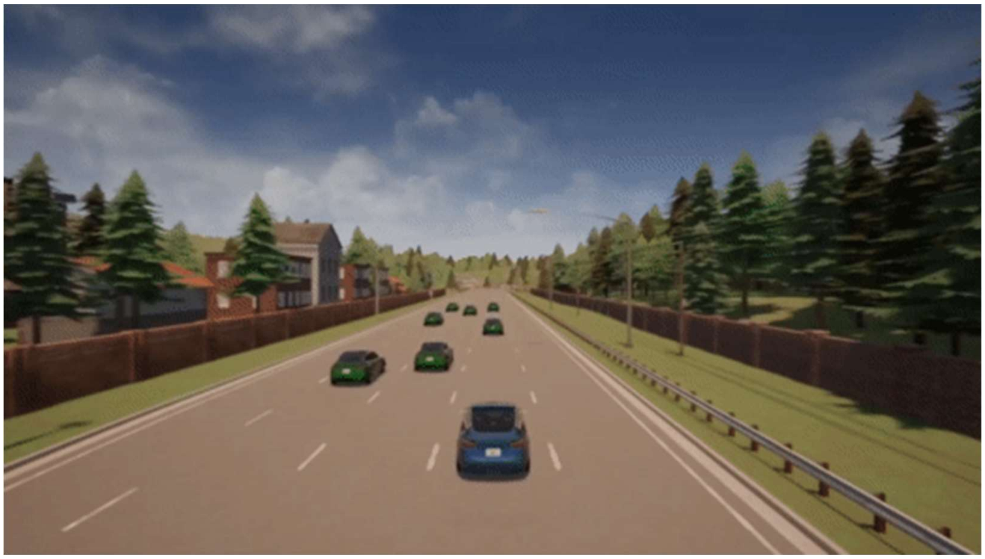


Figure 5: A vehicle merging in front of you

In addition to long-tail event generation, NADE can also provide systematic analysis of safety performance. This includes counterfactual simulation that compares what happened with what would have happened in relevant situations. Long-tail events can also be augmented to meet the user's needs. Other important analyses performed by NADE include safety measures, accident type, accident responsibility, and a VA strength and weakness analysis.

As the space of all possible scenarios is huge, the execution of all these scenarios is costly in terms of time and budget, which is why the researchers of [28] proposed to implement a tool that allows identifying high-risk test scenarios that are most likely to reveal the failures of a system.

For this, they follow three steps:

- **Step 1**
Use multi-objective search to obtain test scenarios that focus on multiple critical aspects of the system and environment at the same time.
- **Step 2**
Reduce the execution time of the search algorithm by proposing a new combination of multi-objective search with surrogate models built based on supervised learning techniques [29].
- **Step 3**
Consists of testing the proposed approach by applying it to an industrial use case that the "PeVi system" (pedestrian detection vision based). It consists of the detection of a pedestrian when vision is disturbed due to metrological conditions for example for that "physics-based simulation platforms" [30] is used.

Just as the researchers who built the PeVi system used the surrogate models to reduce the execution time of the proposed approach other researchers in [31] proposed the surrogate models) during the simulations by applying an iterative approach for testing autonomous vehicles.
In order to test the behavior of a model it is necessary to optimize the objective function (cost function) with the search for a maximum or a minimum, for this it is necessary to use a lot of

simulation in order to reach a convergence. As it is not always obvious to achieve this optimization, the solution is to approximate the cost function with another function that will be optimized. This last one represents the cost function in a region of interest (i.e., the region where the vehicle is malfunctioning)
The scenario consists of a car driving in a straight line and an obstacle whose position is variable as shown in Figure 6. An error in the field of view of the sensor was introduced. The goal of the proposed algorithms was to find the test case with the worst accident evaluation.

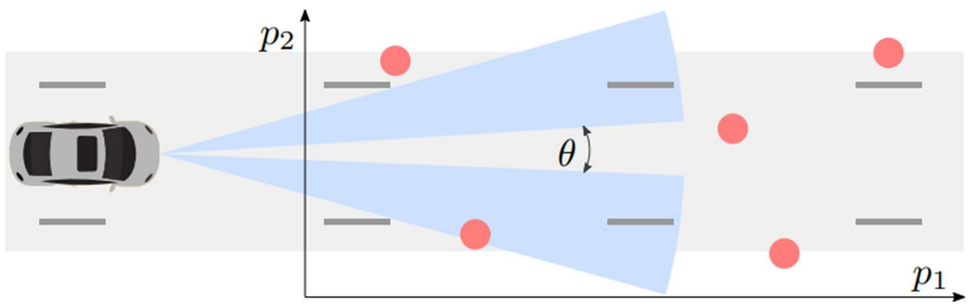


Figure 6: Scenario setup [31]

3.4.2 Functionality-Based Testing (called sub-systems tests in PRISSMA)

Functionality-based testing allows to test vehicles based on 3 functionalities [32][33]: detection/recognition functionality such as vehicle recognition, traffic sign recognition, etc. as shown in Figure 7, decision functionality with respect to the recognized information and action functionality.

Figure 7: Recognition function [19]

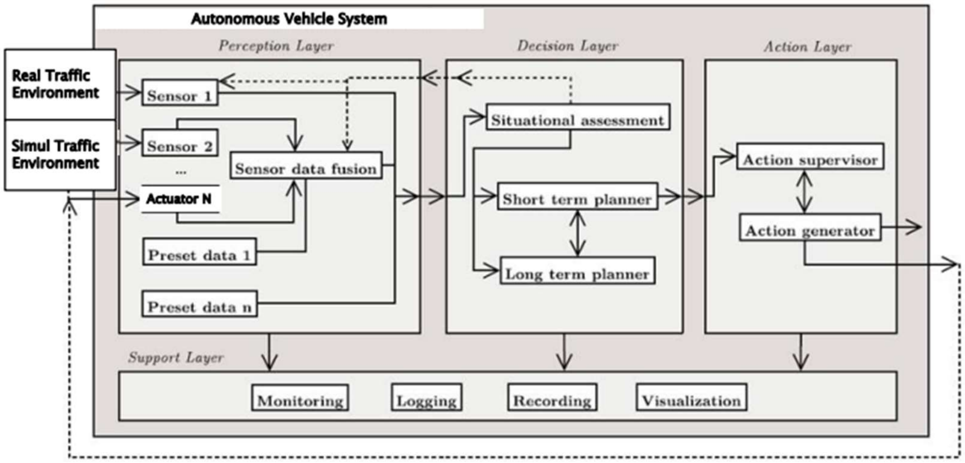


Figure 8: General architecture of functionalities of AV [34].

Figure 7 shows a general architecture of an AV from a functional point of view, so testing a vehicle based on functionality comes down to testing the different artificial intelligence methods that have been presented in section Use of AI in AVs.

Scenario-based testing vs functionality-based testing

	Benefit	Shortcomings
Scenario-based testing	- Complete tests when running a scenario - Ability to detect failures through critical scenarios	-Very large space of scenarios - Important runtime of all scenarios - Limitation of simulation platforms, hence the need to perform real tests - Danger of testing on public roads, hence the need for test tracks
Functionality-based testing	- Possibility to quantitatively evaluate a part of the intelligence implemented on the AV	- The tests are performed separately - Lack of complete vehicle intelligence testing - Lack of a standardized databases that ensure fair comparisons for functional testing of autonomous vehicles

In UTAC POC for WP3 PRISSMA, we address scenario-based tests because they are more relevant and representative of type approve tests.

This State of the art confirms that most of papers cover AI for virtual tests and not AI for physical tests. Nevertheless, some ideas and axes of these papers have inspired us to conceive our physical tests in UTAC WP3 POC, similarly to WP1 inputs, as explained previously in chapter 1.4 « Inputs from WP1 for methods and metrics to evaluate IA repeatability, robustness, overfitting ».

4. Drafting of scenarios/protocols of the 1st POC/tests

4.1 Preamble:

All the following scenarios will be performed on the three different vehicles equipped with ADAS intelligent functions described before.

Firstly, the scenarios will be performed in a subjective way (no equipment) in February 2023 to see which scenario is relevant or not. This part will last 0.5 day per vehicle.

Objective testing, with equipment, during March/April, will follow this protocol.

The ENCAP protocols will be taken as model for all the testing. Depending on the functionalities available on each vehicle, the protocol used will be AEB protocol or AD protocol.

ENCAP has been taken as a model because this is a well-known and mastered protocol; also, their requirements are stricter than Regulation Protocols and can reveal the weakness of an AI system.

Furthermore, most of the vehicle with ADAS and without AI can handle the Regulation Protocol quite easily, whereas the ENCAP Protocol can point out some weakness of the systems. The goal is to challenge the AI system with harder situations.

Below the specifications of ENCAP Protocol that we will follow:

	CPFA	CPNA	CPNCO	CBLA	
Section	7.2.1	7.2.2	7.2.3	7.3.4	
Type of test	AEB			AEB	FCW/ESS
VUT speed [km/h]	10-60			25-60	50-80
VUT direction	Forward			Forward	
Target speed [km/h]	8	5		15	20
Target direction	Coming from Farside	Coming from Nearside		Forward	
Impact location [%]	50	25,75*	50	50	25
Dummy Articulation	Yes – as per test speed				

	VUT	EPT	EBT
Speed	+ 1.0 km/h	± 0.2 km/h	± 0.5 km/h
Lateral deviation	0 ± 0.05 m	0 ± 0.05 m for crossing scenarios 0 ± 0.15 m for longitudinal scenarios	
Lateral velocity		0 ± 0.15 m/s	0 ± 0.15 m/s
Relative distance			
Yaw velocity (upto T_{STEER})	0 ± 1.0 °/s		
Steering wheel velocity (upto T_{STEER})	0 ± 15.0 °/s		

4.2 Critical scenarios and repeatability

The first category of testing is about Repeatability, the goal is to perform many repetitions of a given scenario, with the same conditions and verify if the performance is similar.

Today, on a vehicle equipped with classic ADAS systems (ex: AEB), we note that the performances are not always repeatable. Here are some examples of repeatability results on ENCAP scenarios:

<u>Pedestrian scenarios:</u>

CPFA-50 AEB	Day	50	8	50/-	AEB	-	-	Y	Impact
CPFA-50 AEB	Day	50	8	50/-	AEB	-	-	N	Avoidance
CPFA-50 AEB	Day	50	8	50/-	AEB	-	-	Y	Impact
CPFA-50 AEB	Day	55	8	50/-	AEB	-	-	Y	Impact
CPFA-50 AEB	Day	55	8	50/-	AEB	-	-	N	Avoidance
CPFA-50 AEB	Day	55	8	50/-	AEB	-	-	Y	Impact

CPNA-25 AEB	Day	35	5	25/-	AEB	-	-	Y	Impact
CPNA-25 AEB	Day	35	5	25/-	AEB	-	-	Y	Impact
CPNA-25 AEB	Day	35	5	25/-	AEB	-	-	N	Avoidance
CPLA-50 AEB	Day	25	5	50/-	AEB	-	-	N	Avoidance
CPLA-50 AEB	Day	25	5	50/-	AEB	-	-	Y	Impact
CPLA-50 AEB	Day	25	5	50/-	AEB	-	-	Y	Impact

<u>Car to Car scenarios:</u>

CCRs AEB	Day	20	0	-/-50	AEB	-	-	N	Avoidance
CCRs AEB	Day	20	0	-/-50	AEB	-	-	Y	Impact
CCRs AEB	Day	20	0	-/-50	AEB	-	-	N	Avoidance
CCRs AEB	Day	20	0	-/-50	AEB	-	-	N	Avoidance

<u>Bicycle scenarios:</u>

ENCAP 2020	CBLA-50 AEB	Day	45	15	50/-	AEB	-	-	N	Avoidance
ENCAP 2020	CBLA-50 AEB	Day	45	15	50/-	AEB	-	-	Y	Impact
ENCAP 2020	CBLA-50 AEB	Day	45	15	50/-	AEB	-	-	N	Avoidance

The goal is to see if the AI on the last ADAS system increase the performances or not, compared to a system without AI.

The following scenarios will be performed 10 times each.

4.2.1 CPNCO-50 (Car to Pedestrian Nearside Child Obstructed 50%)

This scenario refers to the ENCAP 2023 protocol:

7.2.3 Car-to-Pedestrian Nearside Child Obstructed

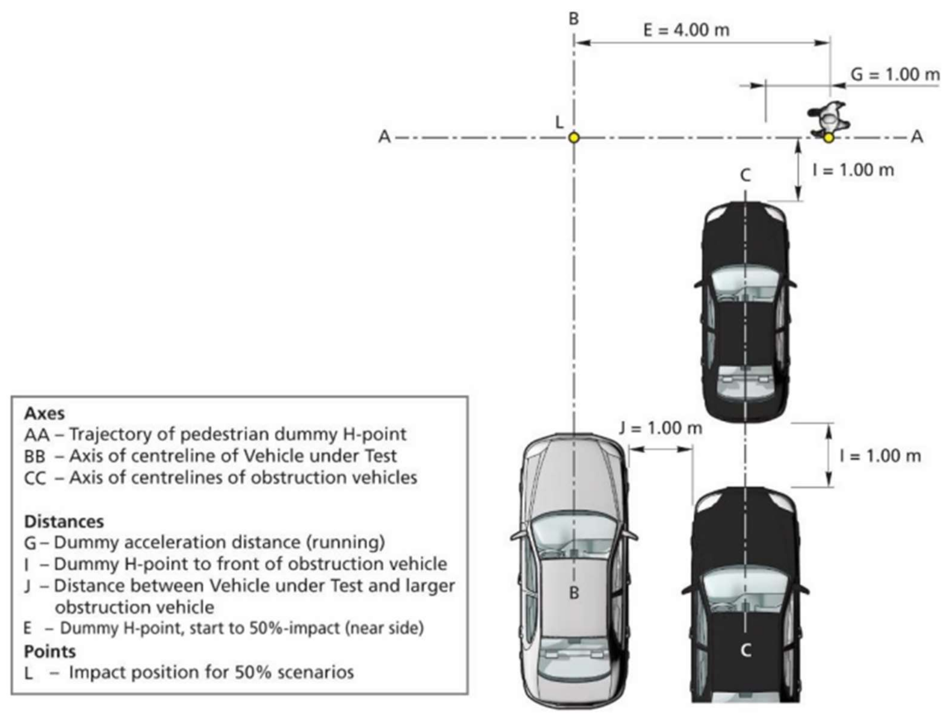


Figure 7-3: CPNCO-50 scenario, Running Child from Nearside from Obstruction

Car-to-Pedestrian Nearside Child Obstructed 50% (CPNCO-50) – a collision in which a vehicle travels forwards towards a child pedestrian crossing (5kph) its path running from behind and obstruction from the nearside and the frontal structure of the vehicle strikes the pedestrian at 50% of the vehicle's width when no braking action is applied.

4.2.2 CPFA-50 (Car to Pedestrian Farside Adult 50%)

This scenario refers to the ENCAP 2023 protocol:

7.2.1 Car-to-Pedestrian Farside Adult

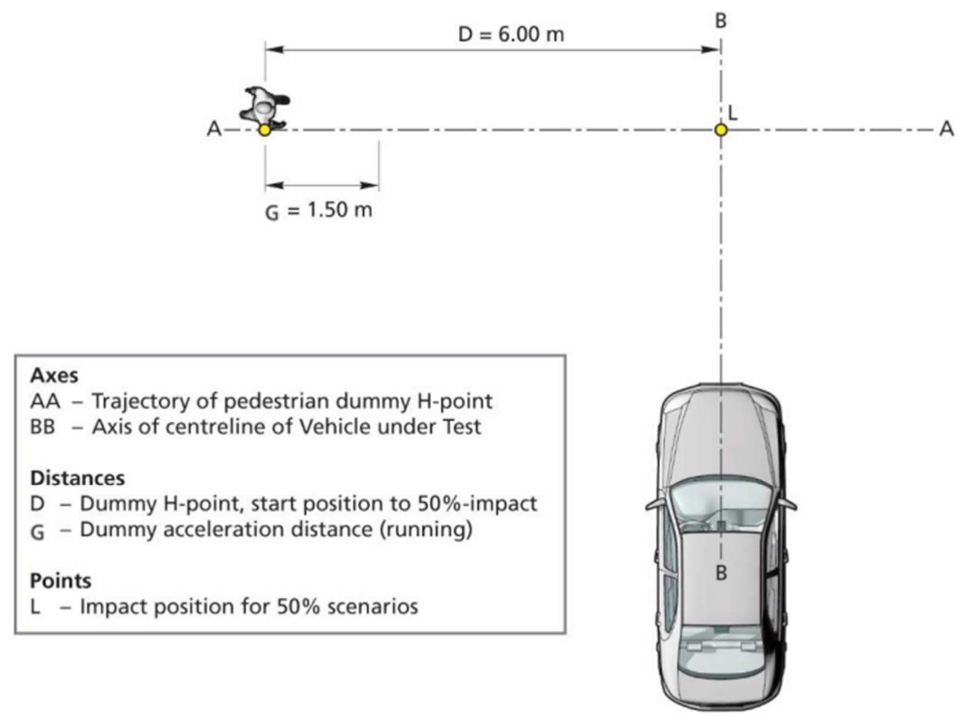


Figure 7-1: CPFA-50 scenario, Adult running from Farside

Car-to-Pedestrian Farside Adult 50% (CPFA-50) – a collision in which a vehicle travels forwards towards an adult pedestrian crossing (8kph) its path running from the farside and the frontal structure of the vehicle strikes the pedestrian at 50% of the vehicle's width when no braking action is applied.

4.2.3 CBLA-50 (Car to Bicyclist Longitudinal Adult 50%)

This scenario refers to the ENCAP 2023 protocol:

7.3.4 Car-to-Bicyclist Longitudinal Adult

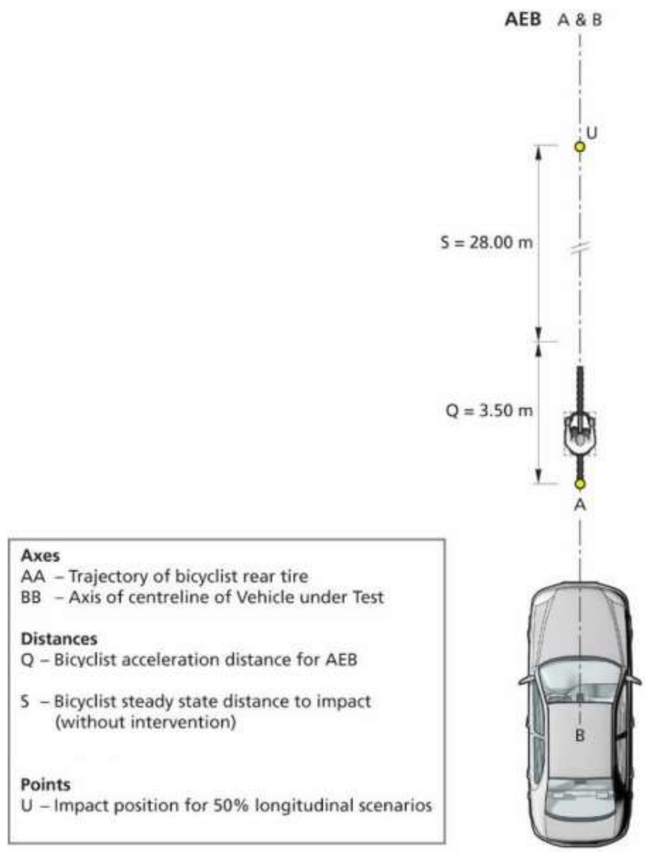


Figure 7-13: CBLA scenarios, Longitudinal Bicyclist

Car-to-Bicyclist Longitudinal Adult 50% (CBLA-50) – a collision in which a vehicle travels forwards towards a bicyclist cycling (15kph) in the same direction in front of the vehicle where the vehicle would strike the cyclist at 50% of the vehicle's width when no braking action is applied.

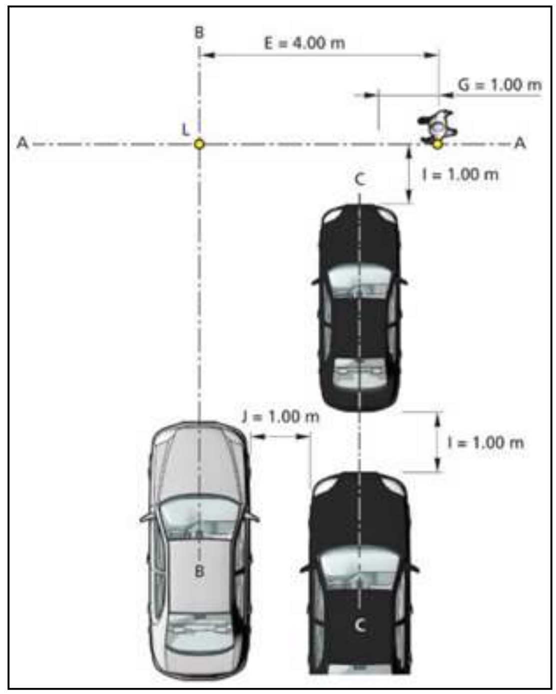
4.3 Critical scenarios and robustness

The second category of testing is about Robustness, the goal is to perform many variants of a given scenario and verify if the performance is similar.

For examples, we can change the speed of the target, the color of the clothes…

4.3.1 CPNCO-50 (Car to Pedestrian Nearside Child Obstructed 50%)

Same scenario as 4.2.1 with different alternative of it. If the obstruction is too harsh, it can be removed.



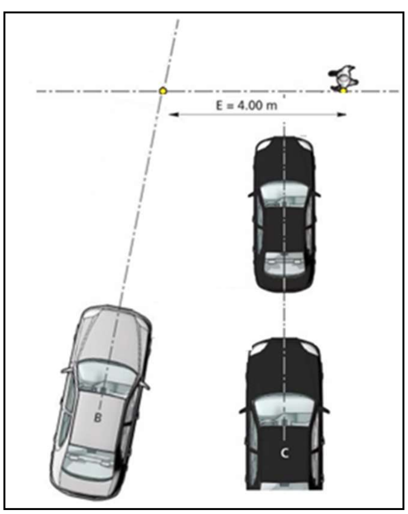
SPEED CHANGING:
- Child running at 6kph
- Child Start @3kph and accelerate @6kph

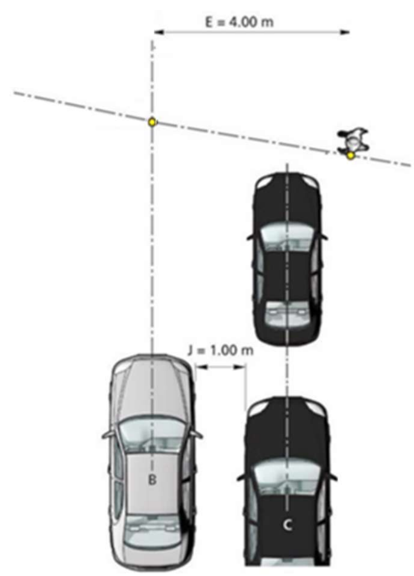
FORM CHANGING:
- Child wearing a backpack
- Child with a stuffed toy

COLOR CHANGING:
- Child with yellow jacket

SURROUNDING CONDITIONS CHANGING:
- Strong Light in front of VUT

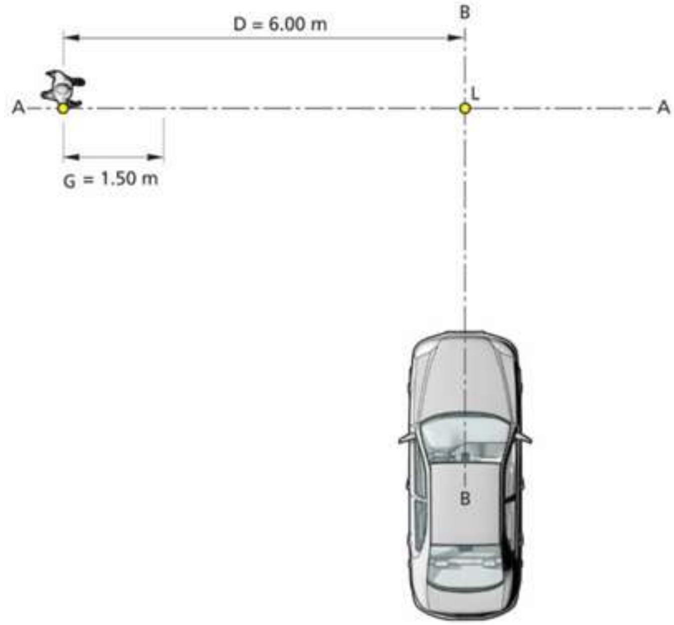
ANGLE CHANGING:
- VUT angle >90° (To be defined) -
Child angle <90° (To be defined)





4.3.2 CPFA-50 (Car to Pedestrian Farside Adult 50%)

Same scenario as 4.2.2 with different alternative of it:



SPEED CHANGING:
- Pedestrian 5kph
- Pedestrian starts 5kph and accelerate 8kph

FORM CHANGING:
- Group of Pedestrian waiting and 1 moving
- Pedestrian with backpack

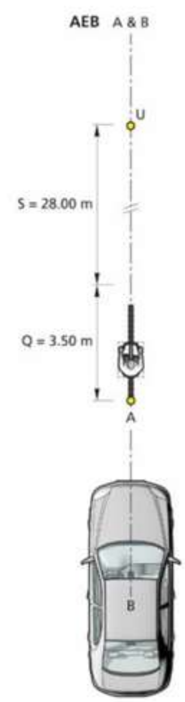
COLOR CHANGING:
- Pedestrian with yellow jacket

SURROUNDING CONDITIONS CHANGING:
- Strong Light in front of VUT

ANGLE CHANGING: (similar as CPNCO with angle changing)
- VUT angle >90°
- Pedestrian angle <90°

4.3.3 CBLA-50 (Car to Bicyclist Longitudinal Adult 50%)

Same scenario as 4.2.3 with different alternative of it:



SPEED CHANGING:
- Bicyclist Starts 10kph and accelerate 20kph (to be confirmed at first tests)
- Bicyclist 25kph? (To be confirmed at first tests)

FORM CHANGING:
- Bike with cargo rack
- Adult with backpack

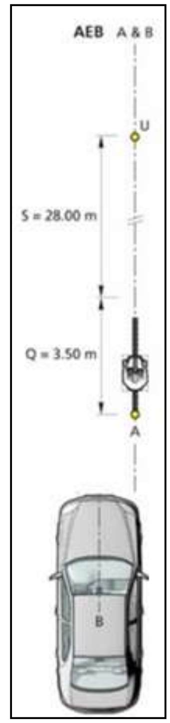
COLOR CHANGING:
- Adult with yellow jacket

SURROUNDING CONDITIONS CHANGING:
- Strong Light in front of VUT

OVERLAP CHANGING:

- 75% (symmetry of 25% usual case)

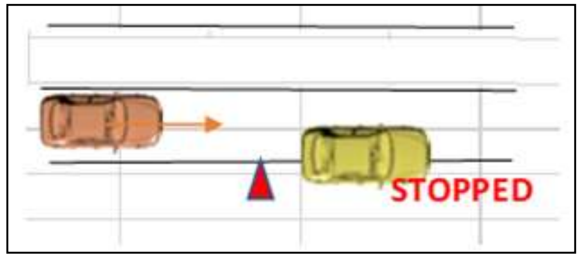
- 5% (bike close to road edge)



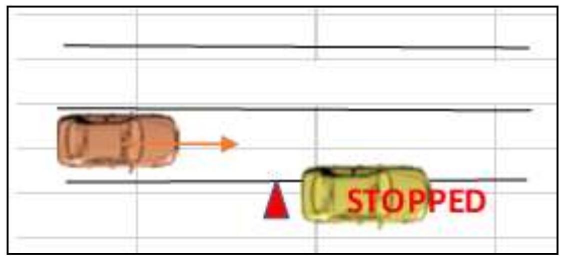
4.3.4 Stationary Car on Emergency Lane

A stationary car is stopped in an emergency lane, with a traffic sign (red triangle), different position of the stopped vehicle:

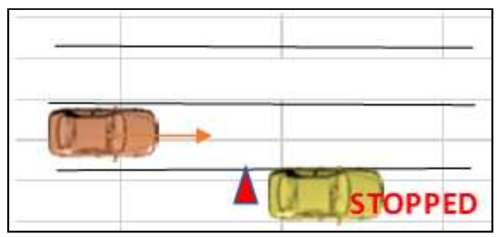
-50%:



- 25% (To be defined):

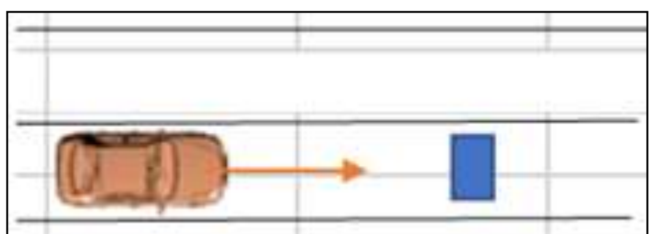


-0% (edge limit):

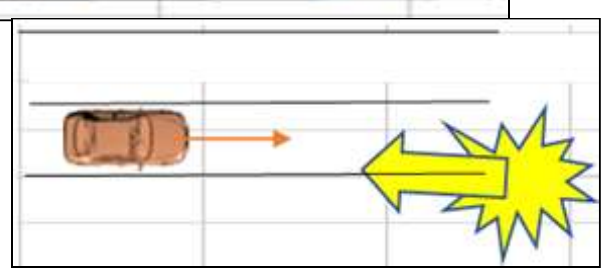


4.3.5 Stationary object or dazzling light on Highway

-Stationary Object:



- Dazzling light (difficult perception):
If possible, it will be performed at the exit of the highway Tunnel



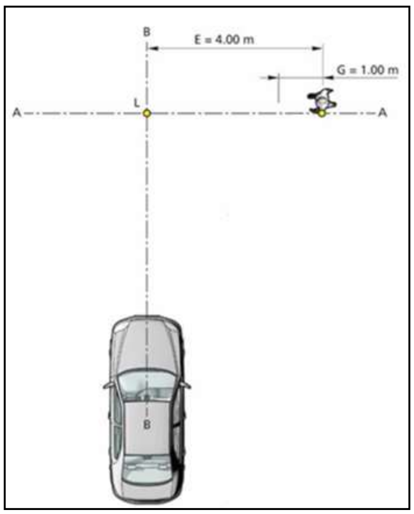
4.4 pre-critical scenarios (anticipating to avoid AEB/critical maneuvers)

The third category of testing is about Anticipation, the goal is to perform some classic scenario by changing some conditions to see if the vehicle can anticipate a potential danger (without activation of AEB).
Each scenario will be repeated twice (2 runs per scenario).

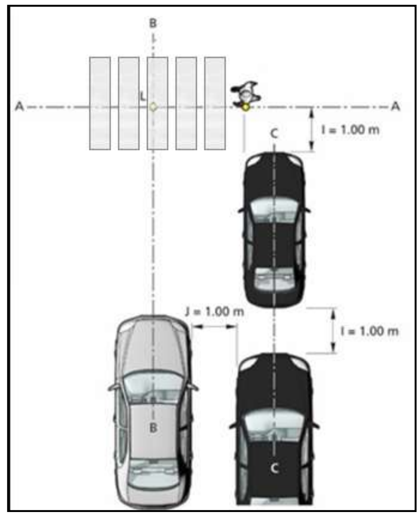
4.4.1 CPNCO-50 (Car to Pedestrian Nearside Child Obstructed 50%)

Same scenario as 4.2.1 with different alternative of it:

-Without Obstruction:



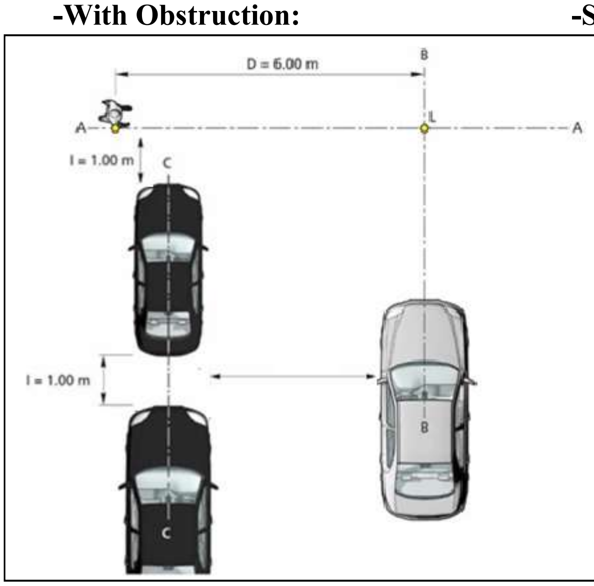
-Stationary Child (edge of pedestrian crossing):



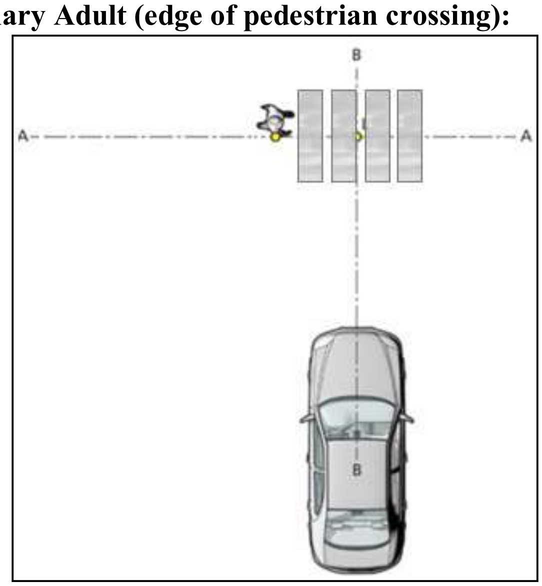
4.4.2 CPFA-50 (Car to Pedestrian Farside Adult 50%)

Same scenario as 4.2.2 with different alternative of it:

-With Obstruction:



-Stationary Adult (edge of pedestrian crossing):

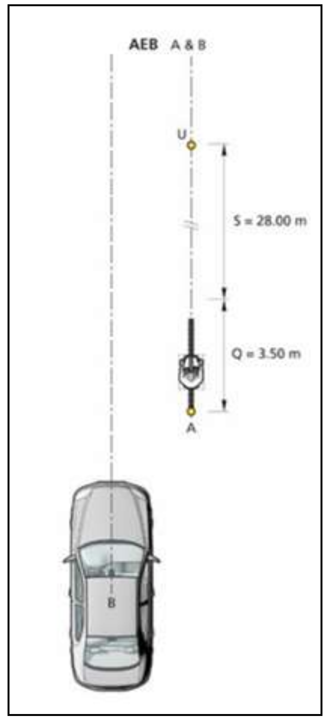


4.4.3 CBLA-50 (Car to Bicyclist Longitudinal Adult 50%)

Same scenario as 4.2.3 with different alternative of it:

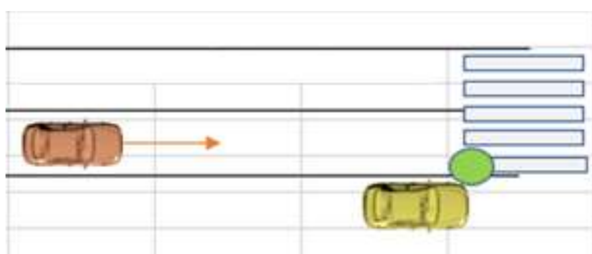
-With different bearing:

-Bicycle close to VUT path (cycling track):

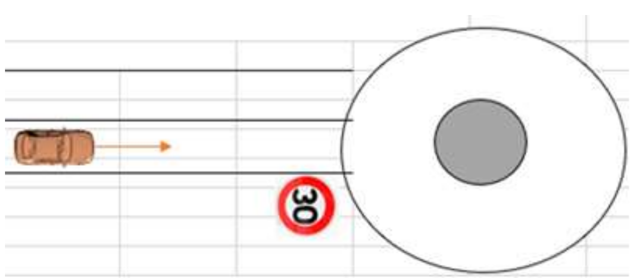


4.4.4 Anticipation without Target

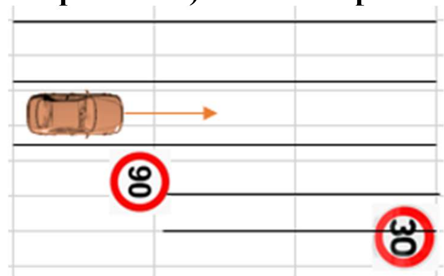
-Pedestrian Crossing, Green Traffic light and obstruction (potentially hidden pedestrian):



-Approach of strong curve (ex: roundabout) with late traffic sign:



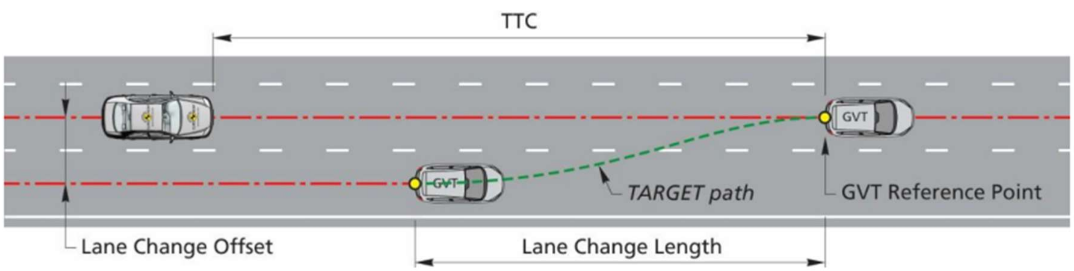
-Traveling on highway (ex: 90kph limited) and lower speed traffic sign visible (ex: exit):



4.4.5 Car to car

- Target Cut-in followed by a braking:

ACC CUT-IN	VUT	GVT	Lane Change Manoeuvre GVT		
			Lateral Acceleration	Change Length	Radius of turning segments
Cut-in					
Cut-in @ TTC = 0.00	50 km/h	10 km/h	0.5 m/s²	14.5 m	15 m
Cut-in @ TTC = 1.50	120 km/h	70 km/h	1.5 m/s²	60.0 m	250 m



The same configuration as ENCAP Highway Assist can be used for the Cut-in part.
This maneuver is followed by a braking of the target with a deceleration of 2m/s² or 6m/s² (same as ENCAP protocol).

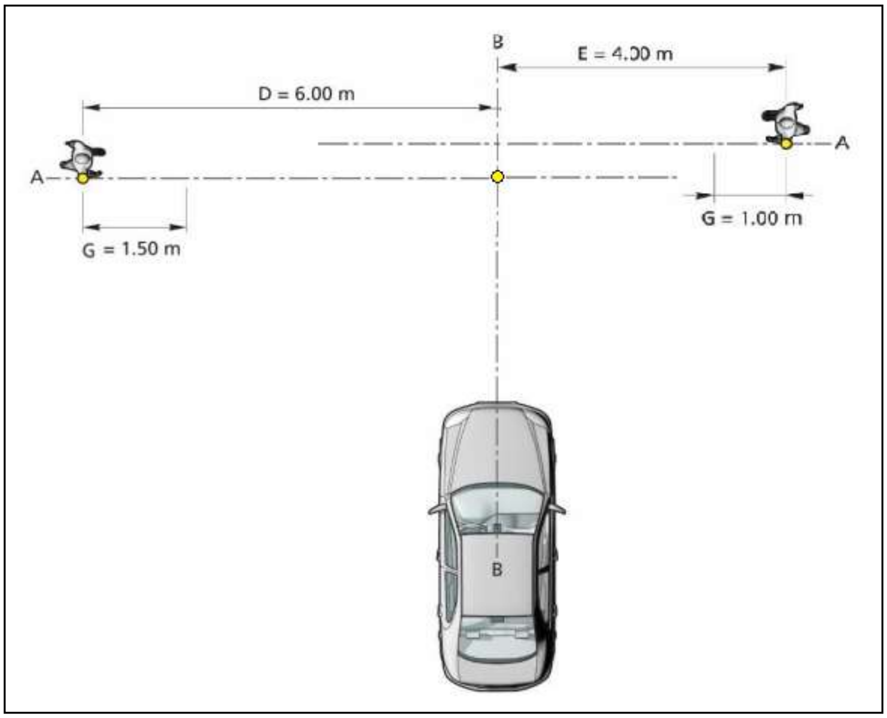
4.5 New random scenarios (to avoid the over-learning of AI (overfitting))

The last category of testing is about random situation, the goal is to perform some random scenario which (in theory) have never been met by the vehicle.

Each scenario will be repeated twice (2 runs per scenario).

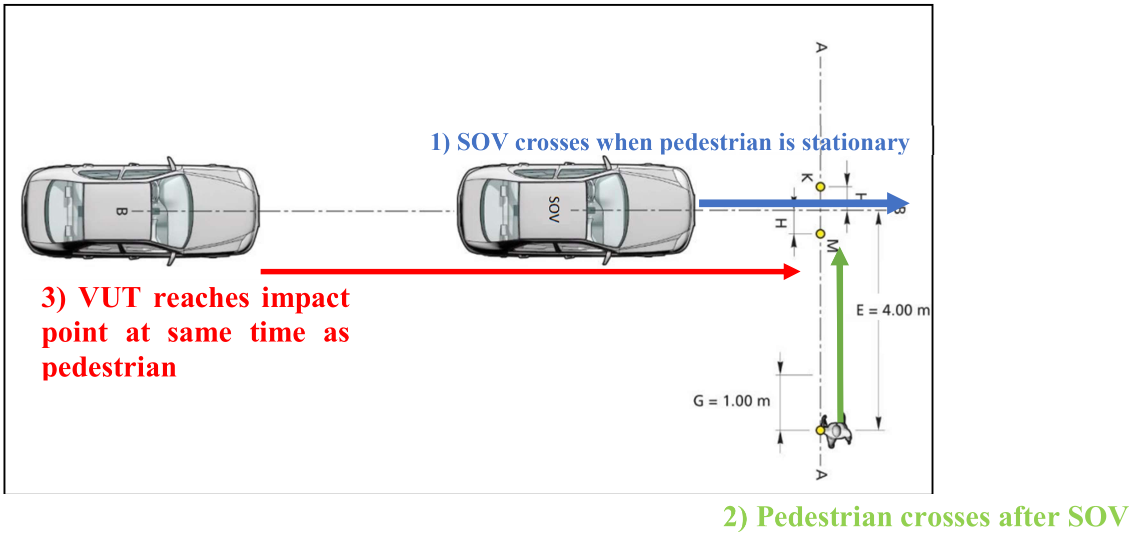
4.5.1 Pedestrians Crossing with two dummies:

Two crossing pedestrians, one from farside, one from nearside, synchronized or not.



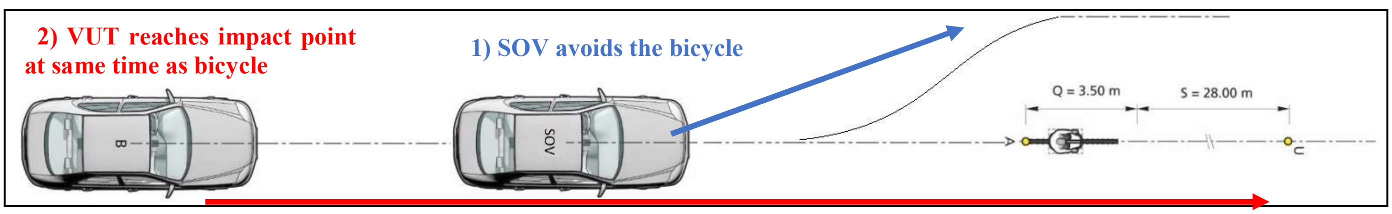
4.5.2 Crossing Pedestrian with VUT preceded by a vehicle:

The VUT follows an SOV (Secondary Other Vehicle) with a distance X, then a pedestrian (adult or child) crosses in front of the VUT. The distance between VUT and SOV depends on the ACC.



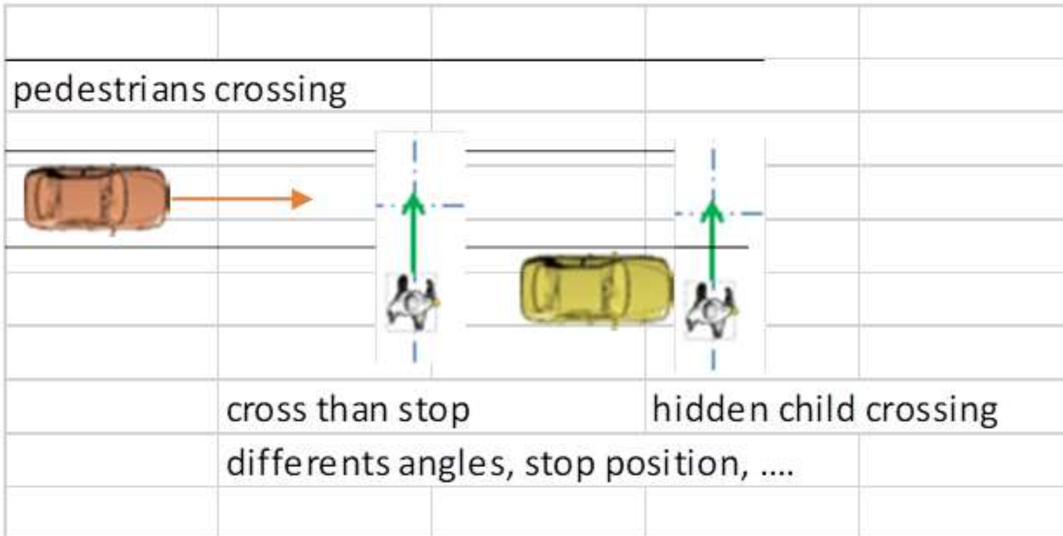
4.5.3 Longitudinal Bicyclist with VUT preceded by a vehicle

The VUT follows an SOV (Secondary Other Vehicle) with a distance to be defined in the same line as a bicycle. At X meters (depending on ACC) of the target, the SOV avoids the bicycle.



4.5.4 Crossing Pedestrian with two dummies, one stops before impact

This scenario is similar as the CPNCO, a second pedestrian is added and starts to cross the VUT path before the stationary vehicle, then stops before the impact.



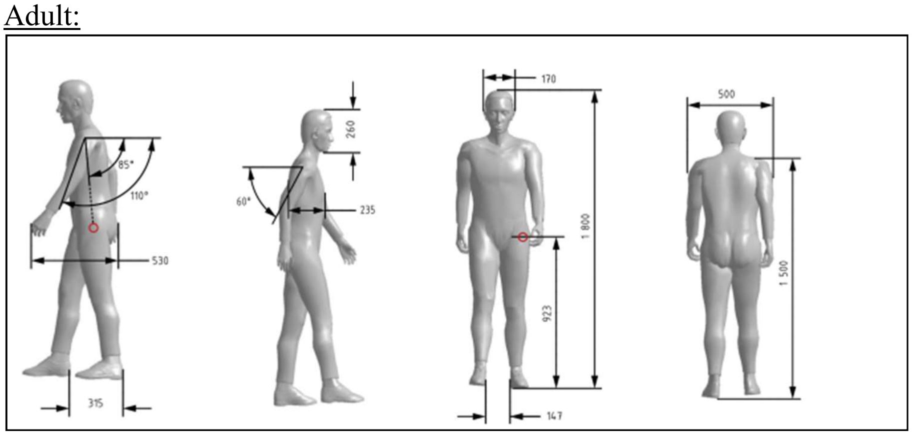
4.6 Equipment for testing and measurement

4.6.1 Targets:

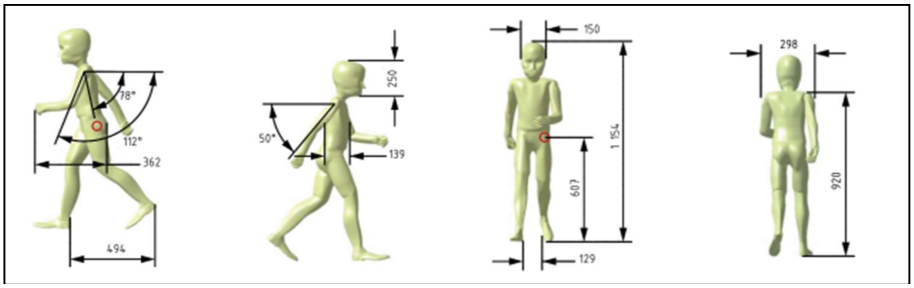
The used targets for the previous scenario are those defined by the **ISO 19206-2_2018** (Pedestrian) and the **ISO 19206-4_2020** (Bicycle). For the Robustness scenarios, the targets will be adapted.

4.6.1.1 ISO 19206-2_2018:

Adult:

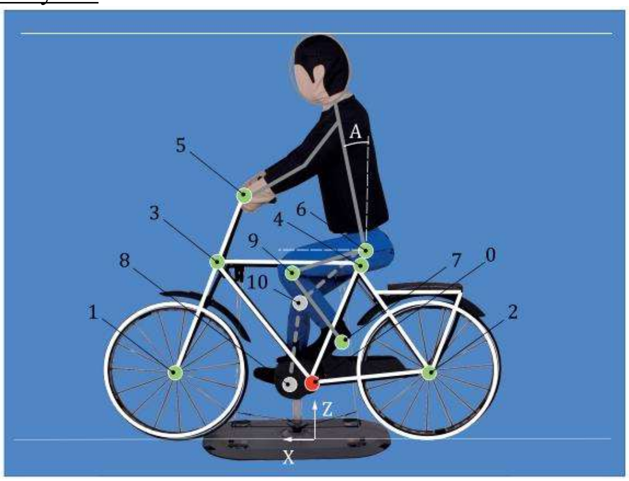


Child:



4.6.1.2 ISO 19206-4_2020:

<u>Bicycle:</u>



Segment	X	Z	Tolerance	Unit
0 Centre of bottom bracket of BT bicycle	0	280	±10	mm
1 Centre axis front wheel	670	340	±10	mm
2 Centre axis rear wheel	-540	340	±10	mm
3 Front top frame	430	855	±10	mm
4 Rear top frame (upper range sloped top tube)	-215	860	±10	mm
4 Rear top frame (lower range sloped top tube)	-145	460	±10	mm
5 Handlebar	310	1 180	±10	mm
6 Saddle	-235	935	±10	mm
7 Lower edge left foot[a]	105	495	±20	mm
8 Lower edge right foot	80	200	±20	mm
9 Knee point, left [b]	150	860	±20	mm
10 Knee point, right	85	700	±20	mm
Total height (for 10° torso angle)	1 865		±20	mm
Total length	1 890		±20	mm
A Torso angle	10 and 30		±2	°
[a] Lowest point of shoe – centre line tibia.				
[b] Knee point: rotation point of knee.				

4.6.2 Propulsion systems:

The propulsion systems used are in accordance with the TB029 of ENCAP.

					Euro NCAP Test Targets			
					Global Vehicle Target (GVT)	Euro NCAP Pedestrian Target Adult (EPTa)	Euro NCAP Pedestrian Target Child (EPTc)	Euro NCAP Bicyclist Target (EBT)
				Supplier	ABD	4a	4a	4a
				Product	Soft Car 360	4activePA Adult	4activePA Child	4activeBS
				Version	DRI Rev G Feb 2020	v4v4	v3v3	v5v5
	Supplier	Product	Version					
Propulsion Systems	ABD	GST100	V1.0 (P8503) & (P8328) with car panel					
	ABD	GST120	V1.0 (P12218)					
	ABD	SPT System	SPT20/SPT20s					
	ABD	LaunchPad 50 & Launchpad 60	V 1.0 (P9226) without extension					
			V 1.0 (P9226) with extension					
	ABD	LaunchPad 80	V 1.0 (P11000) without extension					
			V 1.0 (P11000) with extension					

4.6.3 VUT equipment:

4.6.3.1 Motion Measurement

MOTION PACK 1		
	*Manufacturer* Oxford Technical Solutions (OxTS)	
	Unit model TO BE DEFINED	
	Sensors Accelerometers (Servo) / Gyros (MEMS)	
	Data output rate 100 Hz	*Coupling method* GNSS / INS

4.6.3.2 Data Recording System

CONTROLER		
	*Manufacturer* Antony Best Dynamics (ABD)	
	Unit model XR Omni	
	Sampling rate 100 Hz	
	Analog input voltage ± 10 V	*A / D conversion* 16 bits

4.6.3.3 HMI Analysis

VIDEO VBOX		
	*Manufacturer* Racelogic	
	Unit model	*Frame rate*

GOPRO		
	*Manufacturer* GoPro	
	Unit model	*Video resolution*

5. References

[1] GAO, Paul, KAAS, Hans-Werner, MOHR, Det, *et al.* Automotive revolution–perspective towards 2030: How the convergence of disruptive technology-driven trends could transform the auto industry. *Advanced Industries, McKinsey & Company*, 2016.
[2] JAMIL, Sadia. Artificial intelligence and journalistic practice: The crossroads of obstacles and opportunities for the Pakistani journalists. *Journalism Practice*, 2021, vol. 15, no 10, p. 1400-1422.
[3] SHI, Weijing, ALAWIEH, Mohamed Baker, LI, Xin, *et al.* Algorithm and hardware implementation for visual perception system in autonomous vehicle: A survey. *Integration*, 2017, vol. 59, p. 148-156.
[4] KATRAKAZAS, Christos, QUDDUS, Mohammed, CHEN, Wen-Hua, *et al.* Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions. *Transportation Research Part C: Emerging Technologies*, 2015, vol. 60, p. 416-442.
[5] SCHWARTING, Wilko, ALONSO-MORA, Javier, et RUS, Daniela. Planning and decision-making for autonomous vehicles. *Annual Review of Control, Robotics, and Autonomous Systems*, 2018, vol. 1, p. 187-210.
[6] SHAFAEI, Sina, KUGELE, Stefan, OSMAN, Mohd Hafeez, *et al.* Uncertainty in machine learning: A safety perspective on autonomous driving. In: *International Conference on Computer Safety, Reliability, and Security*. Springer, Cham, 2018. p. 458-464.
[7] LI, Jingyue, ZHANG, Jin, et KALOUDI, Nektaria. Could we issue driving licenses to autonomous vehicles?. In: *International Conference on Computer Safety, Reliability, and Security*. Springer, Cham, 2018. p. 473-480.
[8] A. Taeihagh and H. S. M. Lim, "Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks," Transp. Rev., vol.39, no.1, pp.103–128, 2019.
[9] MA, Yifang, WANG, Zhenyu, YANG, Hong, *et al.* Artificial intelligence applications in the development of autonomous vehicles: a survey. *IEEE/CAA Journal of Automatica Sinica*, 2020, vol. 7, no 2, p. 315-329.
[10] CIREGAN, Dan, MEIER, Ueli, et SCHMIDHUBER, Jürgen. Multi-column deep neural networks for image classification. In: *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012. p. 3642-3649.
[11] ZENG, Yujun, XU, Xin, SHEN, Dayong, *et al.* Traffic sign recognition using kernel extreme learning machines with deep perceptual features. *IEEE Transactions on Intelligent Transportation Systems*, 2016, vol. 18, no 6, p. 1647-1653.
[12] CHEN, Chenyi, SEFF, Ari, KORNHAUSER, Alain, *et al.* Deepdriving: Learning affordance for direct perception in autonomous driving. In: *Proceedings of the IEEE international conference on computer vision*. 2015. p. 2722-2730.
[13] BOJARSKI, Mariusz, DEL TESTA, Davide, DWORAKOWSKI, Daniel, *et al.* End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.
[14] BOJARSKI, Mariusz, YERES, Philip, CHOROMANSKA, Anna, *et al.* Explaining how a deep neural network trained with end-to-end learning steers a car. *arXiv preprint arXiv:1704.07911*, 2017.
[15] ALCANTARILLA, Pablo F., STENT, Simon, ROS, German, *et al.* Street-view change detection with deconvolutional networks. *Autonomous Robots*, 2018, vol. 42, no 7, p. 1301-1322.
[16] VISHNUKUMAR, Harsha Jakkanahalli, BUTTING, Björn, MÜLLER, Christian, *et al.* Machine learning and deep neural network—Artificial intelligence core for lab and real-world test and validation for ADAS and autonomous vehicles: AI for efficient and quality test and validation. In: *2017 Intelligent Systems Conference (IntelliSys)*. IEEE, 2017. p. 714-721.
[17] NOTOMISTA, Gennaro et BOTSCH, Michael. A machine learning approach for the segmentation of driving maneuvers and its application in autonomous parking. *Journal of Artificial Intelligence and Soft Computing Research (JAISCR)*, 2017, vol. 7, no 4, p. 243-255.
[18] HARDY, Jason et CAMPBELL, Mark. Contingency planning over probabilistic obstacle predictions for autonomous road vehicles. *IEEE Transactions on Robotics*, 2013, vol. 29, no 4, p. 913-929.
[19] LI, Li, HUANG, Wu-Ling, LIU, Yuehu, *et al.* Intelligence testing for autonomous vehicles: A new approach. *IEEE Transactions on Intelligent Vehicles*, 2016, vol. 1, no 2, p. 158-166.
[20] KOOPMAN, Philip et WAGNER, Michael. Challenges in autonomous vehicle testing and validation. *SAE International Journal of Transportation Safety*, 2016, vol. 4, no 1, p. 15-24.
[21] NHTSA, Preliminary Statement of Policy Concerning Automated Vehicles, May 2013, http://www.nhtsa.gov/staticfiles/rulemaking/pdf/ Automated_Vehicles_Policy.pdf, accessed Oct. 2015.
[22] Road vehicles -- Functional Safety -- Part 3: Concept Phase, ISO 26262- 3:2011, Nov. 15, 2011.
[23] ZÖLDY, Máté, SZALAY, Zsolt, et TIHANYI, Viktor. Challenges in homologation process of vehicles with artificial intelligence. *Transport*, 2020, vol. 35, no 4, p. 447-453.

[24] GOODFELLOW, Ian, BENGIO, Yoshua, et COURVILLE, Aaron. *Deep learning*. MIT press, 2016.
[25] ALETI, Aldeida, *et al.* Identifying Safety-critical Scenarios for Autonomous Vehicles via Key Features. *arXiv preprint arXiv:2212.07566*, 2022.
[26] FENG, Shuo, YAN, Xintao, SUN, Haowei, *et al.* Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment. *Nature communications*, 2021, vol. 12, no 1, p. 1-14.
[27] LIU, Henry et FENG, Yiheng. *Development of an augmented reality environment for connected and automated vehicle testing*. University of Michigan, Ann Arbor, Transportation Research Institute, 2019.
[28] BEN ABDESSALEM, Raja, NEJATI, Shiva, BRIAND, Lionel C., *et al.* Testing advanced driver assistance systems using multi-objective search and neural networks. In: *Proceedings of the 31st IEEE/ACM international conference on automated software engineering*. 2016. p. 63-74.
[29] WITTEN, Ian H. et FRANK, Eibe. Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record*, 2002, vol. 31, no 1, p. 76-77.
[30] TASS International. PreScan simulation of ADAS and active safety. https://www.tassinternational.com/prescan. Last accessed: March 2016.
[31] BEGLEROVIC, Halil, STOLZ, Michael, et HORN, Martin. Testing of autonomous vehicles using surrogate models and stochastic optimization. In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017. p. 1-6.
[32] HUANG, WuLing, WEN, Ding, GENG, Jason, *et al.* Task-specific performance evaluation of UGVs: Case studies at the IVFC. *IEEE Transactions on Intelligent Transportation Systems*, 2014, vol. 15, no 5, p. 1969-1979.
[33] LI, Li, WEN, Ding, ZHENG, Nan-Ning, *et al.* Cognitive cars: A new frontier for ADAS research. *IEEE Transactions on Intelligent Transportation Systems*, 2011, vol. 13, no 1, p. 395-407.
[34] HUANG, WuLing, WANG, Kunfeng, LV, Yisheng, *et al.* Autonomous vehicles testing methods review. In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2016. p. 163-168.
[35] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, “A brief review of domain adaptation,” CoRR, vol. abs/2010.03978, 2020. [Online]. Available: https://arxiv.org/abs/2010.03978
[36] D. V. Vargas and S. Kotyan, “Model agnostic dual quality assessment for adversarial machine learning and an analysis of current neural networks and defenses,” CoRR, vol. abs/1906.06026, 2019. [Online]. Available: http://arxiv.org/abs/1906.06026

Chapter 2: CEREMA/LNE POC

1. Introduction

More and more intelligent systems on vehicles use AI (e.g., visual or mixed navigation, sign recognition, road tracking and obstacle detection). The qualification of these systems requires verification in all kinds of scenarios, including, for example, taking into account degraded weather conditions. For cost and safety reasons, these qualification tests cannot be carried out in real conditions, as some tests may present risks or have frequencies of occurrence too low to allow the collection of large series of data. For this reason, sensor simulation tools and degraded weather conditions (physical, numerical or hybrid) must be implemented. Additionally, simulation can be purely virtual (integrating sensor models, as in LEIA 1) or, for more realism, can combine the physical system with simulated inputs, as is done in LEIA 2. These simulation tools need to be validated and qualified. In particular, it is necessary to verify on them:

- The repeatability of a test on the same tool.
- The reproducibility of a test from one tool to another (fog/rain characteristics, pedestrian/panel detection).

LNE and Cerema have different tools (for AI systems evaluation) at their disposal that need to be qualified:

- PAVIN fog and rain platform for producing artificial fog and rain [1].
- Cerema noise models for numerical simulation of fog (partial digital simulation):
  - Use of PAVIN Platform data as input initially without fog.
- LEIA 1 and 2 platforms for artificial intelligence evaluation:
  - Replay of videos recorded in the PAVIN Platform in LEIA 2.
  - Full digital simulation (sensor + weather) in LEIA 1.

The following section presents the protocol that was put in place to validate these tools.

2. General definition of the protocol

The proposed protocol is as follows. First, an AI-based algorithm applied to the intelligent vehicle, representative of the state of the art, is chosen. This algorithm will be used as a control for the qualification of the Cerema and LNE simulation tools. A metric to evaluate this algorithm will be chosen. Then identical datasets will be prepared using the different simulation tools available. These datasets will have to include adapted scenarios to evaluate the identified algorithm. In addition, they will include data in clear weather and foggy conditions, but also repeated scenarios to verify repeatability. Finally, the algorithm and the associated metric will be applied to all the datasets. A comparison of the scores obtained for each dataset will allow to verify repeatability and reproducibility from one simulation tool to another.
Among the families of algorithms identified, it seems that visual navigation and road tracking are not very suitable for testing on the PAVIN platform, nor for replay. Therefore, we propose to focus on pedestrian detection or sign recognition. The stereo camera ZED2i (See Figure 9) from StereoLab has been chosen (https://www.stereolabs.com/zed-2i/) and purchased by Cerema for the data acquisition. Indeed, the latter will allow the testing of monocular detection and recognition algorithms (by taking only one channel) but also stereoscopic. This will allow

to propose a database in agreement with the literature. Cerema will also make acquisitions in parallel with a thermal camera. This will allow to label the images of the ZED2i camera in dense fog conditions, thanks to a preliminary geometrical calibration.

Figure 9 : StereoLab's ZED2i camera.

3. Detailed definition of the tests

3.1 Physical tests

The objective of the test scenarios defined by Cerema is to collect videos containing 100 individual pedestrians moving in a scene subjected to various weather conditions (clear weather and two types of fog), lighting conditions (day or night) and seasons using clothing representative of summer or winter. To ensure the repeatability of the measurements, each pedestrian's journey is made twice for each configuration of the scene, weather conditions and pedestrian clothing. The dataset of tests therefore includes a total of **2 runs x 100 pedestrians x 3 weather x 2 lighting = 1200 videos**.

The three types of weather conditions chosen are:

- Clear weather: it allows to have a reference scene without disturbances due to the presence of fog.
- Medium fog: the visibility is of 23 m allowing to modify the general aspect of the objects of the scene by leaving detectable all the elements of the visible scene.
- Heavy fog: the visibility is of 10 m allowing elements of the background to disappear for stereo camera but not for thermal camera.

For each weather condition, there will also be two types of lighting considered:

- Daytime condition with the greenhouse opened on the sides to capture as much natural light as possible (See Figure 10).
- Night condition with the greenhouse totally closed (See Figure 11).

Different objects are placed in the scene to reproduce an urban scene. They remained in the same position for the duration of the tests to ensure reproducibility and to allow comparison of the datasets under different lighting and weather conditions. Here is a list of the objects used:

- Shrubs: A ficus in the background and a large planter with two shrubs in the left foreground.
- Wooden picnic table in the foreground right.
- Orange traffic cones (x3 positioned in line and at equal distance).
- Vehicle (Renault Megane).

- Some traffic signs (Speed limit 60, speed limit 50 and a wildlife crossing sign).
- Ground marking strips: crosswalk and dashed marking.
- Four calibrated targets (a large black and a large grey (50 x 50 cm), a small white and a small black (30 x 30 cm)).

Figure 10: Daytime scene of the PAVIN platform for the PRISSMA tests.

Figure 11: Night scene of the PAVIN platform for the PRISSMA tests.

To add a seasonality in the scene (summer/winter), the pedestrians have been dressed with 3 clothes characteristics of high or low temperatures such as: hats, caps, shorts, pants, coats... and as much as possible, a variability of the color of the clothes has been respected (bright colors, dark or light colors). Wigs have also been used to increase the number of female pedestrians. To break the pedestrian silhouette, accessories have been used to constrain the pedestrian detection algorithms: balloon (soccer and rugby), backpack, computer shoulder bag, tote bag, hiking bag, walking sticks, open or closed umbrella, wooden board, cardboard box, snowboard, green plant, survival blanket, headlamp.

Figure 12: Instrument layout for PRISSMA tests

To obtain a well-characterized dataset, measurements are performed with the following sensors:

- Stereo camera (ZED 2i model) (depth and RGB channels of the image).
- WIFI camera (TAPO C310) (visible image of the scene).
- SWIR camera (Xenics).
- LWIR camera (Xenics).
- Weather sensors of the PAVIN platform (Temperature, humidity, visibility ...).

The first day has been devoted to the installation of the elements of the scene, the instruments and their calibration.

For the night measurements, Audi A3 LED headlights have been added at the level of the cameras respecting the height of a classic vehicle. The headlights were turned on in low-beam as well as for the headlights of the vehicle in the scene. The height and angle of the headlights have been adjusted beforehand to ensure proper road lighting.

The different instruments were positioned at the beginning of the greenhouse (See Figure 12). The bottom of the greenhouse has a transparent rectangular opening covered with a black cover during the night tests. This avoids glare caused by car headlights placed at instrument level.

Estimated data volume:
To facilitate acquisition, the SWIR, LWIR and Wi-Fi cameras have been launched continuously for each half-day of measurements. As for the stereo camera, it was launched for each pedestrian crossing. In addition, to remain as close as possible of an acquisition frequency of 15 images per second, the LWIR camera acquisition frequency has been set to f/2 and the SWIR camera acquisition frequency was set to f/4.

Instrument	Stereo Camera	WI-FI Camera	LWIR Camera	SWIR Camera
Weight/minute (Mo/min)	1045	380	600 (f/2)	235 (f/4)
Tests duration (min)	1200	1680	1680	1680
Total test weight (To)	1.2	0.6	1	0.4

Estimated total test duration:

The database includes 100 different pedestrians (clothes and accessories), moving along an identical route of a duration of approximately 1 minute depending on each pedestrian's walking pace. Each route is repeated twice to test reproducibility. Each pedestrian evolves in the two lighting conditions (day and night) and for the three weather conditions (clear weather, fog visibility of 10 m, fog visibility of 23 m), which corresponds to 2 x 2 x 3 x 100 = 1200 one-minute sequences, i.e., nearly 20 hours of testing. This is not including additional time, such as change of weather conditions (Clear / Fog) or lighting conditions (Day / Night).

3.2 Simulation tests

The tests performed in pure simulation are done by the LNE with the objective to simulate conditions similar to the ones found in reality. The simulation work is mainly done using 4DVirtualiz simulator which is a digital twin software devoted to robotics and the automotive field. However, there are some limitations to this software:

- The availability of some items such as road signs, or the quality of texture on other items.
- The representation of the transparent roof: 4DVirtualiz does not seem to support this transparency, so the roof appears nonexistent.
- The photorealism of reflected light or indirect illumination, this is due to the visual motor used (Ogre3D).

Several simulation scenarios are created by importing the 3D model of the PAVIN environment in 4DV. Animated walking pedestrian models are added in the environment. A model of the ZED2 camera is available in 4DV software, allowing us to retrieve images with the same simulated lens and resolution as the physical tests.

The annotation process is automated: the coordinates of the pedestrians can be retrieved in the simulation and a transformation is applied to calculate the bounding box in the simulated image. This annotation is stored in a csv file with the following columns: (frame number or image name, x,y coordinates of the center of the box, width and height of the box in pixels).

The output of simulated data generation is static images with XML files containing the bounding box of the pedestrian.
The objective of the simulation test is to reproduce the physical test conditions and generate the digital twin of its dataset. However, software and hardware limitations may lead to the generation of a smaller number of images.

Figure 13: Example of image perceived by the simulated ZED2 camera using 4DV software

To compensate for these limitations, we propose to use different simulators in the mixed simulation phase, as described below.

3.3 Mixed tests

To narrow the gap between simulation and physical tests, we propose to use "mixed tests".
In this type of test, the physical device is evaluated using synthetic data presented in a physical manner. In our case, images are projected on a screen in front of the physical camera.
This type of tests offers two main advantages:

- Any 3D environment or simulator can be used, regardless of whether it includes a model of the camera being tested.
- Using the real camera, the generated image has the same lens deformation and acquisition defects as in physical tests, reducing the simulation-to-reality gap.

4. Database

The database will finally include, as represented on Figure 14:

1. *'Real_Clear_0'* set of images acquired within the Fog and Rain Platform, under favorable weather conditions, day and night, containing different scenarios with pedestrians and traffic signs. In daylight conditions, special attention have been paid to the background of the room to make it representative of a clear realistic background.

2. *'Real_Clear_1'* set identical to 'Real_Clear_0' acquired under exactly the same conditions. This repetition of the base will verify the repeatability of the nominal scenario.
3. *'Real_MediumFog_0'*, *'Real_MediumFog_1'*, *'Real_HeavyFog_0'*, *'Real_HeavyFog_1'* sets identical to *'Real_Clear_0'*, but with medium and heavy fogs of different densities. This repetition (0 or 1) of the base will verify the repeatability of the fog.
4. *'CeremaNoiseModel_MediumFog'* and *'CeremaNoiseModel_HeavyFog'* sets identical to *'Real_Clear_0'*, but with an addition of simulated fog on the initially fog-free images of *'Real_Clear_0'*. This set will allow to check the validity of the *CeremaNoiseModel* against the Real fog sets.
5. *'LeiaReplay_Clear'* set identical to *'Real_Clear_0'*, by placing the camera in front of a replay of *'Real_Clear_0'* in the LEIA 2 platform.
6. *'LeiaReplay_HeavyFog'* and *'LeiaReplay_MediumFog'* sets identical to *'Real_Fog'*, by placing the camera in front of a replay of *'Real_Fog'* in the LEIA 2 platform.
7. *'LeiaSimul_Clear'* set identical to *'Real_Clear_0'*, by placing the camera in front of a full simulation of *'Real_Clear_0'* in the LEIA 2 platform.
8. *'LeiaSimul_MediumFog'* and *'LeiaSimul_HeavyFog'* identical to *'Real_ Fog'*, by placing the camera in front of a complete simulation of *'Real_Fog'* in the LEIA 2 platform

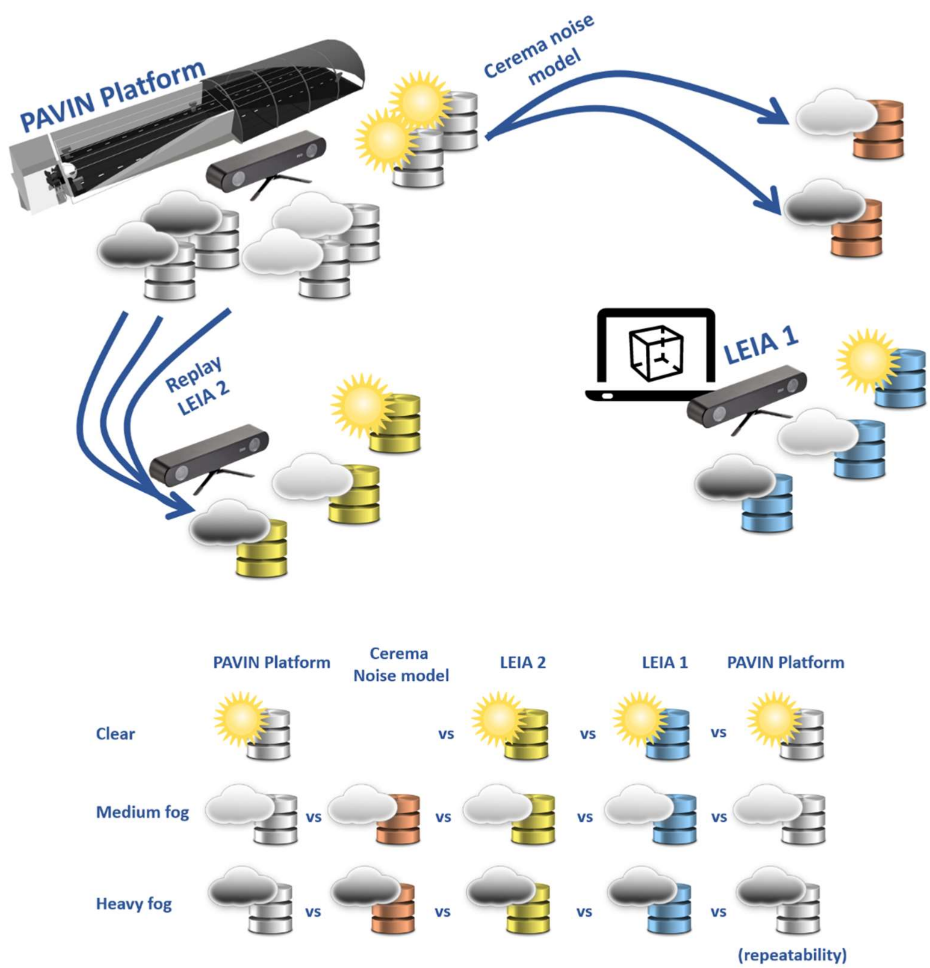


Figure 14: Overview of the database

5. Ongoing work

All the tests in the PAVIN platform have been carried out in October 2022. The post-processing of the data has consisted in a hand labeling of the pedestrians by Cerema. For this purpose, Cerema has developed a hand-labeling tool. It has been used to do a precise hand labeling of the pedestrians in the 'Real' database images which will be used as the ground truth. The hand labeling of the images with fog was carried out using the thermal camera data and then transposed on ZED 2i camera images. Indeed, in foggy weather and in front of a distant pedestrian, the pedestrian is sometimes invisible on the camera ZED 2i. Going through the thermal camera allows to get the ground truth.

LNE will replay the dataset into Leia2, and simulate the dataset into Leia1, thanks to the 3D digital twin of the PAVIN platform.

Cerema applied the object detection algorithm so called YOLOv3 on the *'Real'* dataset and will apply it on the other datasets (*'CeremaNoiseModel'*, *'LeiaReplay'*, *'LeiaSimul'*). Different metric such as precision and recall (on intersection of union) will be used to evaluate the impact of fog on this widely used object detection algorithm. In a second step, the scores obtained by the algorithm in the different cases (*'Real'*, *'CeremaNoiseModel'*, *'LeiaReplay'*, *'LeiaSimul'*) will allow to verify if the numerical and artificial fog simulations are reliable and valid.

6. Expected results

Different results are expected from these different datasets. First, the '*Real*' database will be used to develop the *'CeremaNoiseModel'* models (part of T2.3 of PRISSMA project). Cerema will work with partially physic-based model to develop a model able to reproduce fog on images initially acquired without fog.

Thanks to the defined metrics, Cerema will then make the comparison of the scores obtained by the AI algorithm between *'Real'*, *'CeremaNoiseModel'*, *'LeiaReplay'* and *'LeiaSimul'* datasets. This will allow to verify if the different simulation tools are well correlated between them. Also, the analysis of the variation of the scores on the 'Real_0' and 'Real_1' variants will allow to verify the repeatability of a test on the Cerema platform (with or without fog).

By the way, the variation of scores between *'Clear'* and the two densities of *'Fog'* will highlight the need for further development of the algorithms in difficult conditions such as fog. Finally, the database will be made public via the PRISSMA project, in order to feed future research.

8. Reference

[1] M. Colomb, K. Hirech, P. André, J. Boreux, P. Lacôte, and J. Dufour, "An innovative artificial fog production device improved in the european project "fog"," *Atmospheric Research*, vol. 87, no. 3, pp. 242–251, 2008, third International Conference on Fog, Fog Collection and Dew. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169809507002037

Chapter 3: INRIA/TRANSPOLIS

1. Presentation

The goal of this proof of concept is to demonstrate Inria autonomous platform [1] (figure 13) to showcase the validation [2] of its perception software stack [3] (figure 14) within generated scenarios from [4], using augmented reality to inject dynamic obstacles in the scene [5]. The demonstration site is Transpolis Fromentaux City area (figure 15) the test preparation and augmented reality simulation will be implemented in the digital twin (figure 16).

Figure 15: Inria autonomous platform based on an electric car

5.1 Experimental platform

For the experiments, a Renault Zoe car (figure 5) has been equipped with a Velodyne HDL-64 on the top, 3 Ibeo Lux LIDAR's on the front and 1 on the back, Spectra SP90 RTK Dual antenna GNSS, Xsens IMU providing vehicle velocity and orientation, a stereo camera and 2 IDS cameras. Data from LIDAR's are fused and synchronized using the IBEO fusion box. The perception system described earlier has been implemented on a PC in the trunk of the car, equipped with a NVidia Titan X GPU, while the previously described automation process has been integrated in the vehicle.

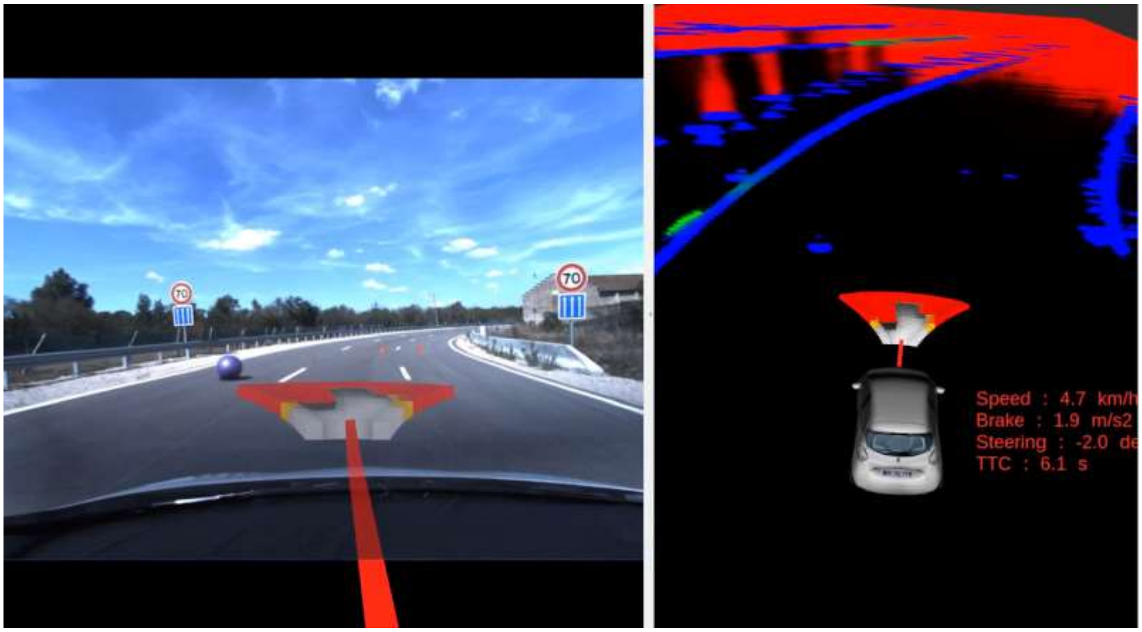


Figure 16: Inria autonomous software running aboard the vehicle

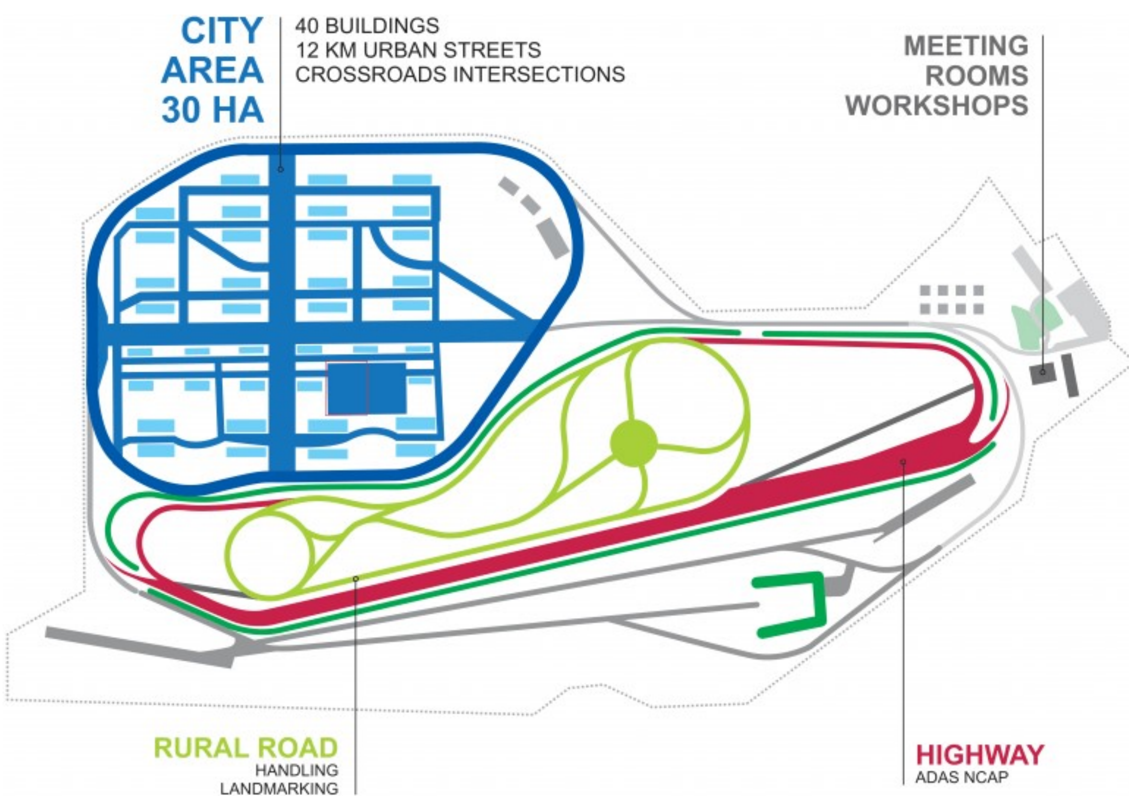


Figure 17: Transpolis Fromentaux testing facility

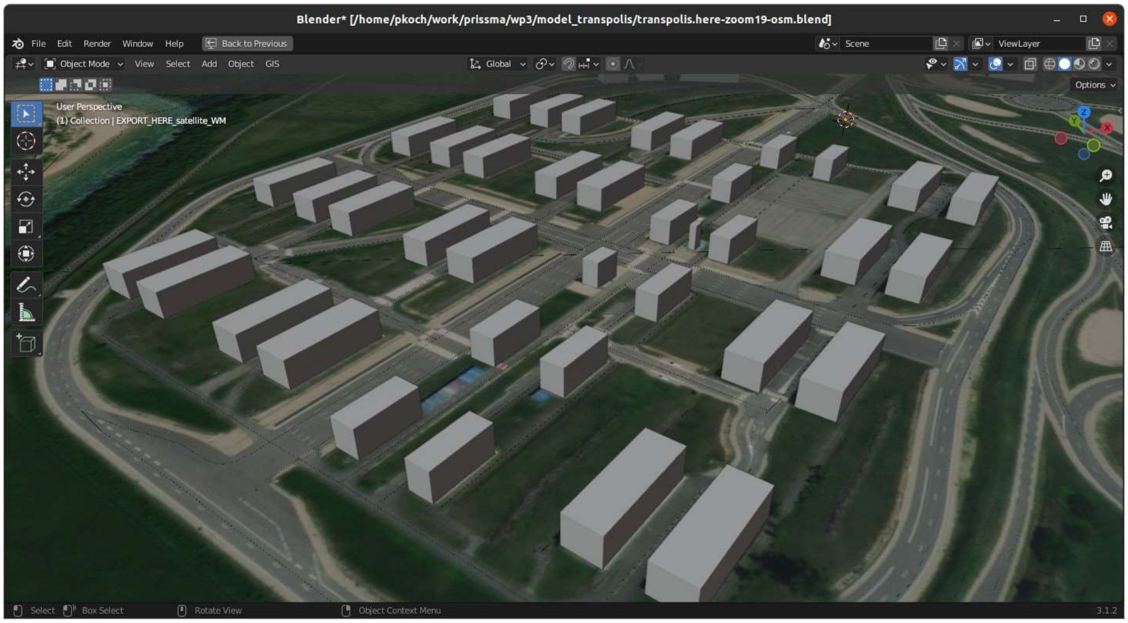


Figure 18: Transpolis digital twin

Figure 19: Experimental Platform: Renault Zoe car equipped with Velodyne HDL-64, 4 Ibeo Lux Li-DARs, Xsens IMU and cameras, and a crash test dummy crossing the dedicated street for the experi-ments.

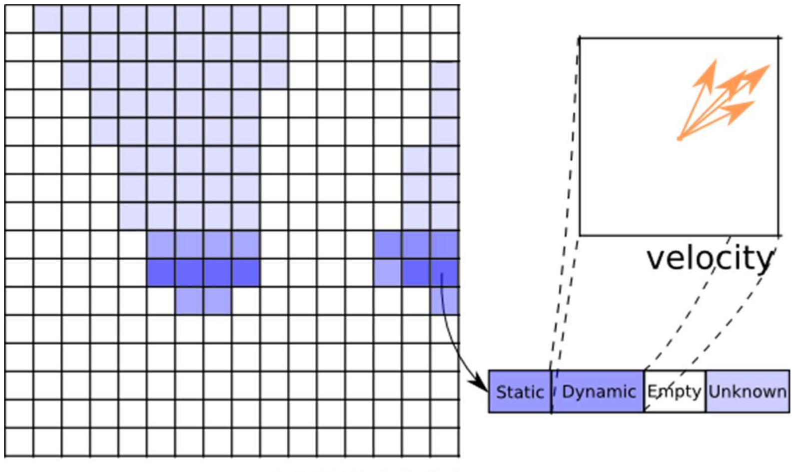


Figure 20: CMCDOT grid structure

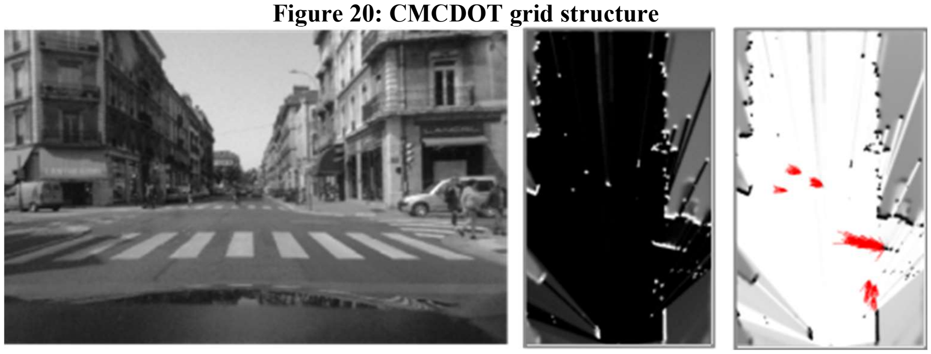


Figure 21: CMCDOT illustration

1.2 CMCDOT

The CMCDOT framework is a broad perception system, based on Bayesian filtering of dynamic occupancy grids (CMCDOT), allowing parallel estimation of occupancy probabilities for each cell of a grid, inference of velocities, collision risk prediction and dynamic object segmentation. From various heterogeneous sensor data, ground form is estimated, instantaneous occupancy grids are generated and filtered using hybrid sampling methods (classic occupancy grids for static parts, particle sets for parts dynamics), into a Bayesian unified programming formalism. Based on this perception framework, navigation systems have been developed and integrated, allowing path finding-and-following, dynamic obstacle avoidance, localization, thus automation of various mobile robots. Also included are communication tools, allowing data fusion from infrastructure systems. The software is composed of ROS packages, which encapsulate the optimized core system on GPU NVidia (Cuda), allowing real-time application on embedded boards (Tegra X2). First developed in an automotive setting, it is now exploited in other areas of mobile robotics, and are particularly suited to highly dynamic and uncertain environment management. Thanks to an important engineering support over the years (notably thanks to IRT Nanoelec), this software has grown to be a core research and development tool of the team, an important technology demonstration and transfer vector, through maintained experimental platforms (most notably automated Zoe) and associated research contracts and software licensing with industrial partners.

1.3 Augmented Reality

On the Gazebo simulator, CHROMA has developed a virtual twin of its Renault Zoe experimental vehicle. This virtual twin generates the same outputs (sensors messages, localization) that the actual vehicle does and reacts to the same commands and has a realistic kinematic and dynamic behavior. This allows to test software in Software-in-the-Loop and Hardware-in-the-Loop. CHROMA has also developed an Augmented Reality framework (figure 20) for testing and validation of software on the Renault Zoe experimental vehicle. This framework provides a flexible way to introduce any virtual element in real time in the data of the LiDAR sensors of the vehicle. Our Augmented Reality accurately handles all possible occlusions between real and virtual elements. The representability of tests scenes generated by the augmented reality framework has been experimentally proven. It is then possible to easily and safely place the whole vehicle and all its software, from perception to control, in hybrid but realistic test scenes. This new testing methodology is intended to be a bridge between Vehicle-in-the-Loop and real-world testing.

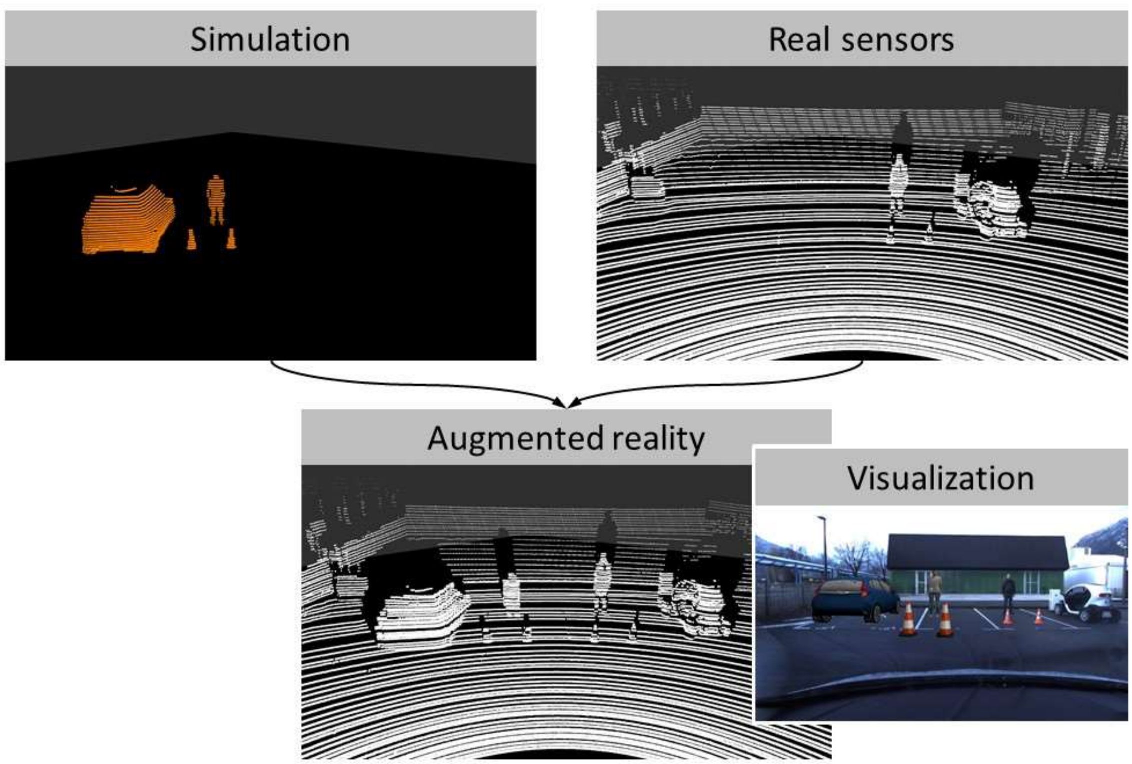


Figure 22: CHROMA Augmented Reality framework

2 Test plan

```
{
  "actors": {
    "zoe": ["zoe/urdf/zoe_gpu.urdf", 30, 5, 0.35, -3.1416],
    "bus01": ["model://bus/meshes/bus.obj", -5, 30, 0, -1.5708],
    "car01": ["model://hatchback_red/meshes/hatchback.obj", -10, 50, 0, -1.5708],
    "car02": ["model://hatchback_blue/meshes/hatchback.obj", 5, -50, 0, 1.5708],
    "truck01": ["model://fire_truck/meshes/fire_truck.obj", 10, -20, 0, 1.5708],
    "person01": ["model://person_walking/meshes/walking.dae", -10, -25, 0, 0]
  },
  "waypoints": [
    [15, 5, 5, "zoe/base_link", "map", {
      "bus01": [-5, 30, 0, -1.5708, 2],
      "car01": [-10, 50, 0, -1.5708, 6],
      "car02": [5, -50, 0, 1.5708, 4],
      "truck01": [10, -20, 0, 1.5708, 0.5],
      "person01": [-10, -25, 0, 0, 1]
    }],
    [-20, 5, 5, "zoe/base_link", "map", {
      "bus01": [-5, 30, 0, -1.5708, 0],
      "car01": [-10, 50, 0, -1.5708, 0],
      "car02": [5, -50, 0, 1.5708, 0],
      "truck01": [10, -20, 0, 1.5708, 0],
      "person01": [-10, -25, 0, 0, 0]
    }],
    [-30, 5, 5, "zoe/base_link", "map"],
    [-30, -5, 5, "zoe/base_link", "map"],
    [-15, -5, 5, "zoe/base_link", "map", {
      "bus01": [-5, 30, 0, -1.5708, 2],
      "car01": [-10, 50, 0, -1.5708, 6],
      "car02": [5, -50, 0, 1.5708, 4],
      "truck01": [10, -20, 0, 1.5708, 0.5],
      "person01": [-10, -25, 0, 0, 1]
    }],
```

Figure 23: Scenario description in JSON format

gazebo_scenario is a software to run a defined scenario from a descriptive input file, containing actors, way-points, and actors' motions (figure 22). The goal of the scenario (figure 21) is to cross virtual dynamic obstacle path to produce safe collision or near collision in augmented reality. In this example the ego vehicle is driving toward a crossroad, crossing the path of other road users, the framework allows us to finely define way-points at which the ego vehicle trigger the motion of the virtual actors, to produce intended collision in augmented reality within the simulation environment which are then injected as sensor input in the autonomous driving software onboard the ego vehicle itself, in a hybrid vehicle-in-the-loop manner.
The figure 23 shows the simulation running in Gazebo on the left, and the ROS visualization software Rviz on the right, in which the view-port is a 3rd person view of the ego-vehicle, over the perception grid. We can observe the dynamic obstacles, such as the bus on the right-hand side of the ego vehicle. The color represents the estimated speed of the obstacles. The first row of screenshots represents the initial step of the scenario, the second row is taken at the mid-term, in a critical situation, a near collision, or a virtual collision in case of perception latency.
The scenario management software take care of sending the path to the ego-vehicle's global planner. It deals with way-points synchronization to execute the plan at each step defined in the description. A way-point is a 2D (x, y) landmark relative to the map model in Cartesian coordinate, the relation between this coordinate system and the world is let to the digital twin responsibility, it should be geo-referenced with regards to a global coordinate system (e.g., WGS84). It is important to note that this geo-registration of the proving ground's digital twin is critical since the actors' trajectories are relative to this reference, in particular, orientation

errors θe can greatly alter the quality of augmented reality interactions. Therefore, a verification must be done on the proving ground, and a calibration might be necessary to converge towards a spatially coherent relationship.

3. Next step

- Data analysis
- Metrics refinement
- Improve perception tuning

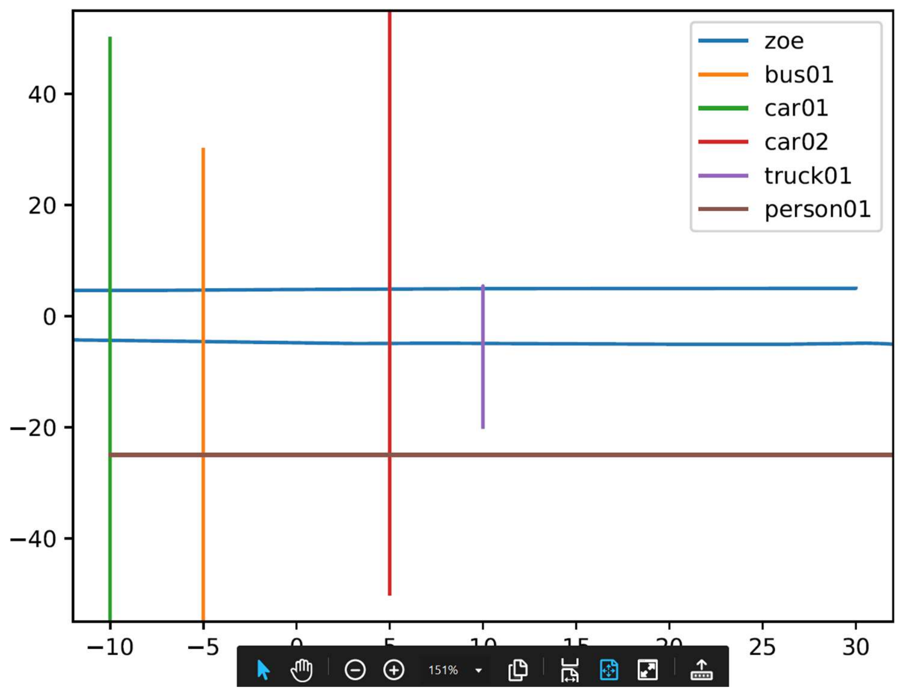


Figure 24: Actors trajectories

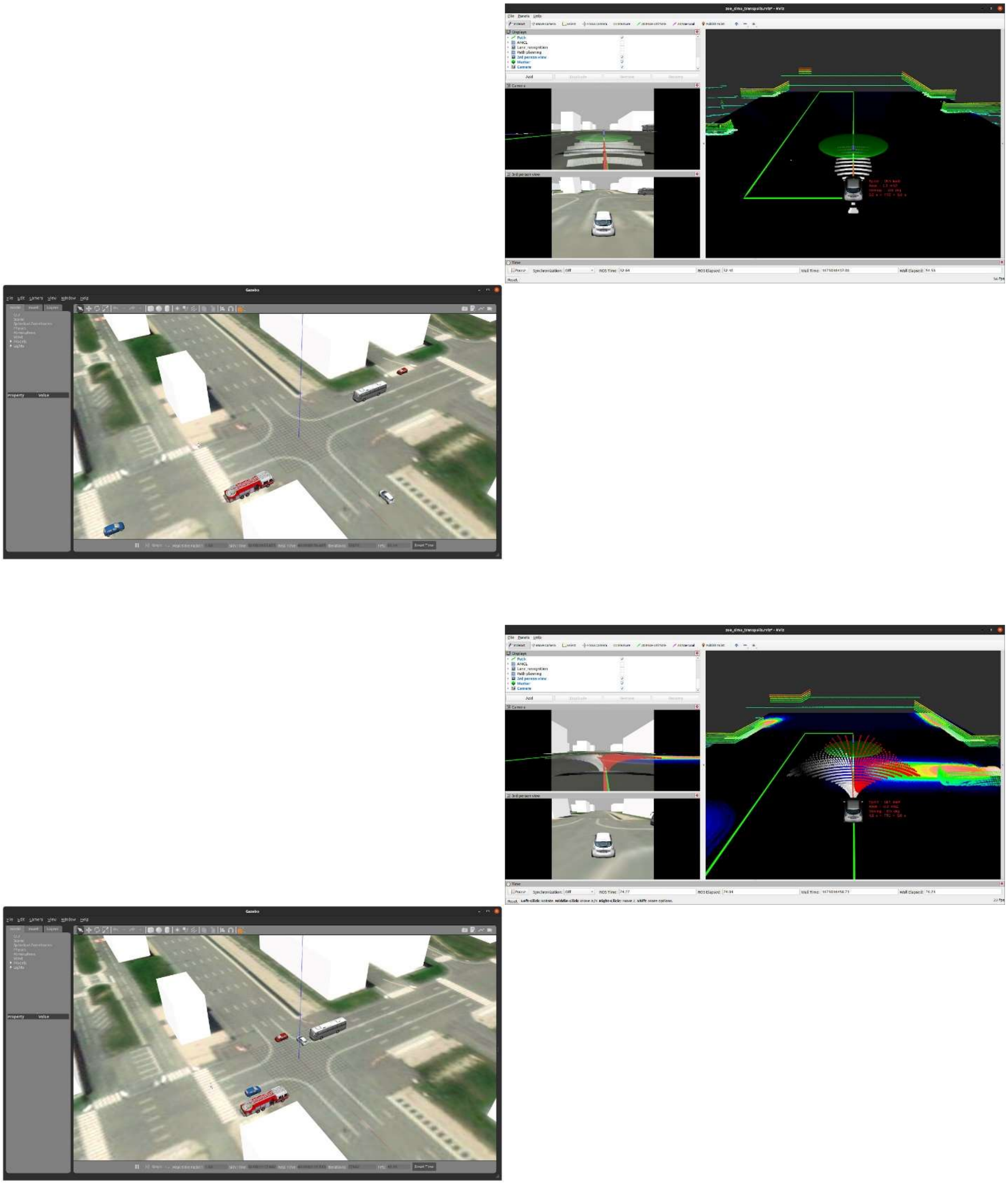


Figure 23: ROS RViz and Gazebo screenshots

4. References

[1] L. Rummelhard, J. Lussereau, J.-A. David, C. Laugier, S. Dominguez, G. Garcia, and P. Martinet, "Perception and Automation for Intelligent Mobility in Dynamic Environments," in

ICRA 2017 Workshop on Robotics and Vehicular Technologies for Self-driving cars, Singapore, Singapore, Jun. 2017. [Online]. Available: https: //hal.inria.fr/hal-01592566

[2] P. Ledent, A. Paigwar, A. Renzaglia, R. Mateescu, and C. Laugier, "Formal Validation of Probabilistic Collision Risk Estimation for Autonomous Driving," in *CIS-RAM 2019 - 9th IEEE International Conference on Cybernetics and Intelligent Systems (CIS) Robotics*, Automation and Mechatronics (RAM). Bangkok, Thailand: IEEE, Nov. 2019, pp. 1–6. [Online]. Available: https://hal.inria.fr/hal-02355551

[3] L. Rummelhard, A. N`egre, and C. Laugier, "Conditional Monte Carlo Dense Occupancy Tracker," in *18th IEEE International Conference on Intelligent Transportation Systems*, Las Palmas, Spain, Sep. 2015. [Online]. Available: https://hal.inria.fr/hal-01205298

[4] J.-B. Horel, C. Laugier, L. Marsso, R. Mateescu, L. Muller, A. Paigwar, A. Renzaglia, and W. Serwe, "Using Formal Conformance Testing to Generate Scenarios for Autonomous Vehicles," in *DATE/ASD 2022 - Design, Automation and Test in Europe - Autonomous Systems Design*. Antwerp, Belgium: IEEE, Mar. 2022. [Online]. Available: https://hal.inria.fr/hal-03516799

[5] T. Genevois, J.-B. Horel, A. Renzaglia, and C. Laugier, "Augmented Reality on LiDAR data: Going beyond Vehicle-in-the-Loop for Automotive Software Validation," in *IV 2022 - 33rd IEEE Intelligent Vehicles Symposium IV*. Aachen, Germany: IEEE, Jun. 2022, pp. 1–6. [Online]. Available: https://hal.inria.fr/hal-03703227

Chapter 4: VALEO/IGN POC

1. Context & Document objectives

Automated driving systems for SAE L4 vehicles (e.g. autonomous shuttles, delivery robots, etc.) may include a localization system.
This system is in charge of providing:

- an **"absolute" position** of the vehicle ("position on Earth"), e.g. latitude/longitude/altitude in a global reference frame (e.g. WGS-84)
- and/or a **"relative" position** ("position on a digital map"), e.g. "on A-street, in lane 1 (at +15cm from lane center), at +2302cm from intersection "B" center.

Other data may also be returned, such as orientation/heading, speed, attitude (yaw, roll, pitch) or time.

Such absolute or relative position may be determined through **various methods/approaches/technologies**.
For example, the SAFAD whitepaper[2] distinguishes the 2 below approaches:

- GNSS-BASED LOCALIZATION
  "This approach consists of GNSS, odometry and correction services to achieve precise global coordinates, and matching GNSS measurements to an HD map to obtain a relative position on the map"
- ENVIRONMENT-PERCEPTION-SENSOR-BASED LOCALIZATION
  "Based on a rough global coordinate obtained by GNSS and odometry, this approach matches real-world features (such as natural or artificial landmarks) or point clouds detected by Environment Perception Sensors with respective features or point clouds on an HD map to localize the automated driving system on the map"

Also, localization systems may rely on **various types of AI** (according to the taxonomy defined by ConfianceAI project):

- Knowledge-based AI
- Data-driven AI
- Hybrid AI

[2] Safety First for Automated Driving, 2019

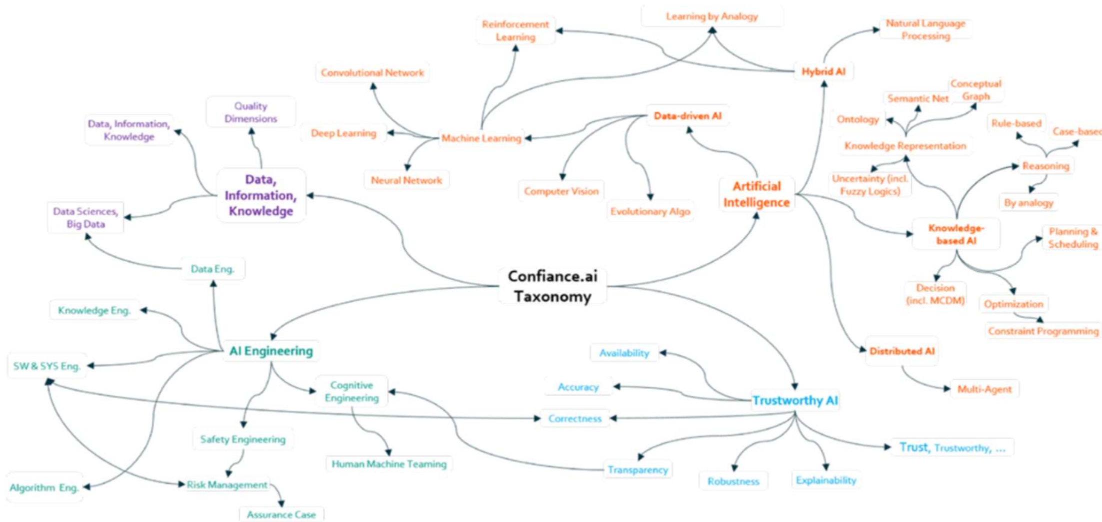


Figure 24: Taxonomy of AI according to ConfianceAI project[3]

Moreover, such absolute/relative position of vehicle may be **used in different ways** within automated driving systems, depending on ADS manufacturers (e.g. to check if the vehicle is within its ODD, etc.).
Consequently, the necessary performance of the localization system may differ according to vehicles and ADS manufacturers.

Hence, this document aims at describing :

- A protocol of test, characterization & validation of such localization systems used by automated vehicles, **in a controlled environment (e.g. test track)**
- A ground truth system enabling such performance & compliance assessment.

Considering that such protocol & ground truth system should be **agnostic** to:

- Approaches/technologies used to determine the absolute/relative position
- Types of AI employed by the localization system
- Usages of localization data within the automated driving system & associated expected performance (in terms of accuracy, integrity, etc.)

2. Physical quantities to consider

To assess performance of the "localization system under test", the target physical quantity to be considered will be the **absolute position** (expressed in a global reference frame, e.g. WGS-84 latitude/longitude/altitude).

Coordinates will be expressed in the legal national reference frame RGF 93 (https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000038203565).

[3] ConfianceAI project (whitepaper, 2022)

The expected precision & accuracy are centimetric.
All the data will be provided in 3 dimensions.

NOTES :
Assessment of **relative position** may be investigated further (but implies additional complexity, e.g. requires the use of a HD Map of the test track, etc.).--> excluded from the 2024 POC

Attitude (yaw, roll, pitch)
These metrics will be natively generated by the IGN during the experimentation process. Precision is depending on the chosen model of the repartition and numbers of the targets on the vehicle.

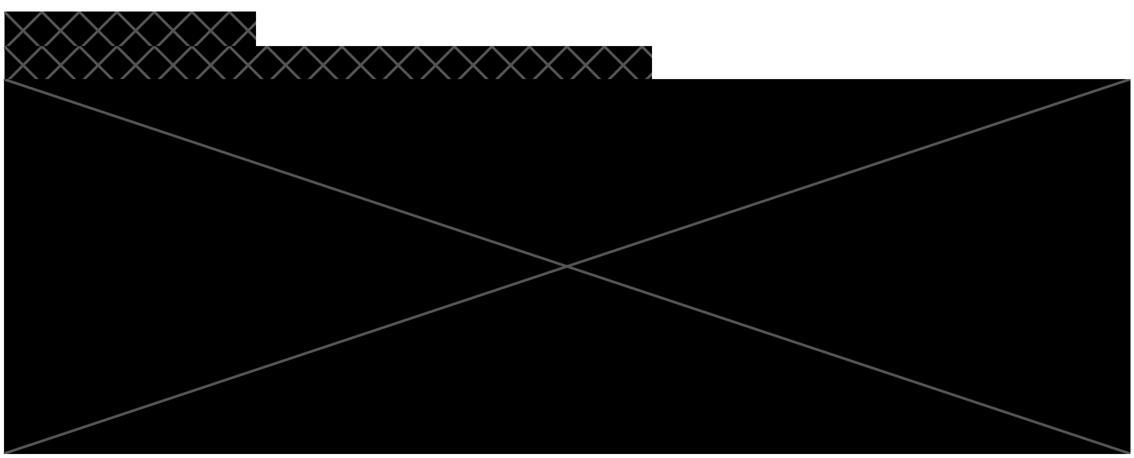
3. Localization system under test

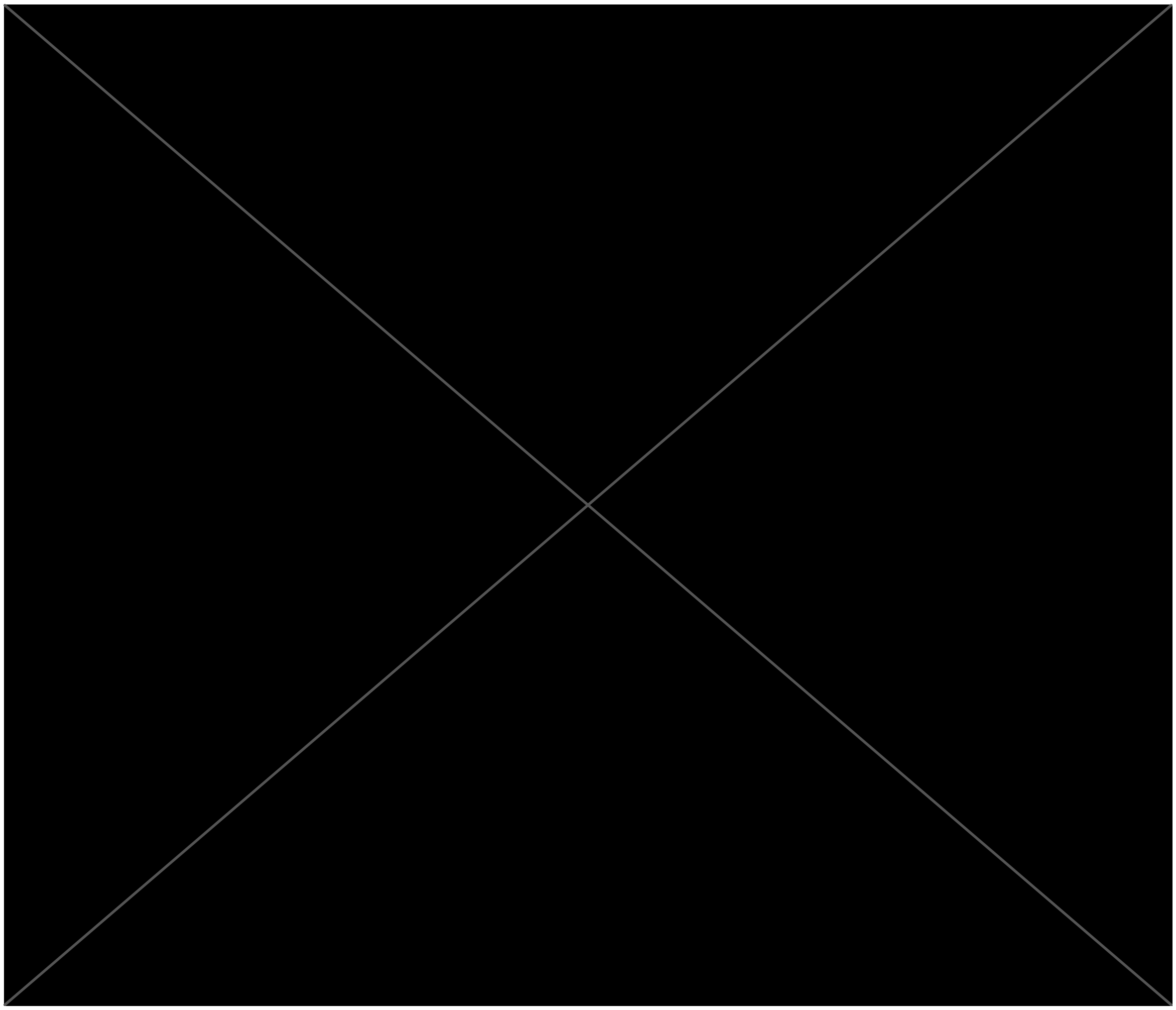
3.1 Data to return

The localization system under test shall determine and return:
- The **absolute position of the vehicle** with an associated timestamp (based on GNSS time synchronization);
- Heading & altitude.

4. Preparation of the assessment

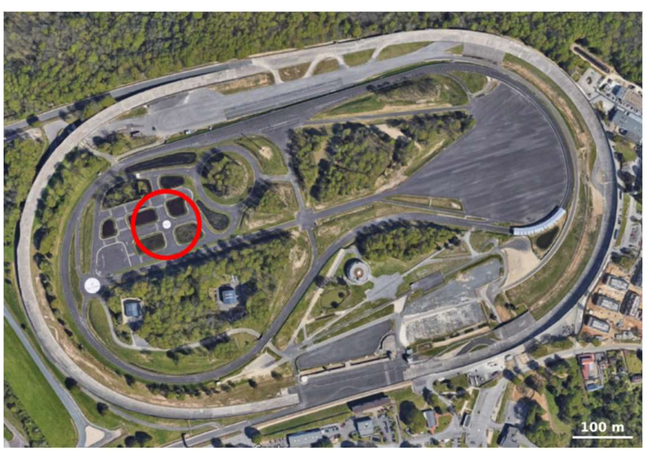
If a specific preparation is necessary to enable the performance assessment, the provider of the localization system for automated driving will have to arrange with the test track manager (and the ground truth system manager).
For example, the localization system under test may need some specific artificial landmarks to be deployed by the test track manager.





5. Test conditions

The POC will take place on the closed road of the UTAC test ground (Montlhery).
The roundabout is the test location that will be equipped with the cameras.



- Climatic & luminosity conditions:

Optimal visibility is required to conduct the photogrammetric acquisition of the ground truth system: daylight, no fog or rain, no GNSS masks.
Optimal, i.e., daylight, good weather (no rain/fog)

- Driving scenarios:
- Scenarios related to roundabout crossing

Prerequisites:

- Max speed of the vehicle to determine according the precision of the trajectography.

The speed of the vehicle is a limiting factor, in relation with the ability of the cameras (specifications = 5-7 images per sec). It will be part of the experimentation to test different speed scenarios.

- The vehicle should always be visible by minimum 4 cameras according with the illustrations (cf next part)

NOTE: Further possible improvement:

- Test/assessment in all conditions within the ODD[4] of the localization system/automated vehicle, including degraded conditions (e.g. night/rain, etc.) => in that case, the ODD supported by the localization system shall be declared by its manufacturer

The IGN protocol for providing the ground truth system cannot be extended to be tested in degraded visibility conditions (the acquisition of specific waterproof material would be necessary).

6. Post-processing of measurements

6.1 State of the art: possibly relevant standards

EN 16803	Space – Use of GNSS-based positioning for road intelligent transport systems (ITS)
ISO/IEC 18305	Information technology – Real time locating systems – Test and evaluation of localization and tracking systems

[4] ODD : Operational Desing Domain

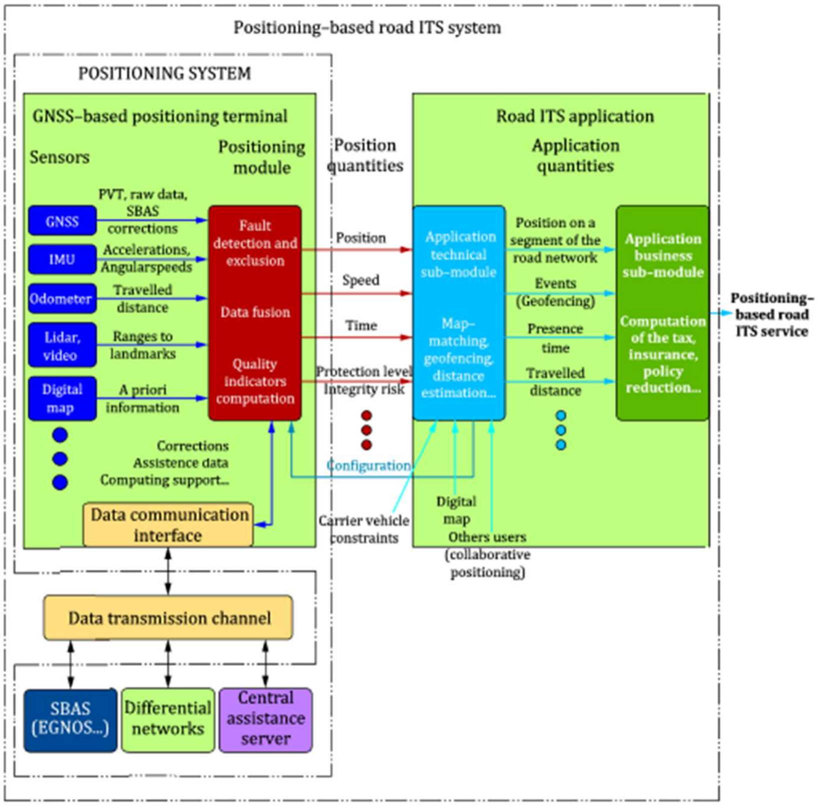


Figure 27: Generic architecture of a road ITS system (EN 16803-1)

6.2 Figure of merit/metrics to consider

The figure of merit to consider (and their associated expected performance) will have to be declared by the localization system manufacturer.

For the POC, the **accuracy** of the absolute position returned by the localization system may be assessed/characterized:

- Horizontal lateral & longitudinal error between the true absolute position (returned by the ground truth system) and the absolute position estimated by the localization system under test
- Mean & standard deviation of the error distribution
- Alternatively, $50^{th}/75^{th}/95^{th}$ percentiles of the error cumulative distribution function.

NOTE:
Other metrics to characterize may be integrity or availability for instance.

6.3 Methods of post processing

This part will be defined afterwards.

7. Ground truth system

The positions and potentially attitudes of the vehicle will be provided according the legal national reference frame RGF93 : geographic (latitude, longitude, height) and projected (easting/northing/elevation)

Ground truth system is defined by the precise implantation of 12 cameras.

The photogrammetric process will enable the definition of the vehicle positions according to a network of targets implanted on the vehicle.

Example of targets implantation on a vehicle

Four sets of 3 cameras are implanted on a precise location and will be synchronized using GNSS precise time.

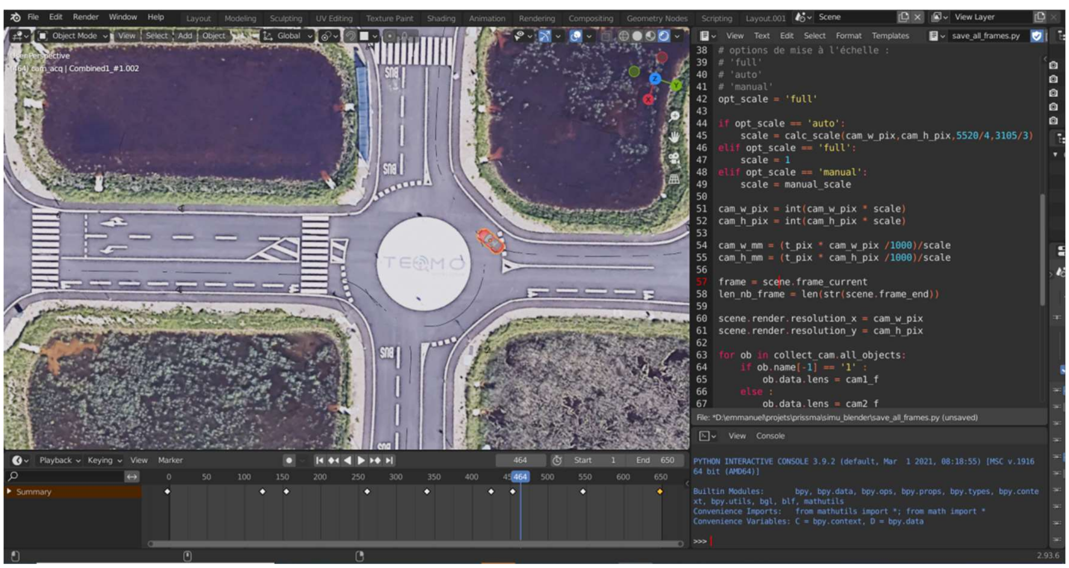
System configuration: coverage of the cameras and data and synchronization links

Process:

2022:
Choice of equipment through calculi and simulations

2023: simulation with Blender to determine the optimal configuration of cameras and targets. The simulation also provides expected accuracy of metrics

2024: onsite POC with Valeo vehicle on UTAC test ground (scenarios, images acquisition).



Chapter 5: SPHEREA POC

1. Context & Document objectives

In terms of V&V campaign, the major problem of AI-based autonomous vehicles is to be able to run enough tests that are representative of the system real operational conditions usage. The qualification of any critical system requires that this system be validated in an operational environment or in an environment deemed equivalent by the stakeholders in charge of verifying the capabilities of the system. This is the case, for example, for Euro NCAP homologation, which consist in reproducing, on test tracks, situations that allow the evaluation of the vehicle's behavior in conditions similar to its use on open roads.
The major problem of the AI-based autonomous vehicle is that the full operational design domain (ODD) cannot be reproduced on a test track. Indeed, some feared events occur only very rarely during the life cycle of a vehicle, and the human consequences are sometimes very serious. It is obviously not possible to put dozens of people and vehicles in danger to reproduce these situations, so the approach is to separate the tests of an autonomous vehicle in the following three categories:

- In its final operational environment (WP4)
- On test track (WP3)
- In simulation (WP2)

The planned test strategy is to segregate the tests according to the possibilities offered by each of these environments:

- realize in operational environment all the possible tests with the real vehicle, according to and what is allowed to do without endangering the goods and the persons around the track
- realize on test tracks the tests impossible to do in operational environment, but still applied to the real vehicle
- finally do extensive test in simulation

In the end, it is hoped that the whole ODD will be covered, with enough tests with the real vehicle

SPHEREA makes the hypothesis that the demonstration of the complete coverage of the ODD by juxtaposition of the tests in Open environment and in closed road will be sufficient only if the simulation can, by using technologies such as the augmented reality and the digital twins, bring on closed road many situations at risk.

In this hypothesis, without testability constraints applied to autonomous vehicle suppliers, the safety demonstration of AI-based autonomous vehicles will only be possible on highly controlled operational domains such as a highway reserved for communicating vehicles, or well-defined routes at limited speeds. Otherwise, many cases with real vehicle will be missing. To be able to create enough dangerous situations representative of the vehicle's operational domain (ODD) on closed roads, the strict black box approach will not be sufficient.

Example: too rarely will natural test conditions allow to evaluate the real behavior of the vehicle when its sensors are subjected to multiple disturbances (cold, fog, electromagnetic reflections) including heavy vehicle traffic.

2. Testability analysis

Introduction

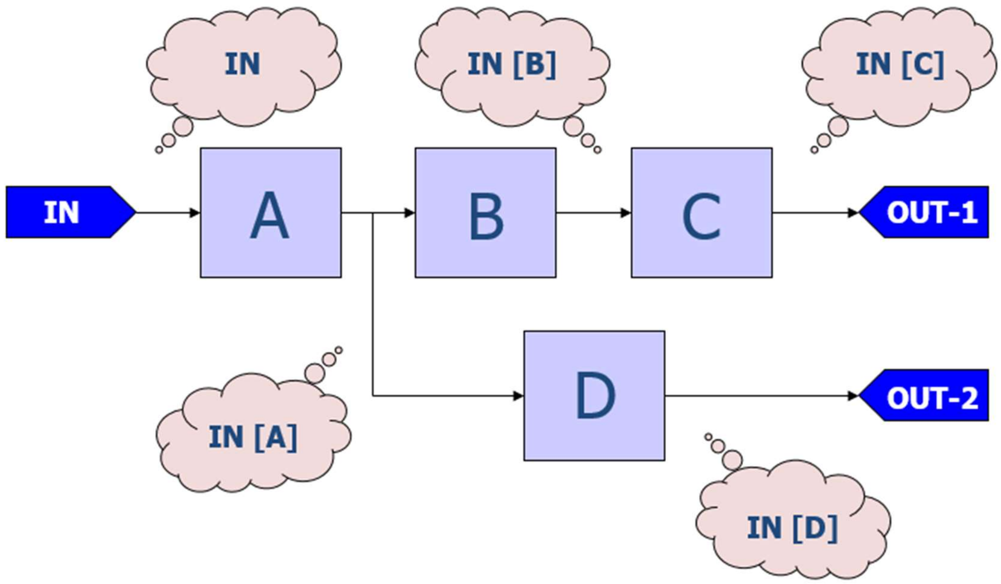
In system engineering, testability is a critical aspect that ensures the system functions as intended and meets the desired requirements. Testability refers to the ease with which a system can be tested and evaluated to validate its functionality, performance, and reliability. In complex systems, such as those found in aerospace, defense, and transportation industries, testability is of utmost importance.
By realizing fault tree analysis, a technique used to identify and evaluate the causes of system failures, testability engineering software provide a set of features that make it easy to develop and test fault tree models, including the ability to define logic gates, analyze data, and visualize results.
One of the major benefits in using such methodology is to detect if a given fault can be isolated by enabling, as soon as the design of the system, much more possibilities to realize some tests. The known impact of poor testability analysis is the low availability of the system in operational conditions, for example:

- NH90 Helicopter: The NH90 is a medium-sized, twin-engine helicopter used by several militaries around the world. However, the NH90 has experienced issues with operational availability, with some reports indicating that the helicopter's availability rate was as low as 25% in certain cases due to a range of technical and logistical issues.
- Siemens Velaro high-speed trains: The Velaro is a family of high-speed trains used in several countries around the world. However, the trains have experienced issues with operational availability due to maintenance difficulties, including problems with the trains' complex systems, such as the train control system and the traction system.
- M1 Abrams Tank: The M1 Abrams is a main battle tank used by the US Army. While the M1 has been in service for several decades, it has experienced issues with operational availability, with reports indicating that only 58% of the fleet was mission capable due to maintenance issues in 2020.

Black box functional chains analysis

In a black box testing approach, the test can only affect inputs and measure the effects on the outputs. In the example below, 2 functional chains are possible:

- FC1: In → A → B → C → D → OUT-1
- FC2: In → A → D → OUT-2



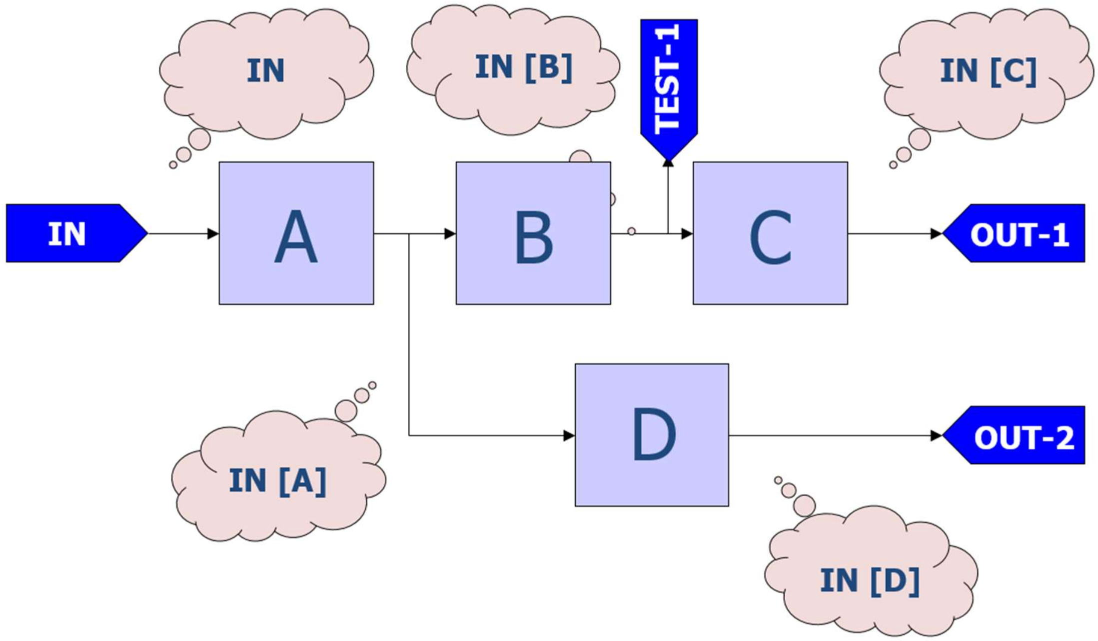
If nothing is added to the design of the system, the isolation of faults of B or C is impossible, resulting in increase of time to repair and therefore a drop in the availability of the resulting system.

Example: If both FC1 and FC2 are invalid, then it is possible to isolate the defect immediately to component A, which is the only component in common with FC1 and FC2. But if FC2 is valid and FC1 is invalid, the only inference that can be made is that B or C have a defect, but it will be impossible to isolate without removing B or C and replace with a known operational component.

Grey box functional chains analysis

Using the result of fault tree analysis, the design decision is to add a TEST-1 point of measure to enable a new chain FC3. The TEST-1 interface is useless during operations of the system, and is used only during the other phases of the lifecycle of the system like design or maintenance:

- FC1: In → A → B → C → D → OUT-1
- FC2: In → A → D → OUT-2
- FC3: In → A → B → TEST-1

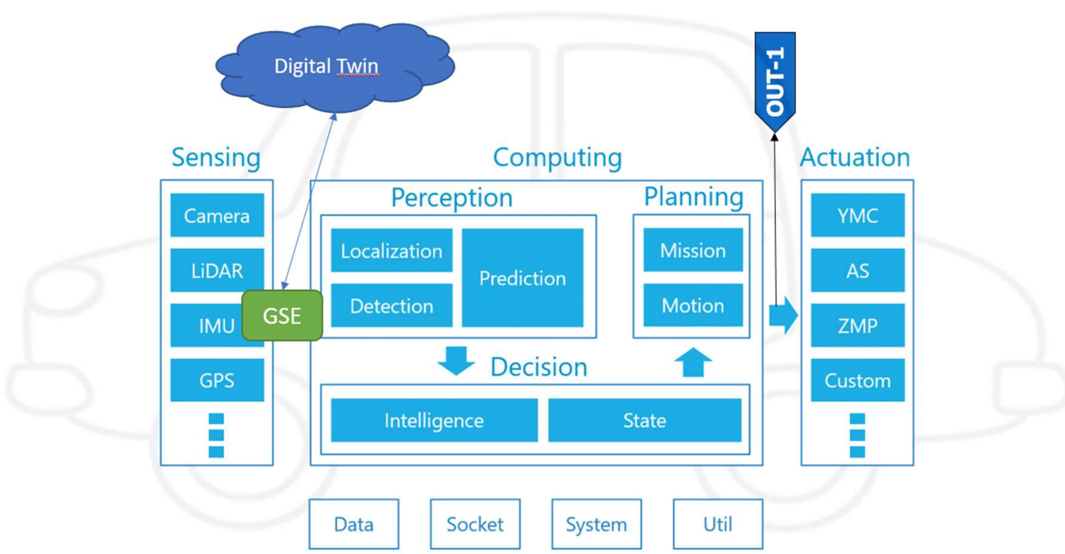


In our previous example, where FC2 is valid, but FC1 is valid, the FC3 enables to immediately isolate the component with defect:

- If FC3 is valid, then C has the defect
- If FC3 is invalid, then B has the defect

Application to the test of ATRS: Vehicle in the loop

Enabling near-accident situations at a large proportion will need special instrumentation. In the field of testability, it is the use of additional break-out test point that enables to add, to the images captured by the perception stack, the perturbations required by the test.



In the new industry of autonomous driving vehicle, this approach is called Vehicle-in-the-loop (VIL). According to Euro NCAP, Vehicle-in-The-Loop is intended to be included in future test protocols. Therefore, some OEMs, Tiers-1, Simulation software providers have launched studies and Proof of concept to explore Vehicle in the loop concept.

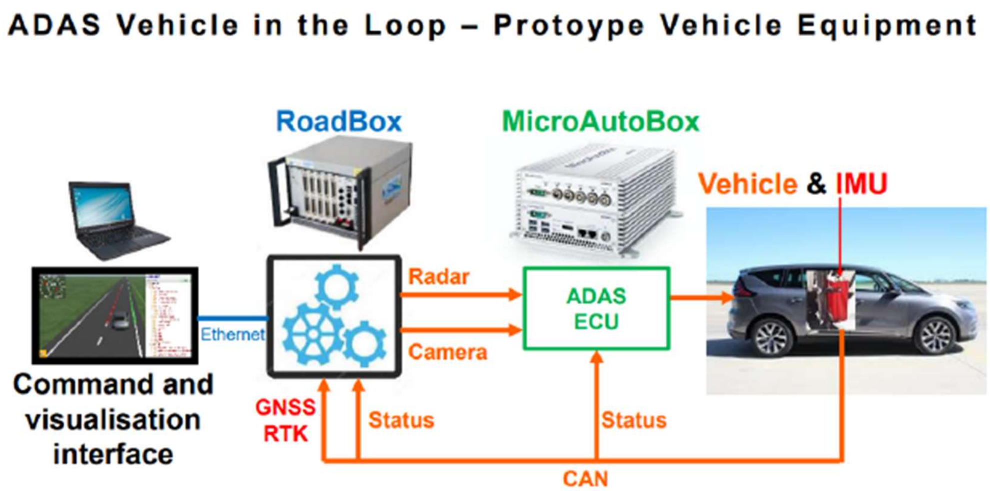


Figure 26:Renault architecture proposal for Vehicle In The Loop

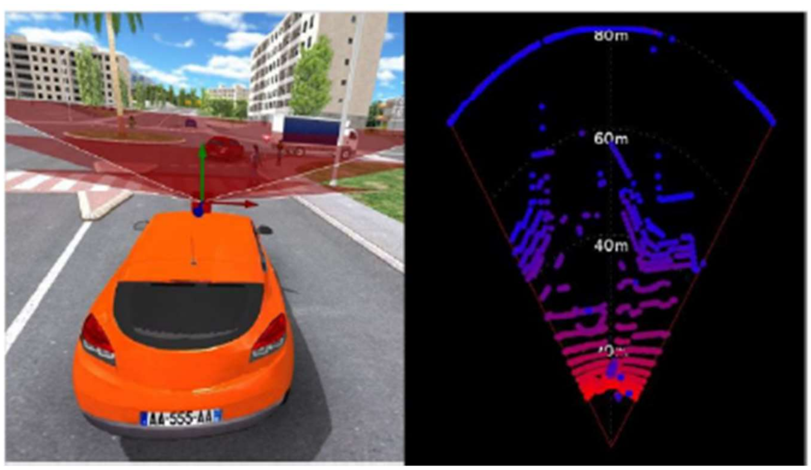


Figure 27:Automotive LIDAR illustration and 2D viewer. (source: ESI group)

Some limitations will be faced during the experimentation and, in particular, the sensors that integrate perception and classification algorithm.

3. Tests and experimentations

Digital continuity and hybrid test system

Digital continuity and hybrid testing are two important concepts in modern engineering that have emerged as a result of the increasing complexity of modern systems. As systems become more complex, it becomes increasingly difficult to maintain a consistent understanding of the system across the various stages of its lifecycle, from design to testing and deployment.

SPHEREA test systems provide Hybrid testing solutions. Itis a testing approach that combines physical testing with digital simulations. This approach allows engineers to test systems in a more realistic and comprehensive way, by combining the benefits of physical testing with the speed and flexibility of digital simulations. Hybrid testing can be used to validate designs, test the performance of systems, and identify potential issues before they arise in the physical system. From PRISSMA WP2 SPHEREA has already proposed the generic modular distributed test system architecture that enables to address the test needs of simulation, test track and open road with the same instrumentation architecture.

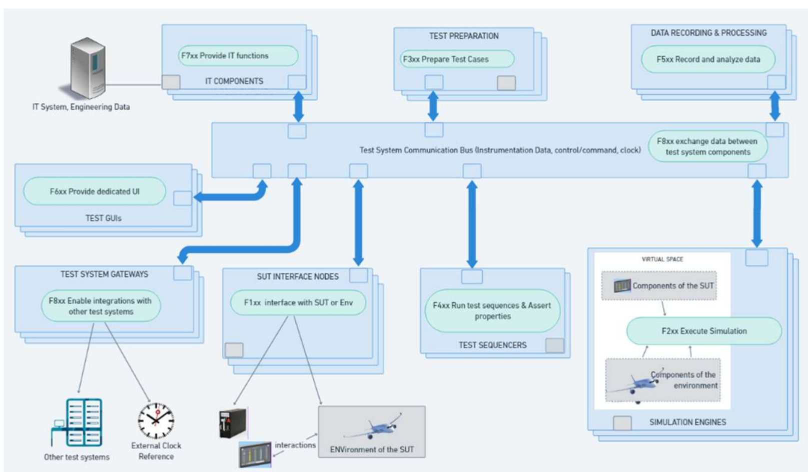


Figure 28: SPHEREA generic modular distributed test system architecture

This architecture has already enabled to integrate road simulation engine.

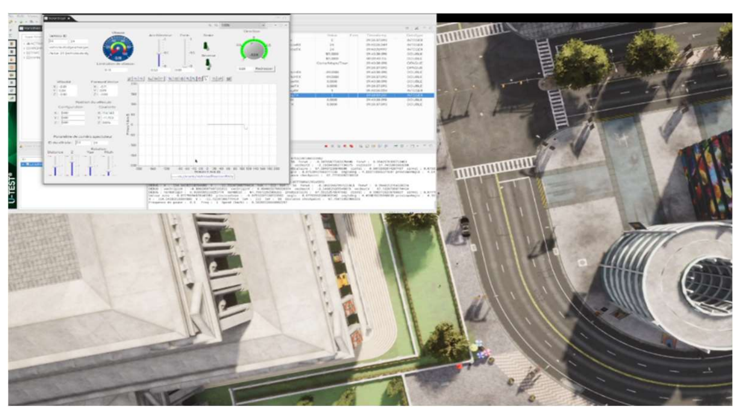


Figure 29: U-TEST® integration with Carla

Axis 1: qualified recorder

Objective: study the impacts of the recording of data in the scope of AI automated vehicle, with a focus on the reliability of recorded data and robustness versus intentional or accidental modification. The RGPD impact should also be addressed.
This axis of experimentation is linked with WP7 (reliability of improvement of AI functional chain after accident or near-accident), WP4, WP5 (robustness versus modifications), and WP3 (comparison between real and simulated test runs).

Axe 2: Digital twin and augmented reality

To enable the happening of rare situations combining multiple perturbations, SPHEREA will study the usage of a ground support equipment to inject effects coming from digital twin overlay of sensors real acquisition. The main challenge address will be the deployment of the simulation on the appropriate architecture:

- On the one hand, the simulation of physical realistic sensors effects require huge calculation power (see the PRISSMA L2.2 on the state of the art). Therefore, one par of the simulation should be computed in cloud computing to stack clusters off computation power
- One the other hand, the latency to communicate with this cloud computation will forbid to simulate some of the most reactive behaviors

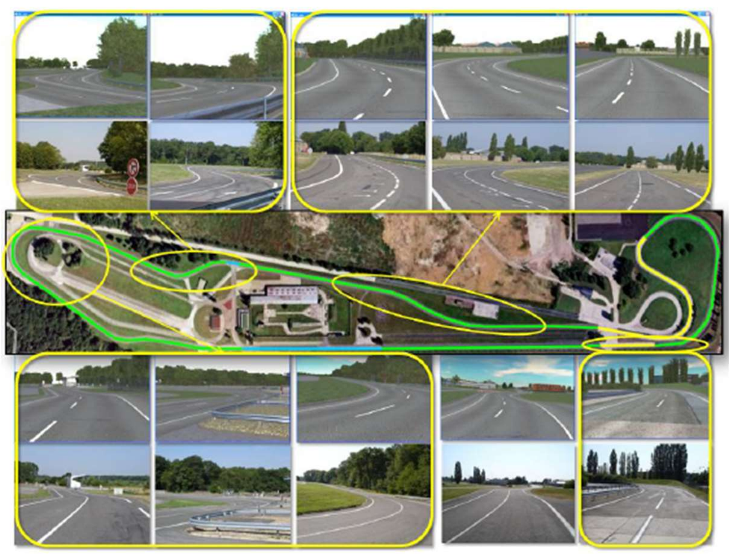


Figure 30: Digital twin of the Satory test track in Versailles. Implementation in Pro-SiVIC™ (source: Univ Eiffel)