International Telecommunication Union

ITU-T Technical Report

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(29 September 2022)

Focus Group on AI for autonomous and assisted driving (FG-AI4AD)

FGAI4AD-03

Automated driving safety data protocol – Practical demonstrators

ITU-T

Technical Report ITU-T FGAI4AD-03

Automated driving safety data protocol – Practical demonstrators

Summary

This Technical Report describes the practical application of the behavioural monitoring and evaluation of autonomous or assisted driving software. Following Workstream 2 (Technical Specification and Demonstration), FG-AI4AD operates as an open collaborative pre-standardization effort to create one or more ITU-T Recommendations in the form of telecommunications and computer protocol specification documents. The specific focus of these ITU-T Recommendations will be in the area of "*the behavioural evaluation of artificial intelligence (AI) responsible for the dynamic driving task*", including the "*in-use assessment of AI driving behaviour using onboard vehicle systems*" expected to become "*an integral part of the field monitoring of assisted and automated vehicles required to ensure continual validation of safety performance*".

The Safety State Framework's methodology for monitoring and assessing traffic behaviour is promising. It objectivates the individual safety performance of all traffic participants. The method is comparable with the assessment of novice human drivers and, therefore, relatively easy to implement. Making it a standard, the commercial use of autonomous vehicles (AVs) will speed up.

Keywords

Artificial intelligence, assisted and automated vehicles, autonomous driving, behavioural evaluation, continuous validation, dynamic driving task, field monitoring, in-use assessment, safety protocol, safety metrics, standardization.

Note

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Change log

This document contains Version 1 of the ITU-T Technical Report on "*Automated driving safety data protocol – Practical demonstrators*" approved at the ITU-T Focus Group meeting held online, 28-29/09/2022.

Editor: Jorrit Kuipers
robotTUNER – Green Dino
The Netherlands
Tel: +31 6 53946943
Email: jorrit@robottuner.com/
j.kuipers@greendino.nl

Acknowledgement

This Technical Report was prepared under the leadership of Mr. Bryn Balcombe, Chair of ITU-T FG-AI4AD (ADA Innovation Lab Limited, United Kingdom).

It is based on the contributions of numerous authors who participated in the Focus Group activities. Due credit is given to the following Focus Group participants representing their respective organizations: Jorrit Kuipers (robotTUNER), Niels de Boer, Teng Sung Shiuan, Ana Isabel Garcia Guerra (CETRAN), James Long (TRL).

Jorrit Kuipers (robotTUNER) served as the main editor of this Technical Report. Stefano Polidori (Cousellor, ITU), Mythili Menon (Project Officer, ITU) and Carolina Lima (Administrative Assistant, ITU) served as FG-AI4AD Secretariat.

The research in clause 6.3 is supported by the National Research Foundation, Singapore, and Land Transport Authority under Urban Mobility Grand Challenge (UMGC-L010). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore and Land Transport Authority.

Copyrights

The authors acknowledge, to the best of their knowledge, that the materials published in this Technical Report have no copyright restrictions that would prevent the free adaptation and distribution of this material in any resulting ITU-T publications.

Table of Contents

List of Tables

List of Figures

Page

Technical Report ITU-T FGAI4AD-03

Automated driving safety data protocol – Practical demonstrators

Introduction

This Technical Report describes the practical application of the behavioural monitoring and evaluation of autonomous or assisted driving software. Following Workstream 2 (Technical Specification and Demonstration), FG-AI4AD operates as an open collaborative pre-standardization effort to create one or more ITU-T Recommendations in the form of telecommunications and computer protocol specification documents. The specific focus of these ITU-T Recommendations will be in the area of "*the behavioural evaluation of artificial intelligence (AI) responsible for the dynamic driving task*", including the "*in-use assessment of AI driving behaviour using onboard vehicle systems*" expected to become "*an integral part of the field monitoring of assisted and automated vehicles required to ensure continual validation of safety performance*".

Unlike supervised learning tasks, the minimal evaluation environment for autonomous vehicles must include more than a labelled dataset. The evaluation environments should incorporate: 1) accurate and efficient simulators to reduce the time and cost of testing; 2) hardware implementations to correlate simulation results with real-world performance, and 3) adaptable benchmarks which include strong baselines supported by competition to drive innovation and reduce overfitting.

This Technical Report describes the set-up of three practical demonstrators:

1) clause 6.1: COLUMBUSS the Netherlands
2) clause 6.2: CETRAN Singapore
3) clause 6.3: Smart Mobility Lab London

In the discussion (clause 7), a common approach for evaluation environments is briefly specified based on similarities between the demonstrators. The conclusions and next steps (clause 8) address future cooperation between the demonstrators for accelerating the commercial use of autonomous vehicles (Avs).

1 Scope

The expectation is that ITU-T Recommendations generated by FG-AI4AD activities should be:

- Globally applicable and globally comparable.
- Independent from AV software architecture.
- Flexible to implement software and hardware.
- Flexible to hosting onboard the vehicle or off-board the vehicle at the edge or in the cloud.
- Provide a framework for assessment of performance but enable individual entities to set acceptance thresholds for performance.
- Policy agnostic – meaning the implementing entity may include; self-certification, enhanced telematics insurance, an independent safety organization, a government regulator or others.
- Applicable to both real-world and virtual-world environments.
- Sensitive to the cost of implementation.

The practical demonstrations of the proposed behavioural evaluation showcase the benefits of standardized field monitoring of assisted and automated vehicles also known as autonomous vehicles (Avs) and the value of continual validation of safety performance to building public trust in the future of mobility.

2 References

References are listed in the clauses.

Current information about the demonstrators can be found at:

www.columbuss.eu

www.cetran.sg

www.smartmobility.london

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Technical Report

This Technical Report defines the following term:

3.2.1 shadow driving: The situation where a human driver is in control of the ego vehicle and interacts with traffic, unaware of an AI-driver (auto-pilot) or AI-examiner making their own decisions (behaviour) transporting the ego vehicle in the operational design domain.

4 Abbreviations and acronyms

This Technical Report uses the following abbreviations and acronyms:

AEB-P	Autonomous Emergency Braking – Pedestrian
ADS	Automated Driving System
ASIL	Automotive Safety Integrity Level
AV	Autonomous Vehicle
AVSC	Automated Vehicle Safety Consortium
CAV	Connected and Automated Vehicle
CCAM	Connected, Cooperative and Automated Mobility
COLUMBUSS	Connected Level 5 Unmanned Buses
DSS	Driver Safety Score
FAIR	Findable, Accessible, Interoperable, Reusable
GLOSA	Green Light Optimal Speed Advisory
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HARA	Hazard Analysis and Risk Assessment
LiDAR	Light Detection and Ranging
NGO	Non-Governmental Organization
ODD	Operational Design Domain
PLD	EU Product Liability Directive 85/374/EEC of 25 July 1985
PMD	Personal Mobility Device
PMS	Pilot Mitigation Score

RSS	Responsibility Sensitive Safety
RTK	Real-Time Kinematic
SAE	Society of Automotive Engineers
SFF	Safety Force Field
SSF	Safety States Framework
SSS	Safety State Score
TRL	Technology Readiness Level
TSV	Traffic Simulated Vehicle
TTC	Time-to-Collision
UTC	Urban Traffic Control
VA	Virtual Assistant
VDI	Virtual Driving Instructor
VRU	Vulnerable Road User
WP	Work Package

5 Conventions

None.

6 Practical demonstrators

The practical demonstrators were initially initiated to implement the data protocols from TR01. Later, for practical reasons the Focus Group decided to limit the work and describe the set-ups of three existing sites that monitor autonomous vehicles (AVs) and evaluate their compliance with the results of the FG-AI4AD 'Molley Problem Public Consultation' (Table 1).

Table 1 – The 'Molly Problem Public Consultation' responses (11-20th October 2020)

No.	**Would you expect…**	**Yes**	**Unsure**	**No**
1	**the software *to be aware* of the collision**	97%	2%	1%
2	**the software *to stop* at the collision site**	94%	4%	2%
3	**the software *to indicate a hazard* to other road users**	97%	2%	1%
4	**the software *to alert emergency services***	94%	5%	1%
5	**the software to recall the *time* of the collision**	99%	n/a	1%
6	**the software to recall the *location* of the collision**	99%	n/a	1%
7	**the software to recall *when* the collision risk was identified**	93%	6%	1%
8	**the software to recall *if* Molly was detected**	96%	3%	1%
9	**the software to recall *when* Molly was detected**	96%	2%	2%
10	**the software to recall *if* Molly was detected as a *human***	91%	6%	3%
11	**the software to recall *when* Molly was detected as a *human***	90%	7%	3%

Table 1 – The 'Molly Problem Public Consultation' responses (11-20th October 2020)

No.	Would you expect…	Yes	Unsure	No
12	**the software to recall *whether* mitigating action was taken**	98%	1%	1%
13	**the software to recall *when* mitigating action was taken**	97%	2%	1%
14	**the software to recall *what* mitigating action was taken**	96%	3%	1%
15	**similar recall abilities for *near-miss events***	88%	5%	7%
16	**expect *driving* to be *prohibited* for software *without recall capability***	72%	15%	12%

6.1 COLUMBUSS – Connected Level 5 Unmanned Buses

6.1.1 Set-up

COLUMBUSS proposes demonstrating that autonomous public transport can reduce transportation costs of people and goods, increase mobility within cities and the countryside, improve the quality of life, and be acceptable to all stakeholders. We intend to demonstrate for all routes in the Metropole Region Den Haag Rotterdam and the Province of Groningen that public transport capacity and use will increase by implementing automated mobility.

COLUMBUSS is an initiative of the Metropole Region Den Haag Rotterdam, the Province of Groningen and robotTUNER. Dutch project partners are RET, Qbuzz, EBS, Delft Technical University, Veilig Verkeer Nederland and Green Dino. COLUMBUSS is open to international collaboration and seeks mirror sites. The Australian Road Research Board, West Midlands (UK) and CETRAN, are associated partners of COLUMBUSS.

The budget for the COLUMBUSS project in the Netherlands is approximately 10 million euros. The available resources are; more than 40 FTE, 15 OV-busses for shadow driving, 5 automated (Level 5) OV-busses, 5 automated micro-busses/MPVs and 1 minibus/taxibus and 2 charging robots. Columbuss Netherlands started in 2022. The lead time is 4-6 years until commercial licensing of autonomous public transport is a fact.

Although over 100 pilots on autonomous transports were effectuated in Europe, ***the impact of autonomous mobility is insufficiently validated***. The main reason is the lack of demonstrators at SAE Level 5 (only passengers in the vehicle) under natural traffic conditions. 'Natural' means an everyday 24/7 all-weather exposure to other traffic at the maximum road speeds for an extended period. Until now, impact studies have had low predictive value. Autonomous vehicles (AVs) are operated at low speed (average far below 20 km/h) and mostly on short tracks (50% of the pilot tracks were shorter than 1500 m). Safety driving skills at higher speeds are limited, and human operators must constantly observe the ODD[1] and frequently interfere to avoid collisions. Tracks are isolated from the rest of the traffic, especially vulnerable road users like cyclists and pedestrians.

AV demonstrations and commercial deployments are hindered by a lack of ***safety assessment standards***. Safety assessment metrics are not yet formulated, tested and accepted. Safety assessment

---

[1] ODD, or Operational Design Domain. The term defines all conceivable overlapping conditions, use cases, restrictions and scenarios that an AV might encounter, even the most esoteric edge cases.

metrics are still under construction by organizations such as the Automated Vehicle Safety Consortium (AVSC) and the UN ITU-T Focus Group AI4AD. The AV industry focuses on the safety envelope concept of absolute safety around the AV. These are absolute metrics that lack predictive strength to determine risk mitigation behaviour.

COLUMBUSS uses ***the competent human driver*** as the reference model for safe driving, a paradigm shift in assessing and licensing AVs. AVs must perform at least equal to their human peers and not twice as safe as the AV industry claims. This approach to driving safety is much easier to implement and execute. It is also easier to accept for society since its base is the existing legal metrics of the competence testing of human drivers. We implement these human-centred metrics in our AV systems and demonstrate and evaluate the improved safe driving skills under natural traffic conditions. Extensive exposure under natural traffic conditions will reveal reliable and realistic opportunities and threads for connected, cooperative and automated mobility (CCAM) solutions, strengthening the value of the impact studies we will carry out.

6.1.1.1 Objectives and ambition

6.1.1.1.1 Objectives

The COLUMBUSS project aims to demonstrate the impact of CCAM in public transport under natural traffic conditions SAE Level 5. The COLUMBUSS partners commit themselves to realizing this overall aim by following the COLUMBUSS-specific objectives (see Table 2).

Table 2 – Objectives of COLUMBUSS

Objective of COLUMBUSS	Measure of success
To conduct pilot experiments on different public transport routes in 2 regions in the Netherlands to determine the impact of autonomous transport in various use-cases, physical environments and human culture.	Autonomous vehicles are implemented on public transport routes and run at SAE level 5.
Create a standardized evaluation platform for objectively determining the driving capabilities of automated vehicles, test the COLUMBUSS system's robustness under natural traffic conditions, and compare it to human drivers and other AVs.	Uptime is minimally 95%.
	Reliability is more significant than 98%.
	Safe driving skills are equal to human bus drivers or better.
	Safe driving skills are better than other Avs.
Provide a public transport solution to maintain or increase the liveability and accessibility of rural and urban areas.	As evaluated by the standard assessment tools, the new transport solution contributes positively to the liveability of the piloted areas.
Assess the economic, environmental and social impact using an extensive evaluation framework built upon assessment tools developed in earlier projects.	An evaluation framework of standard protocols is completed and provides a scientifically underpinned, reliable tool for assessing the various impact elements of autonomous transport.
Involve all relevant stakeholders (users, user interest groups, transport operators, procurers of public transport, certifying bodies, system and knowledge providers) in designing, evaluating, evaluating, optimizing and presenting the pilots.	All stakeholders are involved in every pilot for maximum success. A predefined protocol supports the selection process and evaluation of stakeholders.

Table 2 – Objectives of COLUMBUSS

Objective of COLUMBUSS	Measure of success
Disseminate and communicate the results of the outcome of the pilot to all stakeholders, including the general public. The access and use of new technology should be easy.	All scientific partners will present results in journals and symposia. All sites and technology suppliers will publish information on the COLUMBUSS website and arrange presentations and demonstrations. All partners contribute knowledge to the open-source COLUMBUSS platform and promote its use.

6.1.1.1.2 Beyond the state-of-the-art

The COLUMBUSS consortium goes beyond state-of-the-art technology and knowledge. An overview of the state-of-the-art and our ambitions is provided in Table 3.

Table 3 – State-of-the-art and ambition of COLUMBUSS

Aspect	State of the art	COLUMBUSS ambition beyond the state of the art
Using AVs leads to better connections between villages, hubs and cities crossing rural areas.	No AVs connect villages, hubs and cities, crossing rural areas in natural traffic conditions.	All demonstrators will demonstrate public bus transport between villages, hubs and cities connecting remote suburban areas under natural traffic conditions 24/7.
Extensive demonstration at SAE Level 5 (only passengers in the vehicle, no safety driver)	Demonstrators only exist in fully controlled environments without other traffic; e.g., Parkshuttle RIVIUM.	All pilots will deploy autonomous driving for a minimum of 3 years in natural traffic. In the third year, we aim at driving fully autonomous SAE level 5, only with passengers on board (if allowed) and remote-supervised.
Lower price per kilometre per person (PPP)	Driver costs amount to 60% of the total costs of transport.	With a free-of-cost open-source AV-software platform and deployment of in-use public buses for mass transport, it is possible to achieve a 50% cost reduction.
		Extra revenues by transporting goods and new services like maintenance in off-peak can reduce the PPP.
Reliability	For predicting an accurate accident risk of the autopilot (accidents per kilometre), many kilometres need to be driven without the interference of a human operator.	The decisions of the auto-pilot will be evaluated over more than 2.7 million kilometres.
		We assessed risk-mitigating actions and evaluated the activities of the human driver/supervisor as well, making corrections for interference by the human driver supervisor possible.

Table 3 – State-of-the-art and ambition of COLUMBUSS

Validity	Demonstrations have always been performed under restricted conditions: low speed, a supervisor onboard, limited priority, etc.	We evaluate the performance and impact of AVs under natural (realistic) traffic conditions; in various environments, at different speeds, integrated into the current public transport network with vulnerable road users and without a supervisor on board (in the third year).
	The sample population of passengers do not include all groups, like disabled people, women and children.	Co-creation and surveys include a balanced mix of passengers and road users with extra attention to vulnerable groups.
	Limited passengers are transported in AVs and evaluated after a ride.	A community of more than 1 000 users, at least 200 per region, will be formed for exposure, co-creation and acceptance surveys.
Vehicle types	Almost all demonstrations are carried out with medium-sized vehicles with a 4-8 persons capacity.	The focus is on different types of buses, from micro-sized vehicles with 1-2 persons to macro-sized vehicles >12 persons and multipurpose vehicles (Figures 1 & 2).
Routes and speed	Demonstrations on low and medium-speed routes in residential and city centres, driving below 20 km/h.	Different speeds, up to 80 km/h, are part of the demonstration in all demonstrator and mirror sites.
Law and regulation	Demonstrations are performed under exceptional or experimental temporary laws.	ARRB Australia uses the COLUMBUSS project to model the legislation law for Australia's autonomous vehicles settling in 2026. The legislation will be used in discussion with European legislators.
	Permits are provided separately for vehicles and drivers, issued by national authorities. Both legal entities do not have a regulatory framework for licensing driverless vehicles. Manufacturers, regional and local authorities and customers need transparency and support for the deployment of AVs, one common regulatory framework for vehicle testing and driver testing.	A shared open-source AV software and knowledge platform, including all stakeholders, will stimulate co-creation. Efficiency increase with a standard open platform for development, testing, licensing and control of and speedup commercial licensing.
	Regulation 157 permits autonomous driving SAE level 3 below 60 km/h in specific ODDs (highway).	COLUMBUSS project will be a model for legislation for autonomous vehicles in Europe SAE level 5 below 80 km/h in all ODDs.
Safety metrics	The AV industry organized in the AVSC focuses on absolute safety around the AV (the safety envelope concept) and scenario testing based on driving	In the COLUMBUSS approach, the competent human driver is the reference model, intending to drive at least as safe as a human peer driver.

Table 3 – State-of-the-art and ambition of COLUMBUSS

	simulation, with the ultimate claim of being twice as safe as human drivers.	
	Currently, absolute metrics are applied. They lack predictive strength to determine risk mitigation behaviour. Risks are high in high-risk situations and low in low-risk situations; only the risk mitigation actions reveal safe driving skills.	COLUMBUSS embraces legal safety metrics for human drivers. The predictive strength for accident risk is valid and robust. Risk mitigating and peers related performance increase the predictive power of the assessment scores of safe driving skills.
Testing	Testing of AVs under natural traffic conditions is done with driving simulation.	robotTUNER developed the concept of **shadow driving** to test on-road under natural traffic conditions.
AV-software	Demonstrations are done with AVs using proprietary software.	All COLUMBUSS demonstrations will use a free-of-cost shared open-source software platform.
Patent safety metrics	COLUMBUSS focuses on applying human safety assessment metrics integrated with digital twin technology. Patent research by national patent authorities resulted in automated tutoring systems used for human driver education and testing not using synchronized digital twin simulation technology.	COLUMBUSS partner robotTUNER submitted a European Patent Application P100268EPPC application date 03-01-2018, renewal date 03-01-2021. The patent is granted for the Netherlands P100268NL00 and is pending for the EU, USA, Canada and India. The patent describes the vital components of the COLUMBUSS AV software; AutoPilot, AutoExaminer and AutoMonitor. The invention 'Autonomous assessment system' describes an assessment system for automated assessment of the behaviour of autonomous actors like self-driving vehicles. Essential for the invention is synchronized digital twin simulation technology for the assessment process. This patent goes beyond the state-of-the-art of patented assessment inventions.
Design methodology	Machine-centred, starting from technology.	Human-centred, starting from human drivers.
Evaluation methodology	No standard evaluation methodology is used.	A standard evaluation methodology is followed with specifications and protocols used by all sites in all pilots.
		The standard methodology will fit the common evaluation framework for large-scale demonstration pilots in Europe. Knowledge gained will be exchangeable with the EU database of relevant scenarios.

Table 3 – State-of-the-art and ambition of COLUMBUSS

Environmental effects	European cities introduce zones related to reducing pollution and improving air quality to control noise and harmful emissions of fossil fuel vehicles. Vehicles are kept out of city centres.	Electrification of in-use vehicles with replacement or with a retrofit lowers harmful emissions.
		Artificial intelligent autopilots have an optimal driving style lowering noise disturbance. Artificial intelligent planners choose optimal routes and schedules, lowering travel times.
	Greener cities with fewer vehicles; more places for living, and city nature	Higher capacity and improved and new public transport services will lower private ownership and the number of commercial vehicles.

Figure 1 – The COLUMBUSS macro bus

Figure 2 – The COLUMBUSS multipurpose microbus with exchangeable modules for transporting people and goods

6.1.1.1.3 R&I maturity

The COLUMBUSS technology platform comprises various components with different technology readiness levels (TRLs). Start and end TRLs are listed in Table 4.

The COLUMBUSS technology supplier is Green Dino BV, Wageningen, the Netherlands.

Table 4 – Summary of the current TRLs of COLUMBUSS' components and the TRLs aimed for at the end of the project

Technology	Start TRL	Justification
Virtual assistant (VA) Virtual driving instructor VDI)	9	VA and VDI have been commercially available on driving simulators since 2003.
Driving simulation	9	Driving simulators have been commercially available since 2003.
COLUMBUSS AutoPilot	6	The AutoPilot was demonstrated in 2021 in several restricted test environments running at different speeds up to 80 km/h. It controlled a Renault Twizy 1-2 cm accurately using data from GNNS with RTK correction and radar with object detection on-chip.
COLUMBUSS AutoExaminer	6	The AutoExaminer software was demonstrated in 2021 in an integrated test environment with the AutoPilot using a synchronized digital twin.
COLUMBUSs AutoControl	6	The AutoControl was demonstrated in 2021 in several restricted test environments. It controlled an electric vehicle with acceptable latency at 6 KM/h in a standard 4G network.
Dino multipurpose vehicle (MPV)	6	In 2021 Green Dino constructed a multipurpose vehicle on the chassis of a Renault Twizy with automatically exchangeable modules for performing various tasks such as transporting people and goods and performing services such as watering public gardens. The MPV is demonstrated using AutoPilot and AutoControl software in several restricted test environments.
Shadow driving module	6	In 2021 the AutoExaminer software was integrated by Green Dino into a stand-alone module without connecting to the host vehicle. The SD module is demonstrated in several restricted test environments with the Renault Twizy and Dino MPV.
AutoPilot module	6	In 2021 the AutoPilot software was integrated by Green Dino into an add-on module with redundant technology for vehicle control. The AutoPilot module is demonstrated in several restricted test environments with the Renault Twizy and Dino MPV.
COLUMBUSS critical scenarios	5	In 2021 a limited set of critical scenarios was created by Green Dino to evaluate the AutoExaminer and Autopilot software in a driving simulation.
Open-source maps	9	In 2021 Green Dino demonstrated an integration of the Auto-pilot module and Openstreetmaps.

Table 4 – Summary of the current TRLs of COLUMBUSS' components and the TRLs aimed for at the end of the project

Virtual assistant	9	–
Driving simulation	9	–
COLUMBUSS AutoPilot	7	The shared open-source COLUMBUSS AV technology and the proprietary AV technology of Green Dino have been demonstrated under natural traffic conditions and validated. Green Dino will improve the open-source and proprietary AV technology based on the outcomes reported at the end of the project. After improvements, the shared open-source software is ready for use by all stakeholders. The improved AV technology will be released in 2025 (TRL 8).
COLUMBUSS AutoExaminer	7	
COLUMBUSS AutoControl	7	
COLUMBUSS Open-source platform	7	
Dino MPV	7	
Shadow driving module	7	
AutoPilot module	7	
COLUMBUSS critical scenarios	7	The critical scenarios are evaluated with the data from the pilots. Based on the outcomes reported at the end of the project, Green Dino will improve the critical scenario descriptions. After improvements, the descriptions will be added to the shared open-source software. The improved critical scenarios will be released in 2025 (TRL 8).
COLUMBUSS APIs for open-source maps	7	The APIs are integrated into the open-source. The improved APIs will be released in 2025 (TRL 8).

6.1.1.2 Methodology

6.1.1.2.1 Overall methodology

The COLUMBUSS project provides two demonstration regions with connected, cooperative and automated mobility (CCAM) performed by 4-7 public transport companies. The pilots demonstrate the potential of the COLUMBUSS technology under natural traffic conditions driving 24/7 at maximum road speed (if applicable), and finally at SAE Level 5. We believe there is no solid business case possible with the current limitations of AVs, like very low speeds, an operator or steward on board, or limited interaction with other road users. Various automated buses will drive as many kilometres as possible to strengthen the predictive value of the impact studies. The local authorities and public transport operators, responsible for defining public transport needs, requirements and operations, will act as launch customers.

The project methodology includes three main elements (Figure 3):

1) two demonstration regions in the Netherlands;
2) a shared technology base implementing a gradual increase in autonomy level; and
3) a standardized effect and impact evaluation methodology involving all relevant stakeholders.

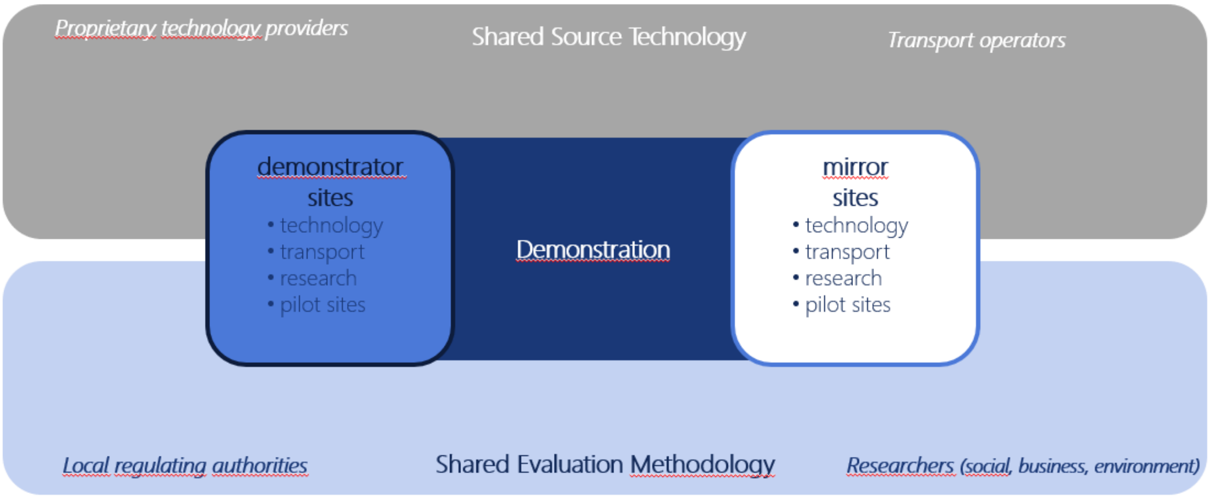


Figure 3 – Main components of the project methodology: shared source technology, demonstration in real-traffic demonstrator and mirror sites, and shared evaluation methodology

The public transport chain is covered, ranging from the villages and residential areas to the city centre (Figure 4). The AVs are obliged to drive only under natural traffic conditions and 24/7 if needed. The final demonstrations (TRL 7) will occur in the third year of the deployment without a human supervisor onboard the vehicle. The AVs will only transport passengers (if allowed) and goods.

Figure 4 – Types of pilot sites, varying from rural to urban areas, and their characteristics

The autonomy of the AV rises step by step to guarantee safety. The demonstration includes autonomous driving at 5-30 km/h in city centres, 30-50 km/h in residential and 80 km/h in rural areas.

All pilots start the first year with *'shadow driving'* for absolute safety. A human bus driver will drive a regular bus while the COLUMBUSS AutoPilot runs stand-alone in the background, not connected to the bus. The COLUMBUSS AutoExaminer constantly assesses the decisions of the AutoPilot and the human bus driver. The COLUMBUSS researchers use the data to construct a bus driver reference model, the 'human peer driver model'. The reference model supports the evaluation of the performance data of the AutoPilot. The suppliers of the COLUMBUSS technology constantly improve AutoPilot until it reaches the safe driving competence level of the human peer driver. The AutoPilot receives control over the vehicle when its safe driving performance is equal to or better than the human peer. First under the supervision of a human driver driving (SAE Level 3) and eventually fully automated (at SAE Level 5) with remote control. When the COLUMBUSS AutoControl detects unacceptable deviations from the human reference performance, automated driving stops and supervised driving, or even shadow driving, starts again. We have named this adaptive automation method 'performance-based automation level selection'.

This new testing method, developed by COLUMBUSS partner robotTUNER and associated partner CETRAN, is comparable to accompanied driving for novice human drivers. Accompanied driving is still the safest way of becoming a safe driver worldwide. The expert human bus driver will supervise the COLUMBUSS AutoPilot until it drives enough kilometres safely to ensure the accent risk.

Determining autonomous public transport's economic, social, environmental, and regulatory impact is equally important as the technology assessment and improvement. Therefore we involve all stakeholders; local, regional, and national governments, public transport companies, AV industry, companies, non-governmental organizations (NGOs), citizens/interest groups, etc.

A standard framework for evaluating all relevant non-technical aspects of autonomous transport will be established, primarily by the knowledge partners supported by transport operators and NGOs. Uniform protocols assess safe driving, traffic conditions, scenarios, business cases, social impact, opinions, etc. These protocols make data collection easier and uniform at the demonstrator and mirror sites. Standard protocols support the dissemination activities as well. Standardization of the deployment process will lower the efforts, cost and duration. Therefore, standardization is crucial for speeding up the roll-out/scale-up of autonomous transportation of people and goods in Europe.

We work towards a comprehensive and holistic view of the relevant legal framework necessary to speed up AVs' commercial licensing; the vehicle regulatory framework, the (human) driver regulatory framework, the consumer protection framework, the machine regulatory, liability, insurance, environmental regulations, data protection, artificial intelligence regulatory aspects, as well as non-binding instruments such as industry standards. These legal instruments interact, and we will investigate them integrally to address legal shortcomings or problems for effectively licensing automated vehicles. In addition, such a holistic approach can result in a standard method for assessing vehicle safety and security. COLUMBUSS partners emphasize using existing legal safety metrics to assess human drivers. This European legal framework supports users' and other stakeholders' adoption and acceptance of new vehicle technologies and speeds up autonomous buses' commercial use because it is well-known. Therefore, we believe it is easier to implement and more reliable than other assessment methodologies. Table 5 gives an overview of the COLUMBUSS project methodology.

Table 5 – Overview of the COLUMBUSS project methodology

Aspect	How this is approached
Implement a series of pilots with sound mobility use cases for autonomous transport under natural traffic conditions.	***Two demonstrator sites*** in the Netherlands. Each demonstrator hosts several pilot tracks covering the transport chain of people and goods from the city centre through residential areas towards villages in rural areas driving at different speeds under natural traffic conditions.
	Mirror sites use the same technology and methodology and support validation.
Test robustness, reliability and safety of highly automated AV systems and services while focussing on user interaction and interaction with other road users; vital enabling technologies (e.g., sensors, connectivity, cybersecurity, AI, big data, space-based services), physical/digital infrastructure support, optimized traffic and fleet management, etc.	Testing will be structured using ***standard protocols*** formulated by scientific and industry partners. Every pilot site uses these protocols for the set-up, data collection, data evaluation, and reporting of the result.

Table 5 – Overview of the COLUMBUSS project methodology

Aspect	How this is approached
User and customer needs for mobility and logistics, paying particular attention to differences in mobility patterns by gender, age, disability and other social groups.	Deployment of AVs can result in more and better public transport. Increased public transport capacity enhances mobility for all consumers, especially vulnerable social groups.
	In collaboration with NGOs, robotTUNER will lead in organizing co-creation sessions to involve citizens, especially the different gender groups, children, the elderly, and people with disabilities.
	Group differences have special attention in the co-creation and evaluation.
More high-quality service is well integrated with physical and digital mobility services like rail transport and MAAS.	Differentiation between macro-, medium- and micro-buses driving at different routes and speeds makes it possible to cover the transport chain of people and goods from door to hub in rural, residential, and city centre areas.
	Attention is paid to the seamless exchange of people and goods between buses and modes: road, rail, air, water, and tube.
Contribution to effective assessment and demonstration of benefits on energy efficiency, traffic flow, safety, user appreciation, etc. based on holistic modelling solutions.	Standard technology and methodology for assessing the impact of CCAM, uniformly applied in every pilot, increases the reliability and validity of assessment and demonstration. A holistic approach is supported by including scientists and experts in technology and social sciences. Together they will make specifications and protocols for executing the pilots, evaluation and dissemination.
Leveraging already existing investments at national and European levels on demonstration activities, optimizing return on investments and creating a solid basis for even larger scale demonstrations and system integration.	COLUMBUSS builds on previous projects executed by the consortium partners and other national and European initiatives such as: Projects with partner involvement: • @north • AVLM • INTERREG ART Forum • CAM Testbed UK
	Uniform specifications and protocols for all activities and equipment will standardize the deployment process and lower the complexity, efforts and cost. An open standard for deployment will support the scale-up of autonomous transportation of people and goods.
Co-creation with users to demonstrate benefits and raise public acceptance/adoption of CCAM.	An interface, the virtual assistant, will be co-created with children, elderly and disabled users, and vulnerable road users who assist them using the AV. GR and rT will, together will local NGOs such as VVN, take the lead in organizing and performing the co-creation activities.

Table 5 – Overview of the COLUMBUSS project methodology

Aspect	How this is approached
The practical contribution of social sciences and humanities (SSH) disciplines and the involvement of SSH experts and institutions, as well as the inclusion of relevant SSH expertise, to produce meaningful and significant effects enhancing the societal impact of the research activities.	SSH experts will develop protocols for co-creation, collecting, and evaluating the data. They will support a holistic view using their legal, social, learning, spatial, human-factors, economic, ethics and AI expertise.
	Business modelling experts of public transport organizations PTOs are involved, like West Midlands, Qbuzz, and RET.
All vehicles used for testing the innovative CCAM concepts **use zero emission technologies**.	We only use electric vehicles for autonomous driving. Fuel-powered buses will be electrified or replaced by electric buses before SAE level 3 is adopted.
	Shadow driving can start with operational carbon fuel-powered buses for a quick start in the first year of the deployment. During the project, these buses are electrified for automated driving.

The operational phase of the pilots is three years, 12 months at minimum per autonomy level, to gather enough data to validate the outcomes. All pilot sites will operate according to the standard working method and experimental set-up following the predefined protocols. Using the same protocols benefits the combination and comparison of the resulting output/data of the pilots. The COLUMBUSS shared open-source platform will share knowledge and combine across demonstrators. This platform will be available and supported after the end of the project for everyone interested.

For safety and feasibility reasons, and due to (justified) regulations, pilots start at SAE level 1 autonomy. Step by step, the level of autonomy rises (Figure 5):

- Level 1: shadow driving. A human driver steers the vehicle, and the COLUMBUSS AutoPilot 'drives' in parallel. The COLUMBUSS AutoExaminator evaluates the decisions of the human driver and the AutoPilot. Based on the results, the local, regional and national legal authorities will decide whether the AutoPilot is ready to take control.
- Level 3: supervised driving. The AutoPilot steers the vehicle. The human bus driver can intervene anytime and remain responsible for avoiding accidents. (SAE level 4 driving is impossible if a supervisor's intervention is necessary for random situations. So most AVs drive at SAE level 3).
- Level 5: autonomous driving (remote controlled). The AutoPilot controls the vehicle. No human driver or steward is present in the vehicle, only passengers. (If national regulations require remotely controlling the vehicle, applied at this level).

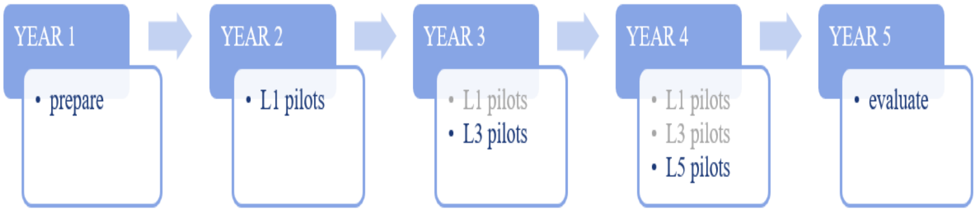


Figure 5 – The level of autonomy of pilots is increased step-by-step, and only after positive evaluation of safety

The automation process is adaptive, meaning lowering the automation level is possible. With the permitting authorities, go/no-go decisions to level up (or down) are made based on a thorough evaluation of the pilot results, auto-assessment of the AutoPilot and human driver interventions. As each pilot follows its own pace, the autonomy level per pilot may differ at certain times. Some pilots might not reach level 5 during the project duration, for example, due to insufficient safe driving skills of the AVs, missing local support base or strict regulations. This also is of great importance for the impact study.

6.1.1.2.2 Related national and international research activities

COLUMBUSS builds on several EU and non-EU-funded projects (Table 6). Some partners participated in these projects and will disseminate the knowledge into COLUMBUSS. We use the knowledge from these projects to support the writing of specifications and formulation of protocols for the pilot set-up and execution and the evaluation, reporting and dissemination.

Table 6 – Overview of EU and non-EU funded projects and how they feed into COLUMBUSS

Project name	Description	How it feeds into COLUMBUSS
@north	The Provinces of Groningen, Friesland & Drenthe study extensive in-practice autonomous transport. E.g., Ommelander Hospital Scheemda. About 10 000 passengers were transported. robotTUNER and Green Dino participate in several @north projects. @Noths' estimated budget is 20 million euros (including the NPG project).	Input WP3-WP6 • User experiences enquiries. • Safety system assessment. • Safety case standardization. • Latency studies in a local 5G network. • AV Interaction with emergency vehicles and control centre. • Communication with the national vehicle authority about safety issues.
ART-Forum ULaaDS	GR participates in ART-Forum – an Interreg North Sea Region Project. The consortium works on an Impact assessment of Automated Road Transport (ART) for people and goods. It also focuses on the capacity building of local and regional authorities in the North Sea Region and aims to facilitate exchange between technology developers and policy-makers. GR participates in the Horizon 2020 project called ULaaDS (Urban Logistics as an on-Demand Service). Three municipalities – Bremen, Mechelen and Groningen – have joined forces with logistics stakeholders and leading academic institutions to accelerate the deployment of innovative, feasible and shared zero-emissions solutions in urban logistics. The role of an autonomous vehicle in the city's logistics is studied within the project.	Input WP1 • Information exchange of safety risks between technology developers and policy-makers in different countries. • Impact assessment on the transport of people and goods. • The role of autonomous vehicles in city logistics. • Cooperation with logistic stakeholders.
CAM Testbed UK	WM and other West Midland partners run CAM testbeds as part of CAM Testbed UK. CAM Testbed UK is a unique facility capable of safely taking ideas from concept to development, virtually and physically.	Input WP1 • Cross-sharing of data between stakeholders. • Concept development.

Table 6 – Overview of EU and non-EU funded projects and how they feed into COLUMBUSS

Project name	Description	How it feeds into COLUMBUSS
	The UK's comprehensive and integrated facilities are world-leading, with the cross-sharing of data and a collaborative way of working.	• Simulation testing. • Collaborative working. • 5G testing. • Different types of AVs.
SPROUT	The SPROUT project studies the driving aspects of urban mobility transition on cities' policies. A ***pilot evaluation framework*** is built to explore changes in urban mobility (i.e., emerging business models, new technologies, and disruptive innovations in mobility). This framework comprises tools for the assessment of • the financial and economic KPIs, • the Environmental & social KPIs, • Policy implementation and user acceptance and • Product's quality and quality-in-use metrics. SPROUT implemented a pilot on autonomous electric pods for cargo hitching to produce an integrated and coordinated regulatory framework, including various policy options in a unified and harmonized framework.	Input WP6 • Driving aspects of urban mobility transition on cities' policies. • Harmonized evaluation framework on cities' policies for urban mobility transition. • Results of a pilot on the deployment of an autonomous electric pod for cargo hitching. • Integrated and coordinated regulatory framework. • Various policy options.
LEAD	The LEAD project creates digital twins of urban logistics networks in six cities to support experimentation and decision-making with on-demand logistics operations in a public-private urban setting. Innovative solutions for city logistics are represented by value case scenarios that address the requirements of the on-demand economy while aligning competing interests and creating value for all stakeholders. Each value case combines several measures (LEAD Strategies): a) innovative business models, b) agile urban freight storage and last-mile distribution schemes, c) low emission, automated, electric or hybrid delivery vehicles, and d) smart logistics solutions.	Input WP2-WP6 • Digital Twins of urban logistics networks. • Requirements of the on-demand economy. • Support experimentation and decision-making with on-demand logistics operations in a public-private urban setting. • Value case scenarios. • Aligning competing interests and creating value for all different stakeholders.
ECOLOGISTICS	ECOLOGISTICS: Low carbon freight for sustainable cities. "The EcoLogistics self-monitoring tool" that cities can use to estimate greenhouse gas emissions from urban transport. The tool also acts as a monitoring tool for cities to compare and analyze indicator data.	Input WP2-WP6 • EcoLogistics self-monitoring tool; estimation of greenhouse gas emissions. • Compare and analyze indicator data.

Table 6 – Overview of EU and non-EU funded projects and how they feed into COLUMBUSS

Project name	Description	How it feeds into COLUMBUSS
LEARN	Logistics Emission Accounting and Reduction Network. The overall goal of LEARN was to empower businesses to reduce their carbon footprint across their global logistics supply chains through emissions measurement, reporting and verification.	Input WP6 • Methodology for emissions measurement, reporting and verification.
AEROFLEX	The AEROFLEX project developed and demonstrated new technologies, concepts and architectures for complete vehicles that are energy-efficient, safe, comfortable, configurable and cost-effective. At the same time, ensuring that the varying needs of customers are satisfied by being flexible and adaptable to the continuously changing operational conditions. The truck, the dolly and the trailer are ready to undertake the test program.	Input WP1 • Customers' satisfaction regarding the design of AVs. • Design of AVs.
Foceta	Modelling and simulation of sensing and perception systems.	Input WP1 • Useful for specification technology methodology requirements and protocols. • Simulation of sensing and perception system.
Urban Smart Park	Improve urban spaces by automated smart parking.	Input WP1 • Safety of complex systems.
Cloud-LSVA	Automatic simulation scenario extraction.	Input WP1 • Critical scenarios.
CoEXist	Prepare shared road networks for connected AV.	Input WP1 • Critical scenarios. • Connected AVs.
HEADSTART PEGASUS MOOVE	These projects have been working on data collection for testing and validation procedures and (accident) simulation. These projects use absolute safety metrics and focus on (private) cars driving on motorways and distributor roads.	Input WP1 • Data collection for testing. • Validation procedures. • Critical scenarios.
HEAT	A project executed in Hamburg, Germany, where low-speed shuttles move through existing traffic on open roads, though on a limited track with an operator/steward on board.	Input WP1 • Driving in natural traffic conditions.
AVENUE	Several pilots have been working towards permanent operations, like the AVENUE project.	Input WP2&3 • Business cases.

Table 6 – Overview of EU and non-EU funded projects and how they feed into COLUMBUSS

Project name	Description	How it feeds into COLUMBUSS
CAMU CAMR ICVDE FTZ	WM is partner in several projects of the Centre for Connected and Automated Vehicles (CCAV), and UK Government and Department for Transport (DfT) with a total budget of 37.7 million pounds: 1) Connected and Automated Mobility Urban testbed. 2) Connected and Automated Mobility Rural Interurban testbed. 3) Connected Vehicle Data Exchange. 4) Future Transport Zones.	Input WP1 • 5G network. • Transport Zones. • Co-creation with citizens.
AVLM CAVIDOR	rT, RET, and GD partner in the AVLM Programme of Metropolitan Region Rotterdam The Hague to demonstrate last-mile autonomous transport. robotTUNER is the lead partner in the CAVIDOR project concerning the certification of autonomous transport in public spaces. AVLM budget is 15 million euros.	Input WP1-WP6 • Certification. • Automation of public macro, medium and microbuses.

6.1.1.2.3 Co-creation

Passengers', road-users and equity owners' opinions influence the successful introduction of AVs in the transport system. Surveys have limited predictive strength without exposure to autonomous driving and interaction with AVs in realistic situations. The involvement of users in developing the technology and services can positively influence the perception and satisfaction and improve the deployment quality. Therefore co-creation activities are organized at all pilot sites (see Table 7).

Table 7 – Overview of co-creation activities carried out in COLUMBUSS

Creation	Description	Effect study
AV-game	Online interactive game for knowledge transfer and training	Knowledge transfer showed in interaction with the AV as a passenger and road user.
Virtual Assistant	Adaptive human-machine interface in the AV	Improved quality of use and passenger satisfaction.
Signs	Led communication lights outside the vehicle	Improved interaction with other traffic and road-users satisfaction.
Simulator	Driving simulation; human-in-the-loop	Improved interaction with road users and enhanced interaction with human drivers and supervisors.
VR	Simulation study	Willingness to share vehicles and lower private ownership; greener the city.

In the simulation study, we use insights from the pilot experiments to visualize a world in VR where shared AVs (in cooperation with transport companies) take a more commonplace part in the multimodal transport system and the landscape/neighbourhoods. We invite citizens to participate in this experiment. We aim to study how different system characteristics would influence residential satisfaction, perceived well-being influences perceived equity, and what results in private and commercial vehicle ownership in **2030**.

Special attention is paid to involving vulnerable groups by co-creation; children, the elderly, the disabled, and vulnerable road users. E.g., partner Veilig Verkeer Nederland has more than 10 000 volunteers who can participate in co-creation and surveys or mobilize their network to participate. West Midlands has created a community of more than 1 000 volunteers participating in co-creation and surveys. Vulnerable groups are sufficiently represented in both volunteer communities.

6.1.1.2.4 Social sciences and humanities

COLUMBUSS prioritizes the users' experience and aims for maximum impact in the applicability and acceptance of autonomous (public) transport. Acceptance of AVs by passengers and road users is crucial for successful commercialization. Therefore, social sciences are a pillar in the validation work package. COLUMBUSS involves engineering scientists and scientists in behavioural and social sciences. They use their expertise to improve the safety metrics of the AutoExaminer, the safe driving skills of the AutoPilot, and security and model driver reference models. They will evaluate the impact of social, environmental, economic and artificial intelligence in (smart) logistics and mobility. Differentiation based on gender, abilities, and age receives extra attention from all involved scientists.

6.1.1.2.5 Gender dimension

Transport of passengers and goods traditionally is male-dominated. Taking gender-sensitive needs and values of transport into account is highly relevant. These values could include speeding, (perceived) safety, ease of use and personal contact. Making transport policy more responsive to the needs and values of women requires a structured approach to understanding their needs, identifying instruments to address these needs, analyzing the costs and benefits of those instruments and establishing an appropriate policy framework. Transport investment planning and design processes should incorporate female representatives.

COLUMBUSS will pursue gender balance in the co-creation activities with citizens, integrating gender equality into the goals of these activities. Gender equality will require female participation, preferably around 50%.

Gender inclusiveness will also be carefully addressed throughout the project's activities, including the language used for communication, outreach, and engagement, gender-neutral language and non-discriminating pictures and visuals.

6.1.1.2.6 Open science

The COLUMBUSS consortium applauds the open science principles. Innovation and implementation will gain speed and quality by sharing knowledge, combining expertise and building on each other's work. The COLUMBUSS consortium actively seeks feedback from third parties through the dissemination activities and the active involvement of human interest parties like our partner VVN ('Veilig Verkeer Nederland').

Open science principles are at the heart of COLUMBUSS: the COLUMBUSS software technology shall be made available as a shared source technology. The consortium will share all evaluation data gathered in the performed demonstrators of this project. This way, all interested organizations can access and start using the technology.

The same goes for the shared evaluation framework. Within the COLUMBUSS project, data is collected through the pilots. This includes information on the performance of the vehicle, (anonymized) data on passengers and users, and data on the vehicle's surroundings (traffic interactions, environmental effects, business data) and pilot set-up (routes, time schedules, spatial data). Protocols for data collection and storage will be formulated.

Relevant raw evaluation data gathered within the project will be made available through data repositories on the COLUMBUSS shared open-source platform, taking data security and privacy protection issues into account. Third-party researchers can evaluate autonomous mobility pilots or performance data accordingly.

All parties generating and/or collecting data and other research outputs will manage these according to the FAIR principles (findable, accessible, interoperable, reusable). In the preparation phase of the project, a data management plan is formulated according to the FAIR convention. The protocols developed in WP6, and the data management plan (part of WP8), will ensure findability and (safe) accessibility of both data and other research outputs, including identifiers and trusted repositories.

6.1.1.2.7 Data management and management of other research outputs

The management of knowledge will be the shared responsibility of all project partners. The project scientists jointly make specifications and formulate protocols to ensure that all generated knowledge is sufficiently and clearly documented in the COLUMBUSS shared open-source platform and well communicated within the project and the outside world. The developed protocols will ensure that reports are adequately detailed, and publications will be encouraged.

Video data of the surroundings of the COLUMBUSS vehicles is collected, anonymized and stored.

In researching the social impact of autonomous transport, data on citizens might include social and financial background issues, age, and possible disabilities. The data will be anonymized and stripped from any recognizable connection to a specific person or class of persons.

Refusal to answer privacy-related questions will be respected. The consortium will adhere to the general data protection regulation (GDPR) set out by the European Commission and implemented by the various national governments. Entities outside the EU will conform to their federal/national legislation on data protection and privacy regulations and respect the EU GDPR. A data management plan will be delivered by month 6.

6.1.1.3 Impact

6.1.1.3.1 COLUMBUSS' pathways toward impact

The COLUMBUSS consortium claims that AV technology used to automate public transport will significantly impact societal challenges in the short and long term. We estimate that the COLUMBUSS assessment safety metrics will be a standard by **2040** and that transport operators are experimenting inside the Netherlands with autonomous public transport using the COLUMBUSS shared open-source technology. In **2050, 50%** of public transport will autonomously use the shared open-source COLUMBUSS AutoPilot as a driver. As a result, residential areas and city centres have **25%** fewer private and commercial vehicles. Cities are healthier because of less dangerous emissions and are greener.

To reach this point, we will demonstrate autonomous public transportation of people and goods in natural traffic from door to door and between villages and cities at SAE Level 5. AV technology will be used and evaluated for three years. The autonomy level adapts to the performance of the autopilot software. The safety risk of autonomous driving in natural traffic is minimized, starting with shadow driving. Shadow driving lowers the barriers to receiving driving permission and enables autonomous driving under natural traffic conditions 24/7.

Several pilot routes with varying speeds are selected, connecting villages, residential areas and city centres in the Metropolitan Region Rotterdam Den Haag (MRDH) and the Province of Groningen. We expect to drive more than **2.7 million** kilometres in three years. This vast number of kilometres under natural traffic conditions and without human interference strengthens the validity and predictive value of the impact studies for the scale-up of autonomous transportation of people and goods in the Netherlands.

We use Dutch standard legal safety metrics for training and assessing AVs' capabilities. This fundamentally different approach provides certifying bodies, such as the Dutch RDW, CBR, and insurance companies, with an easy and efficient methodology to quantify the risks and severity of accidents. Using standard legal safety metrics for human drivers enhances the trust of other road users and passengers in the adequate driving capabilities of AVs.

The COLUMBUSS solution enables for the first time the study of the impact of autonomous transportation without disturbing limitations, as opposed to many experiments and pilots that are performed under (semi-) laboratory conditions, such as transportation at low speeds, in restricted areas and tracks, and mostly with limited interaction with other road users. The COLUMBUSS technology and pilot partners supply three types of autonomous buses for deployment on all pilot routes. The whole public transport chain is covered. Various transport challenges will be solved, making a proper assessment of social, economic, and environmental impacts possible. COLUMBUSS will demonstrate the benefits of fully autonomous public transport to all target groups. These groups are:

- Governments (local, regional, national).
- Certifying bodies.
- Insurance companies.
- AV-manufacturers.
- AV-service providers.
- Consumers and consumer interest groups, representative groups on traffic safety.

6.1.1.3.2 Breakthrough in proper safety metrics

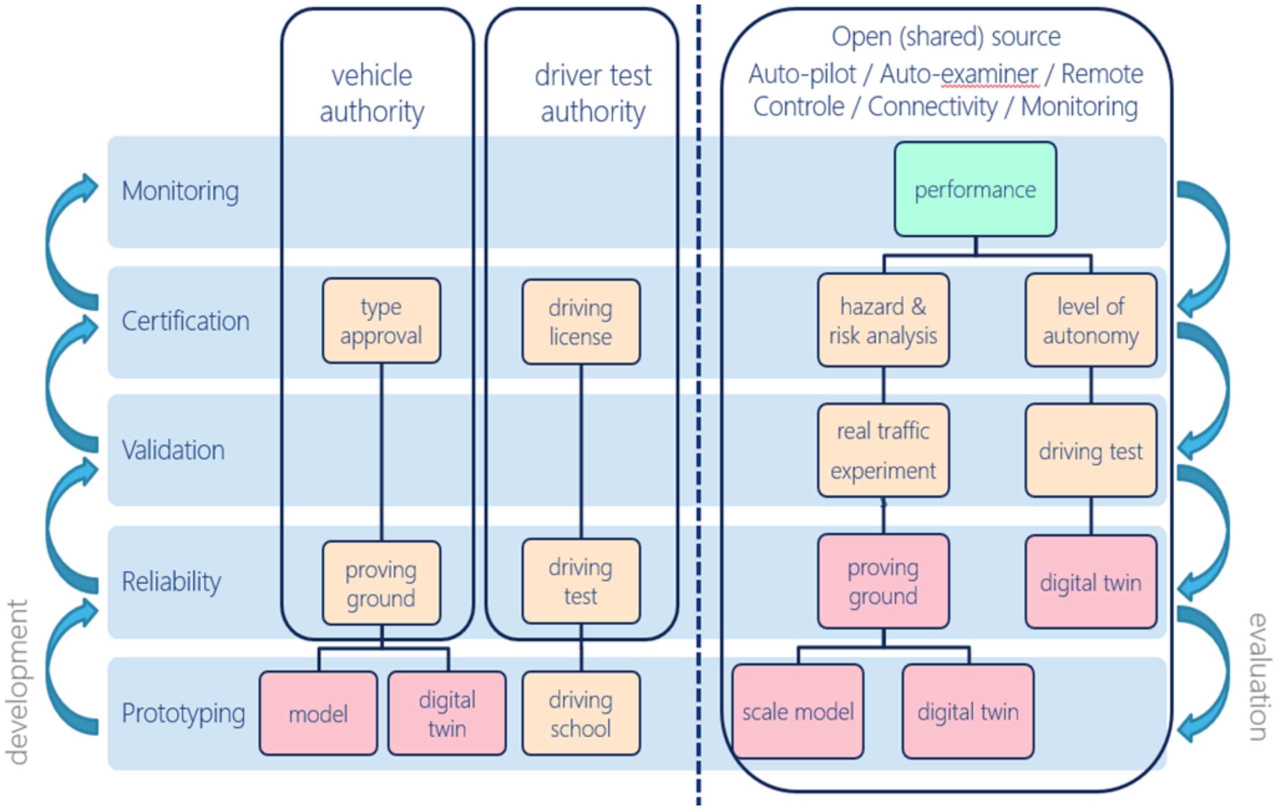


Figure 6 – Left: currently, vehicle and driver are certified separately. Right: COLUMBUSS combines both, including monitoring, certification and validation, assuring reliability and providing room for prototyping

The consortium partners believe that the COLUMBUSS approach focused on human-oriented safety metrics forces a breakthrough in the deadlocked discussion on proper safety metrics in licensing AVs and provides a standard platform that *bridges the gap between vehicle permitting and driver licensing* (Figure 6). We think this will speed up commercial licensing and scale up the roll-out of AVs. Legislating authorities, like RDW, are aware of the importance of the COLUMBUSS approach. They welcome experiments as performed within COLUMBUSS. E.g., COLUMBUSS associated partner Australia and New Zealand Driverless Vehicle Initiative will use COLUMBUSS outcomes to model legislation for licensing AVs. This legislation should be settled in 2026. After legislation in Australia, legislation in Europe is expected to follow.

6.1.1.4 Summary

The COLUMBUSS project demonstrates the social, economic, and environmental benefits of autonomous public transport in two regions of the Netherlands. The project legacy is an open-source platform with AutoPilot, AutoExaminer and AutoControl software free of charge for usage in public transport. This legacy will scale up the commercial use of automated public transport inside and outside the Netherlands and catalysts for the electrification of fuel-powered vehicles and decrease personal ownership of cars.

6.1.2 Metrics

6.1.2.1 Safety states framework – SSF

The safety states framework (SSF) is a metric for objectivation of safety during operations performed by human or artificial pilots. It automates the selection and assessment of behaviour[2] (Figure 7). The SSF applies to the road, rail (tube), air and water traffic systems.

The targeted pilots are part of traffic systems. Traffic systems optimize pilots' travel, balancing time and costs. Traffic rules describe procedures to achieve balanced conditions; **the safety states**. Pilots tune their performance during their journey to meet these safety states. This tuning is information management. Variances on safety states lower the system stability/ performance. Pilots are trained to mitigate risks for system stability. The training of human pilots is based on explainable information management with driving tasks, embedded in traffic laws. For auto-pilots, the SSF uses explainable AI based on human driving tasks.

The SSF calculates the complexity of traffic situations for configured points/ states on the selected route and possible routes and forecasts the risk of collision using a **safety state score (SSS)**. The SSS parameter is an expression of the complexity of a future state in case the pilot continues its current behaviour. A high SSS correlates with a low complexity state, and a low SSS correlates with a high complexity state. Pilots can use this information to optimize behaviour and change routes if applicable. The current driver performance is represented using the **driver safety score**[3] **(DSS)**. The DSS values faults and violations and cumulates multi-factor scores in one safety score. The DDS is part of the virtual instructor tutoring system of Green Dino BV, explained in clause 6.1.2.2.

The SSF uses the pilots' mitigation performance for the behaviour assessment. The parameter for mitigation is called **pilot mitigation score (PMS)**. The PMS expresses the difference in safety between the forecasted SSS with the current SSS and the performed behaviour reaching the future point on the route or adjusted route. The performed behaviour is the difference in DSS between the two states. A low PMS correlates with dangerous behaviour, and a high PMS correlates with low dangerous behaviour.

> For example, to avoid a collision with a pedestrian, the pilot adjusts the route and collides with a fence. The forecasted SSS of the collision point on the route was higher than the current SSS, although the SSS of the current point increased due to the collision with the fence. The PMS increased between the two states of the current points on the route. This enables assessing dangerous behaviour in low complexity situations with high SSS and safe behaviour in high complexity situations with low SSS. The DSS decreased less than forecasted.

The SSS value expresses the probability of a future collision based on the current behaviour; the higher the risk, the more likely the pilot will collide in the case mitigating action is not taken. This SSS parameter is not enough to assess the pilot's behaviour because other traffic can also perform mitigating risk actions. Therefore the SSS also measures the effect of the pilot's behaviour on the complexity expressed in the PMS. The pilots' state change should not contribute to increased risk beyond the selected thresholds and should represent the awareness of possible future collisions defined in the DSS.

The traffic systems have rules for pilots with indicators and thresholds for desired and minimal performance. These rules are embedded in law and named *driving procedures* for the road traffic system. The driving procedures are the algorithms of the road traffic system for operating vehicles.

---

[2] Patents P100268NL00, P100268EPPC, P100268USPC, P100268INPC, P100268CAPC, copyright i-DEPOT 134346.

[3] The Driver Safety Score is a validated multi factor metric for assessment of human drivers and copyright i-DEPOT 116024.

The PMS parameter relates the individual performance (DSS) in selecting and executing driving procedures to *peer performance* to classify variance. The classification is a percentile score.

The SSF uses the performance of human pilots as a baseline reference to assess auto-pilot performance. Arguments are: 1) human performance will influence the peer group averages (both humans and AI) for a long time, 2) it will take a long time to have enough exposure data to prove the validity of new thresholds, 3) driver assessment agencies will not be able to change to a new assessment methodology for AI drivers easily/quickly.

Two indicators/ parameters for risk mitigation behaviour by human vehicle operators are validated and widely accepted. These parameters are viewing (active field of view) and speed. Both correlate; higher speed results in a narrower field of view (tunnel vision). Driving at higher speeds increases the need for accuracy. Accuracy forces the operator to focus and narrows the active field of view. More information needs to be processed: workload increases. Driving tasks cannot be selected or performed at specific workload levels without faults. Viewing has four main components: the selection, the identification (focus), the processing of identified information and the reaction/ response. In the case relevant information is missed or not processed correctly, and or the reaction is not proper in respect to the driving procedures, the risk value increases, and the driver safety score decreases.

External parameters such as environmental conditions like blocking vision (buildings, weather) influence the calculated risk value, but also internal parameters of the operator such as training history (software updates), history of performance/ incidents, and soundness of the information processing system.

☑ Raycasts
☐ Predictions
☐ Fast Trajectory
☑ Manage Risk
☑ Intersection Plans
☑ Additional Plans
☑ Emergency Brake
☐ Direction Control

Preset plans: 1 2 3 Default
Slow above 0.8 , stop above 1
Keep distance margin [m] 1
Go off-lane to evade risk? ☑
Risk difference threshold 0.1
Cross mid-line threshold 0.2
Minimum distance to intersection for evade [m] 1

Risk levels: Warning 0.5, Danger 0.75

TimeScale 1.00
TimePoint 1.2 s

Cyclist Intersection

Figure 7 – Remote monitoring and risk assessment with the SSF: the ego vehicle stops the overtaking manoeuvre of the cyclist based on the forecasted risk (grey and orange ghost cars). It passed the maximum risk threshold. It remains behind the cyclist because cars could enter the crossing from the left road at the intersection

Click on the link for a video demonstration of the SSF https://youtu.be/8t2DB-koQ9o.

Click on the link for a video of the first implementation of the SSF in the Green Dino truck and bus simulator https://youtu.be/Qbt73TWmzq4.

Click on the link for a video of the first implementation of the SSF in the Green Dino multipurpose vehicle https://youtu.be/WwFxIRVEWuc.

6.1.2.2 Virtual driving instructor

Green Dino offers driving simulators with automated feedback for training and assessing learner and experienced drivers[4] (Figure 8). The computerized feedback system works with driving tasks and instruction levels. The development of the virtual driving instructor (VDI) started in 2003. Up to now, more than 200 000 drivers have been trained and assessed. The automated feedback system is validated by 4 PhD researchers (de Winter, de Groot, Huizinga, Kuipers) and classified as reliable and valid.

Figure 8 – Driving simulator, type Drive Master LT

The main principle behind the tutoring system is the execution of driving tasks. The driver must automate a driving procedure to pass a driving task. The didactical method is called procedural training and assessment. Every (complex) driving task is constructed from smaller sub-driving tasks. A sub-driving task is separated into small steps.

6.1.2.2.1 Adaptive instruction

Too much or too few instructions will slow down the automation of behaviour. The separation of driving tasks and adaptive instruction is used to balance the feedback with the performance of the learner driver. The driver is ready for a new driving task in case the active driving task is automated: the learner driver performs the procedure step by step without mistakes and feedback from the VDI.

The VDI uses three instruction levels (levels of automation). The three instruction levels are:

1) always do on instruction; the student gets precise instruction on every step in the driving task procedure;
2) do with less instruction; the student receives indirect instruction;
3) do without instruction; the student gets no instruction and is only corrected.

---

[4] The software copyright protected i-DEPOT 116024.

Too much or too few instructions will slow down the automation of behaviour. Adaptive instruction is used to give balanced feedback. Separation of driving tasks is also necessary to enhance the automation of driving tasks. A new driving task should only be activated when the active driving task is partly or wholly automated. Otherwise, the workload for the student will be too high, and the learning process will slow down. Assessing driving tasks in real-time is necessary to make adaptive instruction possible.

Show Report Options

Strength/Weakness Report

Click on the lesson date or score to view the lesson results.

Explanation

Your Strength/Weekness Score is 6.5.
You are learning less fast than the average student.

If you have not mastered the drivingtask (score les than 5.5) you should repeat the lesson.
The score of the average student is 7.5.

Explanation of the used colors.	
	You did not perform this drivingtask yet.
🟥	You are learning less fast than the average student. You should do the lesson again.
🟨	You are learning just as fast as the average student.
🟩	You are learning faster than the average student.

Pauline van E		1 21 Jul	2 21 Jul	3 21 Jul	4 25 Jul	5 25 Jul	6 25 Jul	7 29 Jul	8 29 Jul	9 01 Aug	10 01 Aug	11 01 Aug	12 05 Aug	13 05 Aug
	6.5	6.1	7.3	10.0	5.0	8.1	8.1	7.6	9.1	7.6	7.8	7.7	7.1	7.9
Vehicle control														
Taking a gentle bend *Too much steer*	21 / 23 7.8 3		2 / 3 4.0 1	3 / 3 10.0 2			2 / 2 10.0 3	6 / 7 5.4 3	6 / 7 7.4 3		1 / 1 10.0 3	1 / 1 10.0 3	1 / 1 10.0 3	1 / 1 10.0 3
Taking a normal bend *You approached the bend too fast*	35 / 37 7.0 3		2 / 2 10.0 1	1 / 1 10.0 1		2 / 2 10.0 2	6 / 7 1.6 3	11 / 11 10.0 3	13 / 14 6.1 3				1 / 1 10.0 3	1 / 1 10.0 3
Taking a right-angled bend *Too much steer*	64 / 79 7.1 3	0 / 2 1.0 1	7 / 8 8.9 3	7 / 7 10.0 3		15 / 22 3.9 3	13 / 16 10.0 3	13 / 17 4.3 3	13 / 14 10.0 3		1 / 1 10.0 3	1 / 1 10.0 3		1 / 1 10.0 3
Use of the accelerator														
Moving off *The engine has stalled or you did not anticipate in good time*	48 / 54 7.2 3	1 / 1 10.0 1	8 / 8 10.0 3		14 / 20 2.4 3	7 / 7 10.0 3	1 / 1 10.0 3	2 / 2 10.0 3	8 / 8 10.0 3	3 / 3 10.0 3	2 / 2 10.0 3	4 / 4 10.0 3	3 / 3 10.0 3	4 / 4 10.0 3
Change up a gear *You let the clutch come up too quickly*	14 / 19 2.5 1				14 / 19 2.5 1									
Change down a gear	3 / 3 10.0 1				3 / 3 10.0 1									
Use of the brake pedal	6 / 6 10.0 3					6 / 6 10.0 2								
Position within the lane *You drove too far right off centre*	454 / 475 8.9 3	6 / 6 10.0 2	42 / 47 5.6 3			87 / 99 6.4 3	50 / 55 9.1 3	88 / 90 10.0 3	124 / 125 10.0 3	14 / 14 10.0 3	29 / 29 10.0 3	29 / 29 10.0 3	14 / 15 3.4 3	19 / 19 10.0 3
Maintain safe distance from car in front	7 / 7 10.0 3							6 / 6 10.0 3	1 / 1 10.0 3					
Moving off after stopping away from traffic *The engine has stalled or you did not anticipate in good time*	3 / 10 1.0 1							3 / 10 1.0 1						
Park	4 / 4 10.0 2							4 / 4 10.0 2						
Keep to maximum speed *You drove too fast*	29 / 31 8.3 3		1 / 1 10.0 1						7 / 7 10.0 3	0 / 1 1.0 3	7 / 8 4.4 3	10 / 10 10.0 3		5 / 5 10.0 3
Crossings (basics)														
Go straight ahead an unmarked junction *No clear priority given to traffic coming from right*	33 / 35 8.1 3	6 / 6 10.0 3								14 / 14 10.0 3	9 / 9 10.0 3			10 / 12 4.4 3
Turn right at unmarked junction *You took the bend with the clutch depressed*	30 / 35 7.7 3	1 / 10 1.1 1								9 / 12 6.9 3	5 / 5 10.0 3		8 / 8 10.0 3	8 / 10 5.7 3
Turn left at unmarked junction *No clear priority given to oncoming traffic*	25 / 34 4.1 3	3 / 10 4.4 1								13 / 15 8.0 3	4 / 6 1.0 3			8 / 12 1.1 3
Turn left at junction with traffic lights	3 / 3 10.0 3		1 / 1 10.0 1								3 / 3 10.0 2			
Turn right at a junction with traffic lights *Wrong use of the indicator*	5 / 6 9.9 3		0 / 2 1.0 1								5 / 6 9.9 3			

Figure 9 – Overview of the driving lessons. The levels of instruction/automation (1, 2, 3) are displayed in the white boxes

Colours based on traffic light colours indicate competence levels. Red stands for the first level of competence, yellow for the second level and green for the third level. When the learner driver has green for all the driving tasks, he or she is ready to drive on the road.

An extra instruction level (0) is available for briefing before the virtual driving starts. Html pages present explanations in text and audio. Videos invite the student to copy the actions of an actor on screen.

In the administration system, the achieved instruction levels per lesson are visible (Figures 9 and 10A). Quick students can reach level 3 within one lesson. An average student will sometimes need two lessons. An overview of faults is available.

The feedback generation of the virtual driving instructor is related to the operation and tactical level of the GDE matrix.

6.1.2.2.2 Strength and weakness report

In the Green Dino simulators, offline assessment (debriefing) is done with a strength and weakness report and failure specification available in the administration system and Internet (Figure 8). The strength and weakness score is calculated based on the performance of other simulator drivers (Figures 10A and 10B). The S&W score is a percentile score: a 5.5 expresses that 45% of the students made fewer mistakes on the task, and 55% made more mistakes.

The student must follow the lesson again if the score is below 5.0. A red cross will appear in front of the lesson, and the same lesson will automatically be added to the queue. If the score is above 5.0, the student is ready for the next lesson, and a green check mark will appear in front of the lesson.

Vehicle control			
Taking a gentle bend *Too much steer*	21 / 23	**7.8**	3
Taking a normal bend *You approached the bend too fast*	35 / 37	**7.0**	3
Taking a right-angled bend *Too much steer*	64 / 79	**7.1**	3
Use of the accelerator			
Moving off *The engine has stalled or you did not anticipate in good time*	48 / 54	**7.2**	3
Change up a gear *You let the clutch come up too quickly*	14 / 19	**2.5**	1

Figure 10A – Strength and weakness score

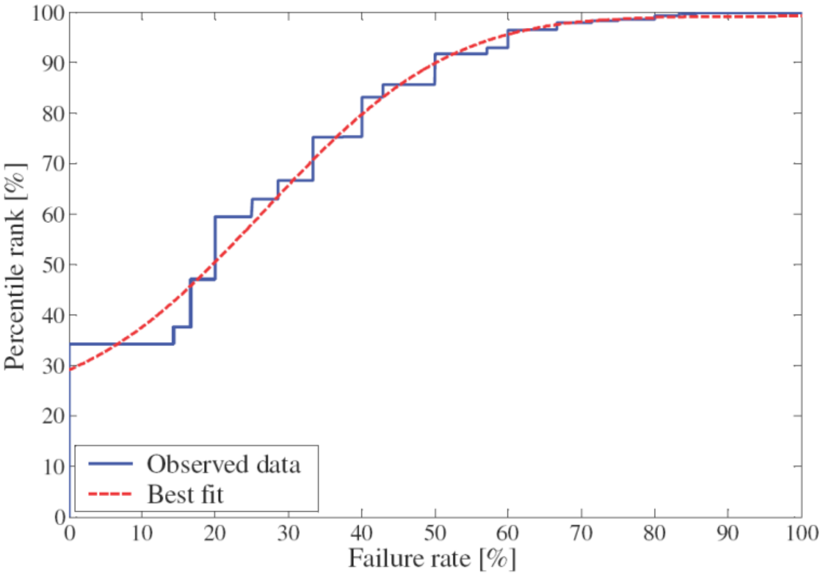


Figure 10B – Reference performance curve on a specific driving task

6.1.2.2.3 Driving style report and safety report

The driving style report superimposes the driving performance from the safety report into a driving style. The driving style report reflects the drivers' competencies and willingness to take risks (attitude) in traffic. The driving style classification supports drivers, instructors and examiners.

The main goal of the driving style report is to classify driving behaviour into predefined profiles, known as the driving styles. The driving style analysis provides comments about aspects of the observed driving behaviour and tips and tricks to improve the personal driving style. It is a multi-variable assessment framework.

6.1.2.2.4 Classification of driving style

Driving behaviour is derived from two main contributors:

1) level of driving skill

2) willingness to take risks

The 'Driving skill level' can be evaluated by considering different aspects of the driving skills. This includes factors related to vehicle control (speed control, steering, braking, shifting gears, etc.) and behaviour in traffic context (obeying traffic signs, following right-of-way rules, using indicators, etc.). The 'Taking risks level' depends on the driver's character and the situational contexts.

The driving style profiles are described using pairs of values estimating levels of the two components of driving behaviour:

The driving style is classified as *'Hazardous'* when the readiness to take risks is very high (VH). It is classified as *'Unskilled'* when the driving skills are very low (VL).

When the taking risk aspect is high (H) and the driving skills are low (L) or average (A), the driving style is classified as *'Overconfident'*. The driving style is classified as *'Aggressive'* when the driving skills are high (H) or very high (VH).

When the willingness to take risks is medium (M), the driving style is classified as *'Confident beginner'* when the driving skills are low (L) and *'Average driver'* when they are average (A). The driving style is classified as *'Good driver'* when the driving skills are high (H) or very high (VH).

When the willingness to take risks is low (L), the driving style is classified as *'Beginner'* if the driving skills are low (L), and as *'Cautious average skilled driver'* if the driving skills are average (A). The driving style is classified as *'Too Careful'* when the driving skills are high (H) or very high (VH).

In case the willingness to take risks is very low (VL), the driving style is classified as *'Cautious Beginner'* if the driving skills are low (L), and as *'Too Cautious Average Skilled Driver'* if the driving skills are average (A). The driving style is classified as '*WayTooCareful'* if the driving skills are high (H) or very high (VH).

6.1.2.3 Safety report

The calculated skill and risk values used for the driving style assessment are stored in the safety report (Figure 11). The summarised safety score is an expression of road safety measured in a certain period. The safety score comprises the summarised scores from different categories of driving tasks.

Safety Report

Click on the lesson date to view the DrivingStyle Profile.

Green Dino	12 Apr
General summary	
Driving skill	8.3
Safety score	5.9
Avoiding risks	3.3
Economical driving	2.3
Summary by categories	
Vehicle control	7.2
Observation and anticipation	7.5
View behaviour	7.3
* before turning left	8.0
* before turning right	4.0
* before going straight on	10.0
* before entering a roundabout	10.0
* before braking	3.5
* scanning	8.2
Keeping safe speed	0.0
Keeping fluent speed	9.6
Keeping traffic rules	10.0
Avoiding traffic accidents	7.7

Vehicle control	
Position inside lane	5.7
Smooth steering	6.1
Precise steering	7.1
Shifting up in time	3.7
Shifting down in time	9.0
Observation and anticipation	
Keeping distance to preceding car	4.8
Smooth braking	7.7
Keeping safe speed	
On straight road segments	0.0
In curves	7.8
When approaching intersections	10.0
* and need to stop	10.0
* turning right	10.0
* going straight	10.0
* turning left	10.0
When crossing intersection	8.2
* turning right	4.9
* going straight	10.0
* turning left	9.6
On roundabouts	2.3

Keeping fluent speed	
On straight road segments	9.4
When approaching intersection	7.0
When crossing intersection	9.9
On roundabouts	9.0
Traffic rules	
Stopping for traffic lights	10.0
Indicators usage on intersections	10.0
Indicators usage on roundabouts	10.0
Accidents (number)	
Collisions with other traffic	0
Onesided collision	0
Offroad	0
Partially offroad	1

Figure 11 – Safety report

6.1.2.4 Effectiveness of the automated driving instruction and assessment

In 2007 J. de Winter, PhD researcher in the Virtual Assistant project, compared the performance on the first exam between simulator students and students who followed only car lessons. Questions to be answered were 1) do simulator students have a higher score on their first exam than students who didn't follow simulator lessons? And 2) is it possible to predict the number of lessons and the performance on the first exam in an early stage of the driving course (after the simulator lessons)? The results showed a 3.7% to 5.7% higher score of simulator students on their first exam than the national average. Driving schools with a Green Dino simulator have a 7.5% higher score on first exams.

In 2019 C. Huizinga, PhD researcher at the Centre of Human Drug Research (Leiden), published a validation study of the driver safety score (DSS) for medicine testing. This clinical study proved the correlation between alcohol levels in the blood and the DSS. Based on this study, the US Food & Drug Agency accepts the DSS for clinical medicine tests.

J. Kuipers (2022) showed that a group of Dutch drivers who received simulator training between 2008 to 2015 had a lower (severe) accident risk after the first year of driving than learner drivers who completed automated simulator training or regular road training without automatic feedback on viewing skills. Drivers who followed eight or more lessons on a simulator and received automatic FieldOfView-training had a 36.8% lower accident risk in the last 12 months of driving than drivers who only followed regular on-road lessons (Figure 13). Drivers who followed hazard perception training and FOC- training on a simulator reported 72.5% fewer accidents risk in the first year of driving than drivers without hazard perception training with equal exposure to traffic (Figure 12). Driving simulator lessons replaced on-road lessons with a factor of 1.6.

Figure 12 – Hazard perception training. A cover situation on a crossing turning left

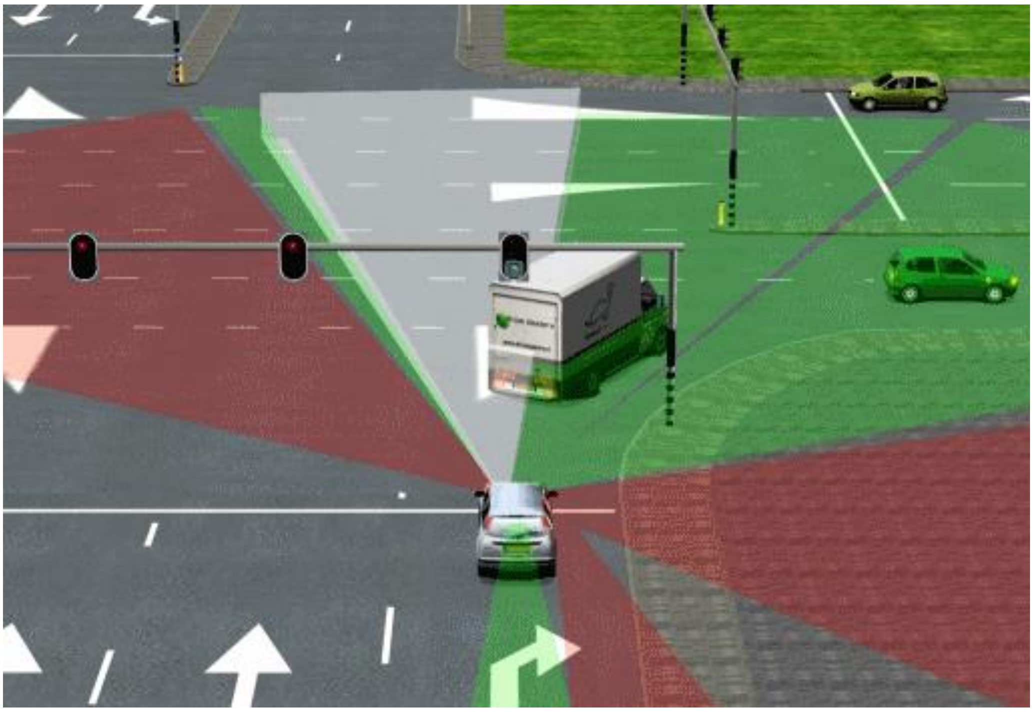


Figure 13 – Assessment of the viewing behaviour with head tracking. The red areas are projected on the screen to inform the student about the viewing faults

The scientifically validated results of Delft Technical University and the Centre for Human Drug Research show that Green Dino's Virtual driving instructor (VDI) is influential for automating training and assessment of learner and experienced drivers.

6.1.2.5 Literature

Winter, J. C. F. De, Wieringa, P. A., Kuipers, J., Mulder, J. A., & Mulder, M. (2007). *Violations and errors during simulation-based driver training. Ergonomics*, 50; pp. 138–158.

Kuipers, J. (2018) *Automated feedback on viewing skills lowers accident involvement*. Proceedings of the 6th Humanist Conference. The Hague.

Huizinga, C, (2019). *Evaluation of simulated driving in comparison to laboratory-based tests to assess the pharmacodynamics of alprazolam and alcohol*. Journal of Psychopharmacology.

6.1.3 Monitoring

6.1.3.1 Perspective

The remote monitoring of the AVs in the COLUMBUSS project happens from the driver's perspective. The ego vehicle provides sensor information for positioning itself and objectivation of the traffic scenario.

6.1.3.2 AI technology

Human peer performance is used to assess the safety performance of the AVs. For the assessment, a bus driver reference model is constructed using performance data of bus drivers collected during daily work and training in a driving simulator (Figure 14). Legal driving task procedures are used as driver-performance data analysis algorithms (see clause 6.2.2). This methodology supports transparent/white box/trustworthy AI.

Emphasis is laid on the effort of the auto-pilot software itself to mitigate the risk, using the safety state framework (SSF) metric (Figure 15). The SSF metric is the, for automated driving, adjusted and improved version of Green Dino's validated driver safety metrics for human drivers.

The intelligence and traffic logic software runs in the game engine Unity 3D. Green Dino is moving the intelligence and logic software stacks to ROS2 to comply with ISO 26262. The migration of the autonomous driving stack to ROS2 will be completed in 2024. When the ROS2 stack is operational, Unity 3D will remain as a (dumb) visualization client, enabling the human interface.

Figure 14 – Green Dino Bus and Truck simulator

Green Dino 1
Green Dino
20-09-2022 08:45

Snelheid 1 km/u
Toerental 100 rpm
Verbruik 0.0 L/100km
Afstand 0.2 km
Gasdruk 0 %
Remdruk 0 %
G-Krachten
Acceleratie 0,00 G
Deceleratie 0,00 G
Links 0,00 G
Rechts 0,00 G

Figure 15 – Bus & Truck simulator with safety states framework (SSF) integration: the bus holds and gives the right of way to the car, the calculated risk for a collision with the car is 0

6.1.3.3 Digital Twin

Unity 3D is used as a human interface to visualize the Digital Twin (DT). The static operational design domain (ODD) is generated daily in Unity 3D before operations using open data sources like OpenStreetMaps (Figure 16). This way the ODD is up2date. Open traffic information is used for the determination of road works/ blocks, etc.

The dynamic ODD is constructed in real-time using sensors, like radars of Continental that can detect objects and their dynamics instead of raw point clouds: https://www.continental-automotive.com/en-gl/Passenger-Cars/Autonomous-Mobility/Enablers/Radars.

Figure 16 – Synchronized Digital Twin driving 6.76 km/h with a Renault Twizy ego vehicle: ODD is constructed with information from OpenStreetMaps and radar (pedestrian)

Detected objects are created as virtual objects in the Digital Twin and synchronized or extrapolated over time (Figures 16 and 17).

Traffic information is used in real-time for the detection of the status of traffic signs/lights and potential dangers like traffic jams/accidents.

Figure 17 – MPV slows down for invisible pedestrian coming from behind the parked van, using the AI in the synchronized DT

Sonar is used for nearby entering of the safety zone of the ego vehicle (by a blockage). A safety zone and blockage are part of the ODD, but have no fixed dimension or identity. Detected zone entrances are synchronized in real-time with the DT.

The sonar sensor that is in use: https://www.robotshop.com/nl/nl/maxbotix-lv-maxsonar-ez0-sonar-module-met-hoge-prestaties.html?gclid=EAIaIQobChMI6_2A9ImR-gIVwp13Ch2b9wkVEAQYASABEgIUK_D_BwE.

6.1.3.4 Calibration

The DT runs synchronized with the ODD. For calibration and synchronization, the ego vehicle's accurate position (coordinates) and orientation are crucial. One or two (large bus) Xsens sensors are used for the location and orientation tracking: https://www.xsens.com/mti-680G?utm_term=mti%20680g&utm_medium=ppc&utm_campaign=3DCA+%7C+North+Amer+%7C+Search&utm_source=adwords&hsa_cam=15269931290&hsa_src=g&hsa_mt=e&hsa_ver=3&hsa_net=adwords&hsa_tgt=kwd-1000087371296&hsa_acc=1306794700&hsa_grp=130165869776&hsa_kw=mti%20680g&hsa_ad=561733607273&gclid=CjwKCAjw1ICZBhAzEiwAFfvFhIJzBErtcN8uoapdCkU2AKuymLI5s5e84FWTpDQoPlCSg7vI2UjSIBoC0m4QAvD_BwE.

6.1.3.5 Remote and safe state controllers

A 4G network is in use for enabling remote speed control and monitoring (Figure 18). Bandwidth and latency issues are mitigated by only transmitting object data and high-level commands. The Digital Twin does not use raw sensor data.

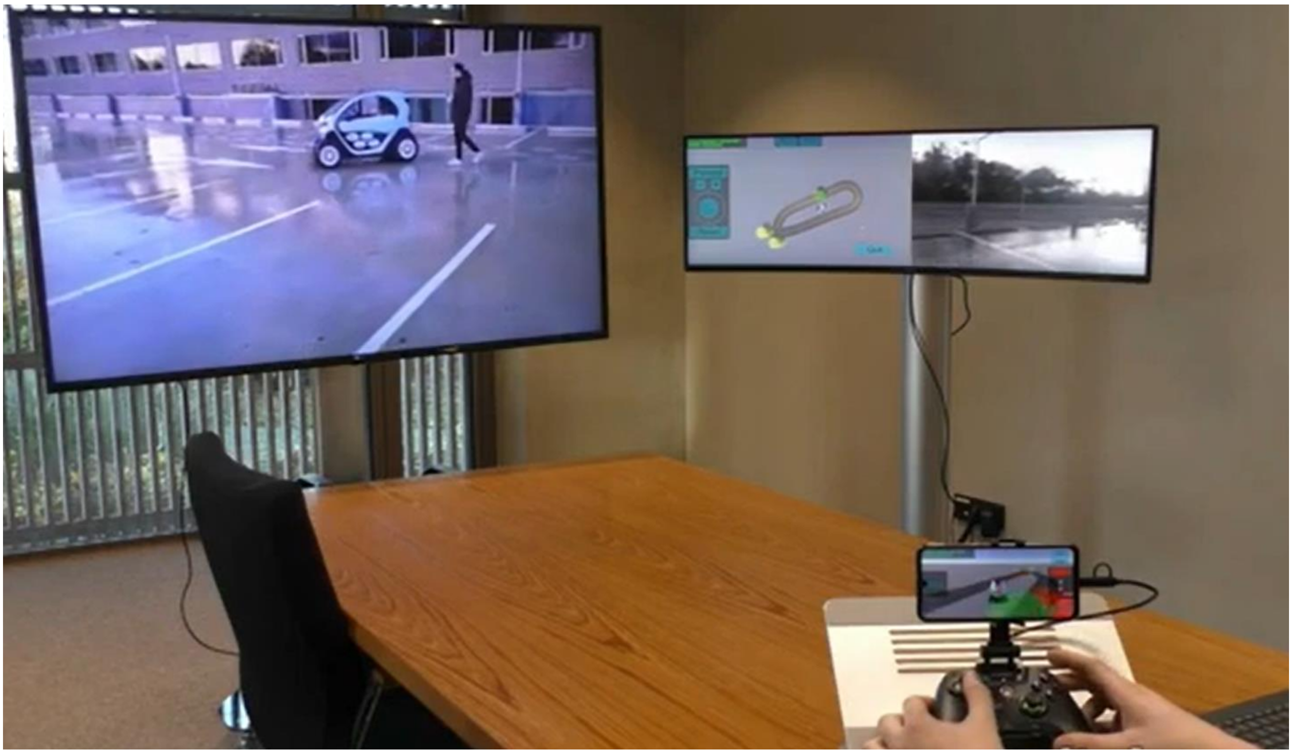


Figure 18 – Remote control in a 4G network. The DT runs remotely on a mobile phone

Green Dino uses a remote AI operator to direct the ego vehicle to proceed with a translation on a predefined path. The remote AI controller first calculates the safety risk for the translation using the SSF (Figure 19). If the action is within the predefined safety margins, it orders the ego vehicle to continue or change speed. If the translation is not within the safety margins, it will command the ego vehicle to change speed. In most cases, to lower speed.

Safety States Framework

☑ Raycasts
☐ Predictions
☐ Fast Trajectory
☑ Manage Risk
☑ Intersection Plans
☑ Additional Plans
☑ Emergency Brake
☐ Direction Control

Risk Management

Preset plans: 1 2 3 Default
Slow above 0.8 , stop above 1
Keep distance margin [m] 1
Go off-lane to evade risk? ☑
Risk difference threshold 0.1
Cross mid-line threshold 0.2
Minimum distance to intersection for evade [m] 1

Risk levels: Warning 0.5, Danger 0.75

TimeScale 1.00
TimePoint 1.2 s

Figure 19 – Remote monitoring and risk assessment with the SSF: the ego vehicle stops the overtaking manoeuvre of the cyclist based on the forecasted risk (grey and orange ghost cars). It passed the maximum risk threshold. It remains behind the cyclist because cars could enter the crossing from the left road at the intersection

The onboard safe state controller is redundant and partly unconnected with the outside world for security reasons. The unconnected part of the controller only looks at the position and orientation of the vehicle body and wheels. The redundant onboard safe state control system's highest priority is directing the vehicle to a safe state on the route. The ego vehicle will go to a safe state if it is not lined up with the predefined path and margins stored in the controllers' memory.

The hierarchy of states is: 1) orientation of the steering wheels, 2) orientation of the vehicle body, 3) position of the vehicle, and 4) speed. The safety state control system's idle state is a fixed position on the predefined path (zero speed). The remote AI controller can only overrule the speed state of the safe state controller. Only speed commands that are authenticated and within the relevant time frame are accepted. The auto-pilot software stack includes the local safe state controller and the remote AI operator.

The 4G network is suitable for remote steering by a human operator at low speeds. However, the human remote control will be limited to changing speed. In case the safety control system goes to the idle safe state, it needs to be reset/start again. This is only possible at the location of the ego vehicle by an authorized human operator.

6.1.3.6 Data storage, cyber security and privacy

All available ego vehicle and ODD information are stored using Microsoft Azure. Only object information and properties are included, no raw data. Playback of recordings in the Digital Twin is supported.

Raw video data is collected in the ego vehicle and stored on portable flash disks. The transport company administrates the flash disks. robotTUNER can only access the raw data with a privacy officer of the transport company on the premise of the transport company. This way, the privacy of individuals is secured.

The data stored on the Microsoft Azure cloud is encrypted and can only be accessed using 'two factor authentication' (Password and Yubikey hardware token).

The hash of data transferred to Microsoft Azure is stored on the Ethereum blockchain. This hash can later be used to prove that stored data was not altered in any way.

All commands received by the ego vehicle (remote AI control and human operator control) rely on end-to-end encryption to ensure that only an authorized client can command the ego vehicle to change speed.

6.1.3.7 Compliance with TR01

The Molly Problem Public Consultation resulted in 16 points of expectation regarding the monitoring of connected and automated vehicles (CAVs) and the capability of the auto-pilot software.

The auto-pilot and the remote control software used in the COLUMBUSS Avs were both constructed by Green Dino. Green Dino stated to be compliant with the results of the Molly Problem Public Consultation outcome. Table 8 lists the compliance for both systems, control and monitoring.

Table 8 – COLUMBUSS compliance with the Molly Problem Public Consultation responses

No.	Would you expect…	Compliance
1	**the software *to be aware* of the collision**	The sonar sensors 'detect' mild collisions. The positioning sensor 'detects' severe collisions. The digital twin, as a third sensor factual, detects/calculates collisions using the static and dynamic ODD information. The digital twin uses extrapolation for the detection of a

Table 8 – COLUMBUSS compliance with the Molly Problem Public Consultation responses

No.	Would you expect…	Compliance
		collision. The sensor configuration forms a redundant collision detection system. Pressure sensors are absolute sensors but are not applicable for automated driving. Collisions are stored in Azure and are replayable in the remote synchronized digital twin.
2	**the software *to stop* at the collision site**	The system design is so-called 'safe state'. The vehicle safe state control unit will go to a safe state without an authorized command for displacement on the route. The remote control software gives this command. If the DT detects or forecasts a collision, all displacement commands are overruled, and the navigation control of the remote operator is shut down. This is also the case with the dysfunction of safety system components. No human or AI operator can overrule the safe state remotely. Overruling is only possible on-site.
3	**the software *to indicate a hazard* to other road users**	The remote digital twin forecasts collisions using the SSF.
4	**the software *to alert emergency services***	This function is not available yet but will be implemented before Level 3 driving on public roads starts.
5	**the software to recall the *time* of the collision**	The ODD is continues stored in Azure included time-stamps.
6	**the software to recall the *location* of the collision**	The Xsense positioning sensor delivers world coordinates used for calibration and synchronization of the DT (that is stored in Azure).
7	**the software to recall *when* the collision risk was identified**	The DT extrapolates the movements of all dynamic objects over time and detects collisions with other dynamic or static objects.
8	**the software to recall *if* Molly was detected**	Using radar object information, Molly would be identified as an object-type 'pedestrian' in the DT.
9	**the software to recall *when* Molly was detected**	The ODD including the Molly object-type pedestrian would be stored over-time with time-stamps in the DT.
10	**the software to recall *if* Molly was detected as a *human***	Using radar object information, Molly would be identified as an object-type 'pedestrian' in the DT. The DT itself classifies the object as human based on parameters like volume, speed and position.
11	**the software to recall *when* Molly was detected as a *human***	The ODD, including the Molly object-type pedestrian, would be stored over time with a time-stamp in the DT.
12	**the software to recall *whether* mitigating action was taken**	Based upon the global risk along the route over time, the SSF calculates the contribution of the

Table 8 – COLUMBUSS compliance with the Molly Problem Public Consultation responses

No.	Would you expect…	Compliance
		ego vehicle to the global risk. This reveals the mitigation performance of the ego vehicle.
13	**the software to recall *<u>when</u>* mitigating action was taken**	The auto-pilot of Green Dino follows legal human driving procedures. Driving procedures are action plans with priority levels. The plan activation by the ego vehicle software using the SSF is recorded and stored in Azure with time stamps. Because the monitoring software also uses the SSF, it stores stamps when specific plans should be activated by third-party software.
14	**the software to recall *<u>what</u>* mitigating action was taken**	Every driving procedure consists of an action list. The list describes the individual actions and their parameters. The activities and parameters are stored in Azure with time stamps.
15	**similar recall abilities for *<u>near-miss events</u>***	The DT with SSF also calculates near-misses in the form of high global risk events. All objects' risks, orientations and positions at that time are stored in Azure with a time-stamp.
16	**expect *<u>driving</u>* to be *<u>prohibited</u>* for software *<u>without recall capability</u>***	COLUMBUSS supports this expectation.

6.2 CETRAN public road demonstrator of metrics based on AV perception system output

6.2.1 Demonstrator goals and objectives

The CETRAN team has been working on a demonstrator to prove that metrics can be obtained from the perception system through continuous evaluation of the object list generated by the AV stack. For this demonstration, the AV stack is effectively a black box and only the object list in Apollo format as generated by the perception stack is used for the evaluation. The metrics evaluated are an CETRAN in-house developed methodology which is a proof of concept only and not proven to give adequate metrics. Both the use of a standards compliant object list and the calculation of standards compliant metrics are deemed to be future work.

According to the CETRAN vision, developing metrics and methods must be split into two separate activities:

- Development of methodology and corresponding metrics for continuous evaluation of real-world trials and operations. This activity is being done in collaboration for International Telecommunication Union (ITU) focus group on Artificial Intelligence for Autonomous and Assisted Driving (FG-AI4AD) which operates under United Nation Economic Commission for Europe (UNECE). In work package 1 (WP1) it is tasked with enabling vehicles to monitor and interact safely with the surrounding traffic environment. Development of these metrics should accept the following boundary conditions:
    1) They will never be perfect, due to infinite number of dynamic situations.
    2) They should be a continuous evaluation during the full duration of the drive.
    3) They need to assume that the automated driving system (ADS) is a black box and vendor specific implementation data cannot be used to create metrics.

4) They could assume that a full detailed HD map is always available, but a trade-off needs to be made based on time available and technical feasibility if a simpler set of metrics based on a non-map dependent is a more realistic implementation for the FG-AI4AD demo.

- Development of metrics for basic driving behaviours. Metrics need to be determined for individual behaviours like the stopping behaviour and associated metric for stopping distance. These individual metrics can then be used to create a set of metrics for a test scenario, either to be applied through simulation or by on-circuit testing, without having to develop an explicit set of metrics for a scenario.

The demonstrator was done with minimal Apollo AV stack, minimal sensor set and a human driver taking the role of path planning and vehicle controls as would normally performed by the AV stack. The sensing and perception system of stack was run in an identical configuration to what would be used in a full AV stack, while GNSS and RTK only was used for localisation due to the unavailability of HD maps.

6.2.2 Technical approach

The focus of this work package (WP) would be to develop a methodology to evaluate the driving risk with the traffic situation. One challenge is to have a continuous evaluation of the dynamic traffic interactions. This means that multiple interactions are continuously evolving, and the risk evaluation methodology should be able to capture this dynamic variation of conditions. The outcome of the risk evaluation methodology should provide a general risk associated to the overall drive with the intent that this could be adopted in assessment of AVs by authorities in the future. This risk evaluation methodology will be based on the application of safety metrics.

A proof-of-concept vehicle-based demonstrator will be developed to apply the use of this driving risk methodology. To achieve this implementation, the approach will be to develop the methodology in intermediate steps, such as using simulation tools in increasing the level of accuracy to reality. It should be kept in mind that a simulated environment will attempt to represent but is unable to fully replicate the true physical environment. Simulation tools are nonetheless utilized due to the ease of scenarios set-up, availability of data and analysis through visualization.

MathWorks' *Driving scenario designer* [10] enables easy design of synthetic driving scenarios, where a road and actor models can be used with a user-friendly drag and drop interface. This is the first step of the methodology development, allowing for iterative adaptations without the concern of data quality that could arise in the vehicle demonstrator.

A second step of the approach is utilizing simulation software that aids in AV development. Two types of simulation would need to run concurrently, one that simulates the vehicle's drive – ADS which would control how the vehicle reacts to scenarios, and another to simulate the environment such as road infrastructure and other road users. Apollo is the choice of ADS simulator, together with SVL from LG for environment simulation. This provides an increased reality of a simulated scenario due to the ADS in the loop, and data errors are expected which could help prepare for the vehicle-based demonstrator.

The final step is the vehicle-based demonstrator, developed by CETRAN to apply the use of this driving risk methodology. This utilizes a Mitsubishi iMiev vehicle which will be driven manually, which has been retrofitted with perception sensor to gather data of surrounding traffic when driving on Singapore's urban roads. The summary of the three steps in the technical approach in developing the methodology to evaluate driving risk is summarized in Table 9 below.

Table 9 – Overview of three configuration sources for data inputs into the driving risk evaluator methodology

Step of development	Driving vehicle	Automated driving system (ADS)?	Environment	Perception
MathWorks – Driving scenario designer	Simulated	No	Simulation: Simple Road configuration	Ground Truth
AV simulation – (Apollo- SVL)	Simulated	Yes – Apollo	Simulation: Complex environment (e.g., San Francisco roads)	Ground Truth
In-vehicle demonstrator	Mitsubishi iMiev	No – Human Driver	Real on-road environment	Sensor hardware

The vehicle will be driven on urban roads to collect data of its interactions with other traffic agents. This provides a practical demonstration of the driving risk methodology proposed. The vehicle will undergo sensor calibration on the CETRAN test track before piloting on urban roads within the Jurong-West area in Figure 20, which has exposure to multitude of traffic interactions.

Figure 20 – Driving area for data collection

6.2.2.1 Set-up of technical approach

The methodology of driving risk evaluation and metrics choice need to be evaluated with data. Data from three configurations are planned to be used, with increasing level of realism. This is further described below. MathWorks' *driving scenario designer* will be the first configuration, followed by AV simulation and finally from real on-road data.

6.2.2.1.1 MathWorks' *driving scenario designer*

MathWorks' *driving scenario designer* [10] allows for design of synthetic driving scenarios, where a road and actor models can be used with a user-friendly drag and drop interface. Setting up of scenarios can be configured for use of the European New Car Assessment Programme (EuroNCAP) or other prebuilt scenarios can be used. The set-up allows for easy adjustments of ego vehicle traits like speed and positions, as well as actor types and their speeds and trajectory. The data from the simulated scenarios are velocities and position of the ego vehicle and actors. As seen in Table 9, this configuration does not include ADS of the ego vehicle drive, the movement of the ego vehicle can be controlled through waypoint definitions. The environment can be simple road networks with simple roads with two lanes such as the cross junction shown in Figure 21. The vehicles are simple box dimensions without detail such as windscreen, lights, or wheels. Perception data can be chosen, but the positions and velocities of the actors can be obtained from ground truth, not requiring data from simulated perception sensors. The data retrieved from MathWorks' *driving scenario designer* of the ego vehicle and actors are perfect and at a low level of reality – free from noise and errors. This is a good initial configuration to test the methodology and tweak the algorithm with basic scenarios without having to debug the source of discrepancies and error in the data collected.

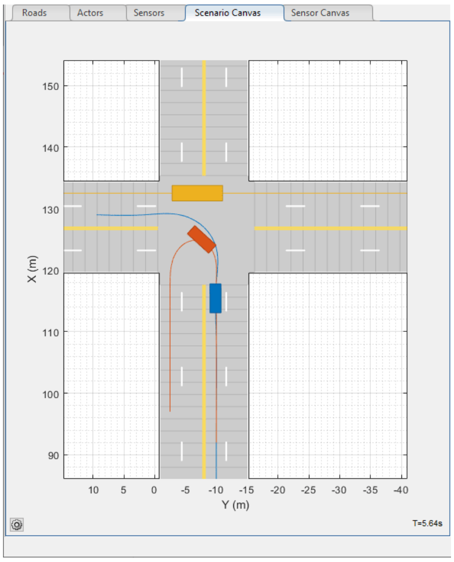
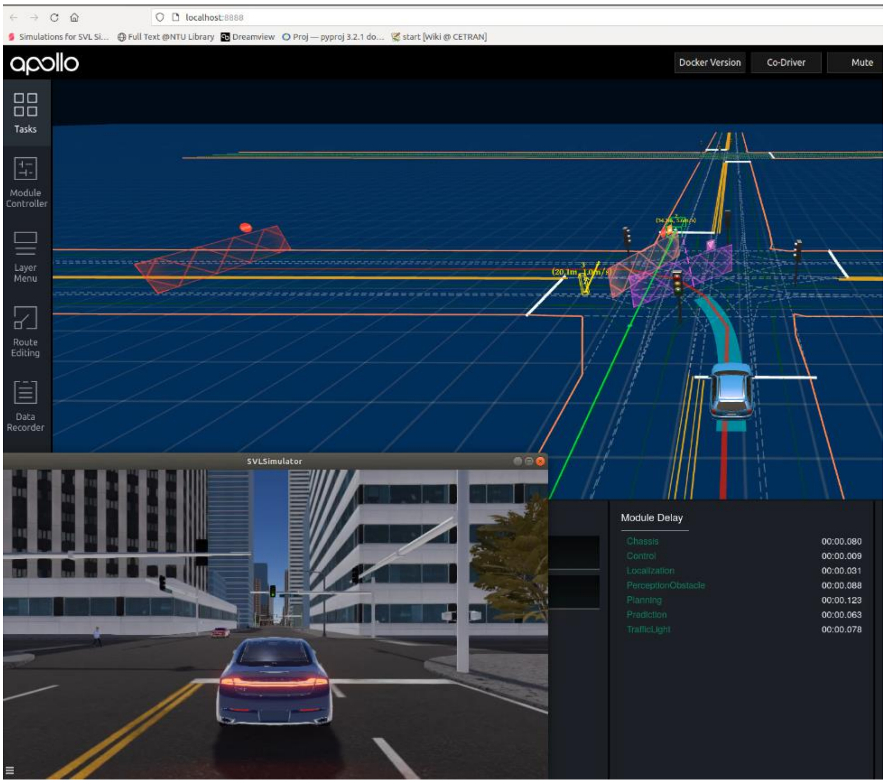


Figure 21 – Example of set-up of scenario in MathWorks' *driving scenario designer* [10]

6.2.2.1.2 Simulation using Apollo-SVL

The next level of realism in data would be obtained from simulation runs. Two types of simulation would need to run concurrently, one that simulates the vehicle's drive – automated driving system (ADS) which would control how the vehicle reacts to scenarios, and another to simulate the environment such as road infrastructure and other road users. CETRAN has prior experience running co-simulations between Apollo and SVL and hence use this combination to carry out simulation runs as the second configuration. The choice of ADS simulator is Apollo, an open-sourced simulation that provides the ability to drive virtually over millions of kilometres to assess, validate and optimize models [11]. The second concurrent simulation is of the environment – where the actors are placed on road maps and networks to move in as in the real world. This is provided by SVL from LG, which was an open-sourced environment that has been sunset since June 2022 [12]. Figure 22 below illustrates the increased level of realism of the environment with road networks with buildings, traffic lights and lane markings, and actor and vehicle features. Apollo ADS is also simulated in the loop, hence the ego vehicle will react if there is an actor in its pathway – differing from the first configuration in MathWorks' *driving scenario designer*. Data from this configuration would be of ground-truth data and provide increased reality on road traffic environments, with ADS in the loop reacting to scenarios created.



Module Delay

Chassis	00:00.080
Control	00:00.009
Localization	00:00.031
PerceptionObstacle	00:00.088
Planning	00:00.123
Prediction	00:00.063
TrafficLight	00:00.078

Figure 22 – Example of Apollo-SVL simulation, providing simulated environments and ADS in the loop driving in configured scenarios

6.2.2.1.3 Vehicle-based demonstrator

The next level of fidelity would be data from real traffic interactions input into the risk evaluator methodology. A vehicle-based demonstrator is developed to collect perception data of the surrounding traffic. CETRAN is utilizing a vehicle equipped with sensors to collect perception data of the surrounding traffic to evaluate the risk of the interactions with traffic during a drive. This vehicle-based demonstrator would utilize a manually driven vehicle with an AV stack driving in "shadow mode", with the driving risk evaluator relying on data output from the perception system. The vehicle-based demonstrator is established as part of the UMGC (Urban Mobility Grand Challenge) project work package 1.3 Demonstrator 2 and does not require any data from path planning and vehicle control systems of the AV stack. Operating the vehicle manually and the AV stack in shadow mode has the following benefits:

- There are no limitations where the vehicle can be driven for the purpose of testing the risk monitor and therefore roads are available for testing where normal AV trial operations are not allowed.
- It eliminates the AV operation risk and makes it easier to perform demonstration rides. In most AV trials in Singapore, there will be a safety driver and safety engineer communication in with each other, preventing discussion by other parties while the trial is ongoing.

Figure 23 – Mitsubishi iMiev vehicle-demonstrator equipped with sensors for data collection

The vehicle-based demonstrator is set up by CETRAN on a Mitsubishi iMiev as shown in Figure 23. Sensors are mounted on the vehicle to provide raw sensor data to the perception module to the ADS stack. The components and corresponding details are listed in Table 10, this includes global positioning system (GPS) and light detection and ranging (LiDAR) used to provide localization coordinates and detection of surrounding objects respectively. The choice of a Velodyne LiDAR is a primary sensor as it is the sensor choice by Apollo ADS software [13].

Table 10 – Details of vehicle-based demonstrator

Components	Model	Component use	Details
Vehicle	Mitsubishi iMiev	To be driven on public roads	Sensor components to be mounted in suitable locations in vehicle
Global positioning system (GPS)	Novatel PwrPak 7D-E1	GPS + onboard inertial measurement unit (IMU)	Localization and speed of vehicle.
	Novatel Global navigation satellite system (GNSS) 502-Antenna	Antenna receiver	Receive GPS signals as inputs for localization data
GPS corrections	SiReNT	GPS real-time kinematic (RTK)	Service by SLA to increase GPS accuracy
Light detection and ranging (LiDAR)	Velodyne Alpha Prime (VLS -128)	Object detection	Range of the detected object by reflection of laser beam
Computer	IPC	Processing Unit	Instrument to run Apollo software's perception module to process raw sensor data

A schematic of the vehicle-based demonstrator is shown in Figure 24, visualizing the data flow from the sensors to Apollo software to the processed output. The PwrPak 7D-E1 outputs raw global coordinates, together with a real-time kinematic (RTK) service called SiReNT by the Singapore Land Authority (SLA) to correct common errors received on the global navigation satellite system (GNSS) module. The output from this sensor is GNSS localization that would provide the vehicle's information of heading, velocity, and position. The Velodyne LiDAR provides 3-dimensional (3D) point cloud, which is then combined with GPS coordinates and time from the GNSS sensor to provide localization and time for synchronization. This results in a 3D point cloud with localization that are inputs into the Apollo software, specifically into the perception module. Apollo version 6 is used for this vehicle demonstrator. Apollo then processes these raw data to generate an object list of the surrounding actors in the vicinity which are commonly seen vehicles and pedestrians in real traffic situations. The object list information is the vehicle's (ego) speed, position and the detected actor's object type, velocity and position. This object list information is used for the driving risk evaluator using the methodology described above to determine the risk of the drive as described in clause 6.2.3.3.

A camera will be installed to record the drive to provide a visual companion to the data collected from the onboard LiDAR, allowing replay of the drive situation to correlate the perception of object list from the sensors to the situation. It will be in a form of a Go-Pro camera installed in the car forward facing on the drive.

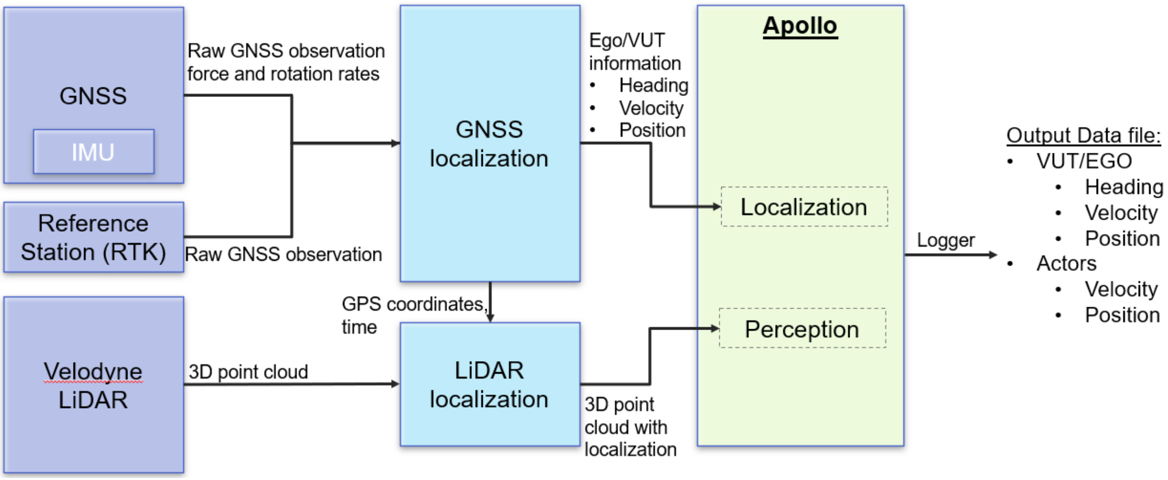


Figure 24 – System architecture of sensors onboard the vehicle demonstrator for object list information

6.2.3 Methodology of driving risk evaluation

As mentioned above, the challenge of developing this methodology is to evaluate the risk of dynamic variation of traffic interactions continuously. Part of the challenge is how to combine a few concurrent individual traffic interactions into one general evaluation. This evaluation should be carried out continuously to capture the changing traffic interactions throughout the drive.

Current metrics can only be applied to specific situations for one interaction. The current state of research shows the lack of metrics that are generic enough to cover all dynamic traffic interactions.

An additional challenge is seen when defining the thresholds of these metrics, as there is subjectivity to the thresholds to distinguish between dangerous and safe interaction risk levels. Part of subjectivity stems from different authorities, countries, and types of road infrastructure. Even within a country, there are varied sources with different recommendations for safe threshold levels based on different considerations of the same traffic situation.

6.2.3.1 Methodology considerations

The current method of evaluating behavioural safety for autonomous vehicles in the Singapore Milestone test framework uses simple clearance distance and static distances as metrics with tolerances developed through trial and error as mentioned earlier. In addition to determine other objective metrics used for evaluation, a better methodology to determine the risk of a drive needs to be developed. This refined methodology can then be adopted into a first pass of the driving risk evaluator that could evaluate a drive continuously. This would improve upon the current status of evaluation based on simple stopping distances. The risk evaluation methodology would require the following considerations to be included:

- Risk as a way of evaluating the drive, with varied levels of safety.

- Continuous identification of interaction scenarios with the subject vehicle (hereafter described as ego vehicle or vehicle under test (VUT).
- Metrics that would be relevant for each interaction scenario, with thresholds assigned to risk levels of each interaction.
- Continuity of risk levels between various interaction scenarios.
- Actor information required includes velocity and position.
- Actor types differentiating pedestrian, motorcycle/cyclist, and vehicles.
- Assessment of risk would be a black-box outcome based on the subject vehicle.

The above points are focused on the interaction of the ego with one other vehicle or pedestrian (also termed actor from here on). The driving risk evaluator would need to further encompass the possibility of evaluating risks with multiple actors at a time. This provides an increased level of complexity to the conceptual risk evaluation method in combining all interaction for an overall risk score.

The following considerations are not considered in this initial methodology as the focus is on the continuous evaluation of the on-road interactions aspect.

- Road infrastructure is not considered.
- Road traffic rules are not considered.
- ADS is treated as a black box. Only information on position and velocity is required.

6.2.3.2 Methodology review

The challenge for continuous evaluation of an autonomous vehicle with multiple actors in the immediate vicinity of the ego vehicle would raise questions such as:

1) Would the actor type determine priority and risk levels?
2) Would a different interaction type have a higher risk?
3) How would an increase in the number of actors affect the total risk evaluated of the drive?
4) Would proximity to the ego lead to a higher priority or risk assigned for the interaction?
5) Would there be granularity of risk levels assigned to differentiate two different drives that are determined safe?

6.2.3.2.1 Industry risk methods

To address the above concerns, current industry standard risk assessment methods were further analysed for relevance. One such industry standard is the hazard analysis and risk assessment (HARA) method determining automotive safety integrity levels (ASIL) levels to establish a risk classification system according to ISO 26262 for functional safety [6], [7]. This method for assessing the risk of malfunction of electronic components of a road vehicle that looks at three components; severity, exposure and controllability uses the risk matrix as a combination to define the risk levels [6], [7]. This method to identify and mitigate risks for determinist systems is not applicable for AI based systems but the concepts of severity, exposure or controllability provide inspiration for the methodology developed here.

Severity	Exposure	Controllability		
		C1 (Simple)	C2 (Normal)	C3 (Difficult, Uncontrollable)
S1 LIGHT AND MODERATE INJURIES	**E1** (Very low)	QM	QM	QM
	E2 (Low)	QM	QM	QM
	E3 (Medium)	QM	QM	A
	E4 (High)	QM	A	B
S2 SEVERE AND LIFE THREATENING INJURIES – SURVIVAL PROBABLE	**E1** (Very low)	QM	QM	QM
	E2 (Low)	QM	QM	A
	E3 (Medium)	QM	A	B
	E4 (High)	A	B	C
S3 LIFE THREATENING INJURIES, FATAL INJURIES	**E1** (Very low)	QM	QM	A
	E2 (Low)	QM	A	B
	E3 (Medium)	A	B	C
	E4 (High)	B	C	D

QM (Quality Management)
Development supported
by established Quality Management
is sufficient.

A **lowest ASIL**
Low risk reduction necessary
B
⋮
C
D **highest ASIL**
High risk reduction necessary

Figure 25 – ASIL determination [6], [7]

The AVSC [2] also looks at safety performance metrics in various categories as shown in Figure 25 as a way to evaluate the risk. The first category to "maintain a safety envelope" would be the method of calculating risk due to interaction with actors and vehicles in the vicinity of the ego. A separate calculation to evaluate the ego's behaviour, such as rapid acceleration or speed, would fall under the second category to "exhibit contextually safe vehicle motion control". This could be assessed independently should there be no surrounding actors, which would reflect on the behaviour of the vehicle despite no risk of interactions.

Studies were also conducted into the severity based on fatality probability due to a collision. It is natural to consider as fundamental the notion of injury risk when it comes to assessing safety performance. This notion is however an entire research field very much dependent on the vehicle under test, the test conditions, the involved actors and an entire domain of physical considerations and impact biomechanics [8].

6.2.3.2.2 Initial methods explored

Some initial concepts were brainstormed for developing a continuous evaluation of multiple interactions on-road. The following sections will walk through ideas that were considered before concluding with the final method adopted.

First method:

An initial method explored the possibility to give priority to pedestrians and riders on personal mobility devices (PMDs), followed by vehicles then objects. This is to support the view that pedestrian and PMD users are vulnerable compared to other road users, resulting in a higher risk value for such interactions. After assigning the risk level based on the metrics chosen, for example an interaction with a pedestrian or PMD user will have an increase in risk level by some percentage, while for a vehicle it is maintained, a decrease of another percentage is applied to an object. The total risk of interactions with three actors would then be an average of the three risk levels assigned to each individual actor.

Increasing the final risk was considered with the number of actors around the ego vehicle. This would reflect the increased attention required of the ego vehicle while driving on the road when there are greater number of vehicles and pedestrians around. These two points cover only points 1 and 3 of the methodology considerations, and the weightage on actor type is subjective.

Second method:

An adaptation of the first method would be to assign units per actor class type to give priority to vulnerable road users. This would be with values such as:

Pedestrians:	4 units
Vehicle:	3 units
PMD:	2 units
Object:	1 unit

This was tested out by averaging 3 pedestrians with a risk level of 3 (out of 4, with 4 being the highest risk), and 1 object with risk level of 4. The combined risk of such an interaction would be

(Units per actor type * No. Actor * Risk Level)/Total Units = (4x3x3 + 1x4)/ (4x3+1) = 3.07

This average method of different risks would dilute a high-risk level of 4 that was assigned to the interaction with an object. The concept of prioritising vulnerable road users should be retained although another way of achieving this meaningfully could be adopted. This was further explored and influenced the eventual methodology.

6.2.3.3 Methodology adopted for first pass of driving risk evaluator monitoring

In the methodology adopted for a first pass of driving risk evaluator, all five elements of the methodology review are considered. Certain actor types such as pedestrian and PMD users would have priority with granularity of risk levels applicable. The driving risk evaluator methodology has five steps in evaluating the risk of a drive. The method is applied at each time-step of the drive, such that there is a combined risk of interactions at each time instant. An overall score would be available for easier assessment of the drive, however how that score and data is used to evaluate the drive would need to be developed based on its adoption use case.

The steps of the methodology of calculating the risk at each time instant are visualized in Figure 26:

1) Identify interaction with each actor in the vicinity.
2) Assign risk back for each actor based on the metrics that are relevant to the interaction.
3) Severity 1 is assigned based on actor type, and the interaction situation that embodies various kinetic energy if a collision were to occur.
4) Based on the ego speed and number of actors, a risk adjustment is done based on all interactions within that time instant, this step applies a second level of severity.
5) Integrating all the interactions and risk for one risk score at that time instant.

Of the five steps, the first three are on individual interactions of one actor with the ego/subject vehicle. Then a score of each individual actor's risk is grouped to collectively assess all interactions with the ego/subject vehicle at that same time instant through steps four and five.

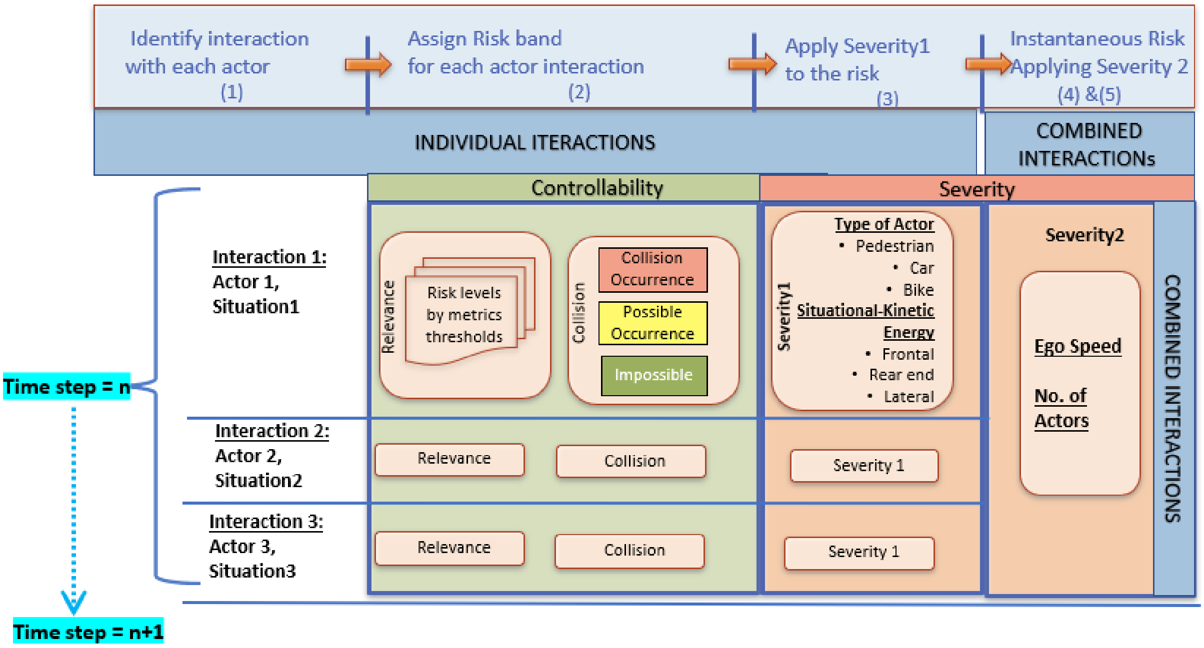


Figure 26 – Schematic of the steps of the driving risk evaluator methodology

Figure 27 shows in detail the methodology of calculating the risk for a specific time instance.

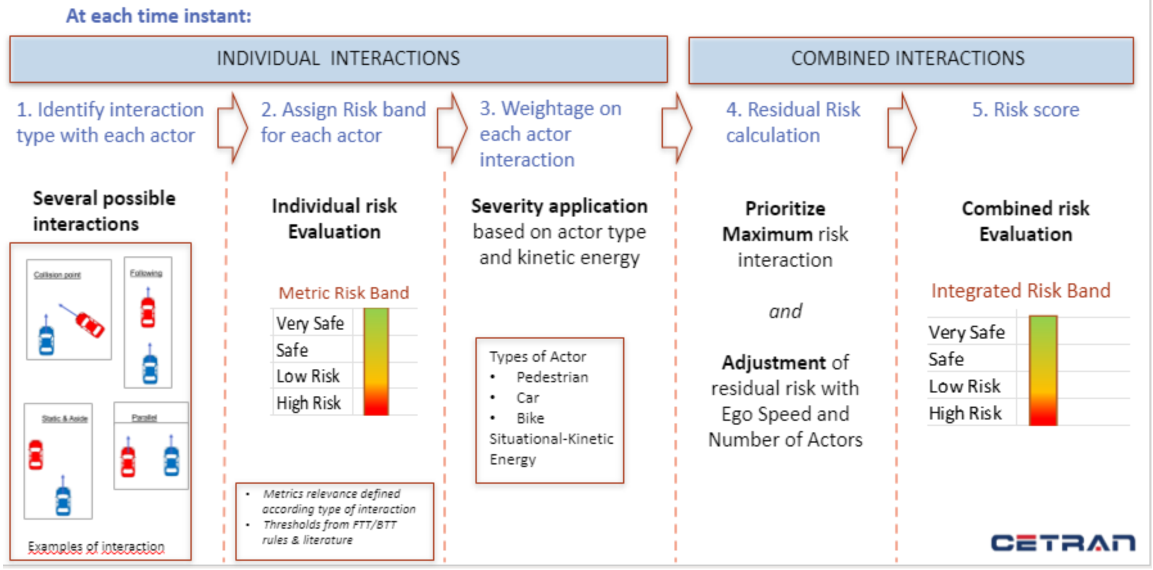


Figure 27 – Description of steps of the driving risk evaluator methodology for a specific time instance

6.2.3.3.1 Step 1: Identify interaction with each actor

In the first step of the driving risk evaluator methodology, an interaction situation type would need to be identified. Actors would only be classified if they are within proximity to be considered. This is generally done in four classes (see Figure 28):

- Collision point: actor and ego are heading into a collision.
- Following: Ego is following behind an actor.
- Parallel: Ego is moving alongside to an actor.
- Static and aside.

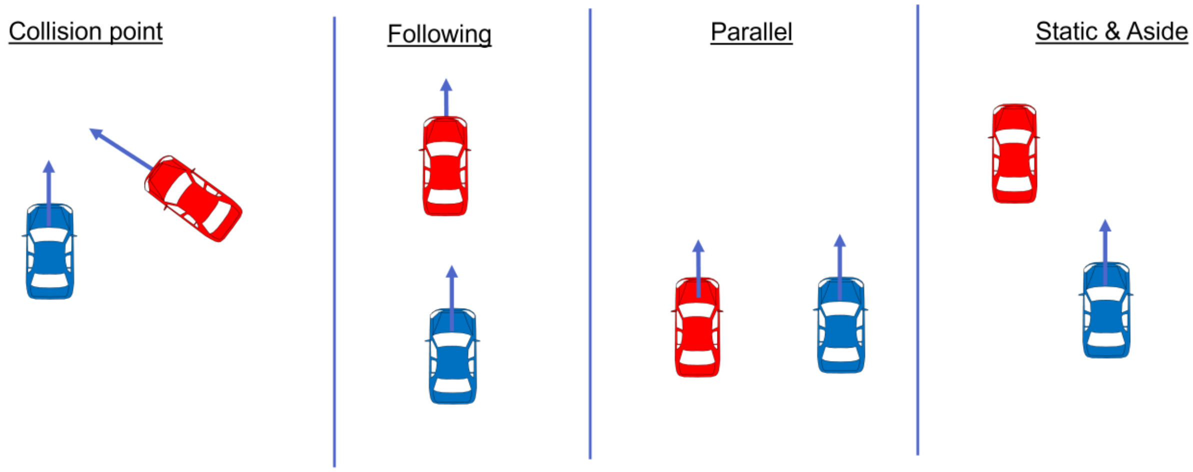


Figure 28 – Interaction situation type with ego/subject vehicle

6.2.3.3.2 Step2: Assign risk level based on relevant metrics to interaction type

When the interaction situation type is assigned for an actor at a given timestep, a set of relevant metrics is then applied for the situation. There are four risk bands or levels for every metric, to reflect safety levels: namely 1. Very Safe; 2. Safe; 3. Low Risk; 4. High Risk. These levels are referred from the literature. The four levels provide more granularity as compared to a safe or not safe criteria. The governing metric for each situation type is listed in Table 11, the criteria for choosing these metrics will be further discussed in clause 6.2.3.4.

Table 11 – Table of governing metrics for each interaction situation type

Interaction situation type	Governing metric
Collision	Time-to collision
Following	Longitudinal clearance in relation to speed
Parallel	Lateral clearance
Static and aside	Lateral clearance

6.2.3.3.3 Step 3: Severity application from actor type and situation kinetic energy

The third step of the driving risk evaluator methodology is to account for the actor type and situational kinetic energy in the interaction. It is a way to acknowledge that pedestrian and cyclists have greater vulnerability if a collision were to occur. Another aspect would be to differentiate the various kinetic energies if a collision were to occur. Studies on differing impacts show that a side-impact results in greater severity compared to a head-on collision as observed in a study in Adelaide [8] by the Australian Road Research Board and Centre for Automotive Safety Research in University of Adelaide that referenced Wramborg's model shown in Figure 29. Figure 29 illustrates three different lines of the probability of fatality in relation to collision speed. It is seen that at a given speed, collision with a pedestrian or a cyclist has the highest severity, followed by side-impact and lastly a lowest probability with a head-on collision. The type of situational interactions with pedestrians is not listed – highlighting that such actor type's severity of collision is independent of the interaction situation. The energy of the collision is reflected in the collision speed, indicating increasing fatality probability with speed. With this third step in this driving risk evaluator methodology, the individual risk of each actor interaction at each time-step is finalized.

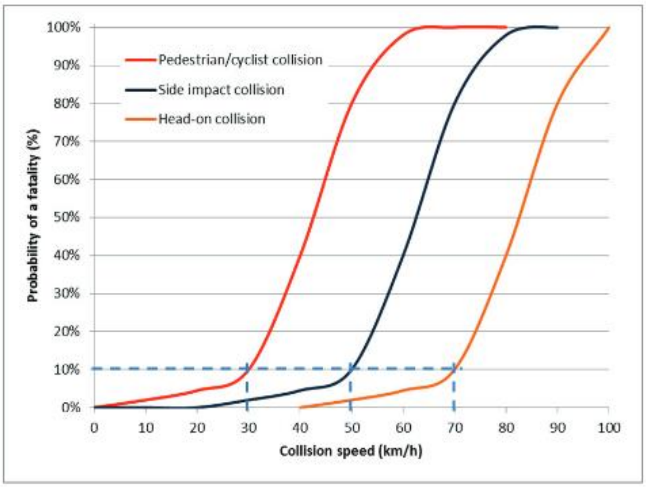


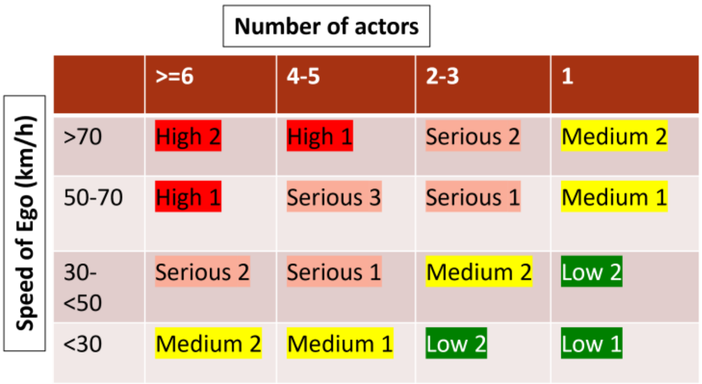
Fig. 1. Wramborg's model for fatality probability vs. vehicle collision speeds. Source: based on Wramborg (2005).

Figure 29 – Wramborg's model that depicts the probability of a fatality due to collision speed, differentiating between actor class and direction of collision [8]

6.2.3.3.4 Step 4: Residual risk calculation based on number of actors and ego speed

In the fourth step of the driving risk evaluator methodology, a combination of the risk of multiple interactions is done. The initial methods explored an averaging method described in clause 6.2.3.2.2 however this final first-pass methodology retains the highest risk of all interactions and seeks to retain the risk of the remaining interactions risk through a risk adjustment method on ego speed and the number of actors. It draws on a common risk assessment matrix used in HARA, and we have determined 9 risk levels as can be seen in Table 12. Two assessment criteria considered are the number of actors and the speed of the ego vehicle to indicate different aspects of severity. Based on the risk levels from Table 12, a weightage is referenced in Table 13 to be applied on the sum of the remaining interaction risks. It should also be emphasised that the highest risk among all the interactions at each time instant has been retained, and this method of risk adjustment applies to the sum of remaining risks (which are lower than the highest risk). The levels in Table 12 are an initial recommendation to apply on the sum of remaining risks.

Table 12 – Zones of risk based on number of actors in the vicinity and speed of the ego vehicle



Number of actors

Speed of Ego (km/h)	>=6	4-5	2-3	1
>70	High 2	High 1	Serious 2	Medium 2
50-70	High 1	Serious 3	Serious 1	Medium 1
30-<50	Serious 2	Serious 1	Medium 2	Low 2
<30	Medium 2	Medium 1	Low 2	Low 1

Table 13 – Application of weightage percentage on sum of remaining risks based on risk zones from above

Risk Zones	Application of %weightage to sum of remaining risks
Low 1	**0**
Low 2	**2**
Medium 1	**4**
Medium 2	**6**
Serious 1	**8**
Serious 2	**10**
Serious 3	**12**
High 1	**14**
High 2	**16**

6.2.3.3.5 Step 5: Integrating all interactions for one risk score at each time instant

The final step of the driving risk evaluator methodology would to be incorporate all individual interaction risks with the ego vehicle into one risk value at that time instance. The earlier step four is a method of retaining the influence of other actor's risk that did not result in the maximum risk. This final score adds the maximum risk with the weightage applied on the remaining interaction risk score.

$$\text{Total risk}_{t=i} = max(risk_a)_{t=i} + [Risk\ adjustment\ weightage\ \% \times \left(\left(\sum_{a=1}^{num\ of\ actors} risk_a \right) - max(risk_a) \right)]_{t=i}$$

6.2.3.4 Final metrics chosen

The methodology adopted for the first pass of a driving risk evaluator thus requires that each interaction situation have the same four levels of risk bands assigned at each time instant. The governing metric for each situation should be assigned thresholds that pose the same level of risk comparable to a different interaction situation. There is also the need to have continuity between the governing metrics when an interaction situation switches. To resolve this, one method would be to translate these metrics and thresholds to a common entity such as clearance.

Some of the thresholds for metrics from Singapore's 'Traffic Police guidelines of Final Theory and Basic Theory of driving' provide only one safe level of thresholds. This provides only a clear pass or fail criteria and does not allow for more intricate levels of risk assignment. An example would be the following distance to a vehicle ahead, the recommendation is the 'two second rule'. However, this would only result in one demarcation between safe or unsafe.

For other metrics, there are multiple thresholds based on details of each unique scenario, with the challenge of choosing the thresholds. There are a wide variety of metrics as researched in the literature review. A down-selection of the relevant metrics would need to be done to suit the methodology developed for the driving risk evaluator.

6.2.3.4.1 Metric for static and aside interaction

For the interaction type of static and aside it is simple to determine that the governing metric is lateral clearance. It is stated in the FTD (226F) [5] that at least 0.5 metre clearance should be kept for static obstacles, and for parked vehicles a maximum of 1 metre. There were intricacies to a clearance threshold from the FTD where with different facing direction of the pedestrian relative to the ego vehicle, different thresholds could apply FTD (242F and 243F) [5]. Furthermore, in Singapore's TR68-1, there are also different recommended thresholds based on the ego speed. If above 30 km/h, a clearance of 1.5 metres, below 30 km/h would be 1 metre, and finally if an actor is not on the road surface, 0.5 metre would be the recommended threshold. With these various thresholds and conditions, the following thresholds, listed in Table 14, were determined for the four risk levels, incorporating the element of speed into severity 1 found in step 3 of the methodology. These levels are a first pass of thresholds chosen that should be validated in scientific studies and improved in future projects.

Table 14 – Thresholds for lateral clearance metric for interaction of static and aside[5]

Risk levels	Lateral clearance
Very safe	> 1.5 metres
Safe	1-1.5 metres
Low risk	0.5-1 metres

6.2.3.4.2 Metric for parallel interaction

In a parallel interaction situation, the governing metric will also be lateral clearance similar to the earlier static and aside interaction. However due to actor movement, the thresholds would be higher than if interaction was with a static actor. Singapore's FTD has different thresholds based on the vehicle's moving speed; with 1.5 metres as the recommended distance when a vehicle is travelling above 30 km/h [5]. This threshold is used for parallel interaction, above which it is in a safe risk level. The threshold of 1 metre if the vehicle is below 30 km/h is used for the low-risk band level. From TR68-1 [3], a threshold of 1.5 metres is simply stated. An additional threshold of 2 metres was added by CETRAN to indicate an increased safety level, with thresholds shown in Table 15. This is an example where four levels of risk bands would enable incorporation of the various thresholds recommended for the Singapore road traffic interaction. The element of ego vehicle speed is later considered in severity based on actor type in step 3 and in a smaller percentage in step 4 when considering all actors' interaction in a time instant. These levels are a first pass of thresholds chosen that should be validated in scientific studies and improved in future projects.

Table 15 – Thresholds for lateral clearance metric used for the interaction of parallel[6]

Risk levels	Lateral clearance
Very safe	> 2 metres
Safe	1.5-2.0 metres
Low risk	1.0-1.5 metres

6.2.3.4.3 Metric for following interaction

For the interaction of following an actor, the governing metric that is utilized for the driving risk evaluation methodology is the longitudinal clearance in relation to speed. From the literature review in clause 6.2.6, there are multiple ways to assess this. From the Singapore traffic road rules, there is a recommended '2 seconds rule', where the longitudinal clearance is dependent on the ego vehicle's speed [4] to allow for braking distance if the vehicle ahead should come to a stop abruptly. Additionally, there is also a recommendation for one car length clearance for every 16 km/h of the ego vehicle. In the CETRAN document "*Assessment Guidelines towards Independent Safety Assurance of Autonomous Vehicles*" (which is under LTA review), there is a recommended 2 metre clearance. In these threshold recommendations, the consideration is only on the ego vehicle's velocity,

---

[5] Simple values chosen from FTD and TR68-1, the choice of threshold values need to be validated by scientific studies in future projects.

[6] Simple values shown for illustration of risk levels, threshold values should be validated in scientific studies in future projects.

other elements of ego vehicle's acceleration, actor velocity and acceleration have not yet been considered.

Time-to-collision (TTC) is commonly used as a metric for a following type interaction. It typically uses the distance between the ego and a traffic simulated vehicle (TSV) and the difference in speed to approximate the time of colliding with the vehicle ahead. In a study by Rutgers Intelligent Transportation System (RITS) Laboratory [9], a proposed minimum time to collision (MTTC) uses a difference in velocities of the ego and actor, difference in acceleration and the current distance using the non-negative solution. This method was proposed as an improvement to conventional TTC calculations, which ignores potential conflicts that would arise from acceleration discrepancies.

In this 'following interaction' category, the metric will utilize the MTTC method when applicable, or else revert to the conventional clearance with respect to velocity such as the '2 seconds rule', and one car length every 16 km/h of velocity. By choosing an average car length of 4.2 metres, it was calculated at 2 seconds the threshold has a larger clearance than one car length for every 16 km/h. The '2 seconds rule' is then used for the very safe risk level. A choice of one car length every 24 km/h was chosen to indicate a high-risk level, through comparisons of clearance across various velocities. A summary of these threshold is shown in Table 16. These levels are a first pass of thresholds chosen that should be validated in scientific studies and improved in future projects. The thresholds for MTTC clearance chosen are derived from literature review recommendations for TTC thresholds. The thresholds for the respective risk level using MTTC are shown in Table 17.

Table 16 – Thresholds of clearance with respect to velocity of the ego vehicle for an interaction of following[7]

Risk levels	Clearance with respect to velocity
Very safe	> 2 seconds
Safe	2 seconds – 1 car length every 16 km/h
Low risk	1 car length every 16 km/h – 24 km/h
High risk	1 car length every 24 km/h

Table 17 – Thresholds of clearance utilizing alternate minimum time to collision (MTTC) method for an interaction of following[7]

Risk levels	Clearance threshold using MTTC (if applicable)
Very safe	MTTC > 5.5 seconds
Safe	MTTC 3-5.5 seconds
Low risk	MTTC 2-3 seconds
High risk	MTTC < 2 seconds

This metric is a first derivation of the risk levels for this 'following interaction' type. Further improvement and consideration would need to be done in future projects to refine these thresholds and the method.

---

[7] Threshold values shown for illustration of risk levels, these should be validated in scientific studies in future projects.

6.2.3.4.4 Metric for collision interaction

In the interaction type of collision point, the metric chosen is based on the well-known time-to-collision (TTC) method. It is to calculate the convergence of the trajectory of both ego and actor which is indicative of a potential collision point. The trajectory of the ego and actor is used to calculate a collision point, and the time for each the ego and the actor is calculated based on the current speed and acceleration. The delta in the time to reach the collision point is calculated and banded into the four risk levels.

This interaction would be the only one where there is a two-step approach in assigning a risk level. The initial risk level is assigned based on the delta in time to collision point. Various sources have differing recommendations for what is a good delta, some with a recommendation of 1.5 seconds, others with 2 seconds [1]. Hence CETRAN has decided to order these differing recommendations into the four risk band levels. These levels are a first pass of thresholds chosen that should be validated in scientific studies and improved in future projects.

Table 18 – Thresholds of risk levels when considering delta time to collision point for an interaction type of collision[8]

Risk levels	Delta time to collision point
Very safe	> 3 seconds (3.5)
Safe	2-3 seconds (3-3.5)
Low risk	1.5-2 seconds (2.5-3)
High risk	< 1.5 seconds (2.5)

The second step in assigning the risk band is to determine the proximity of the ego vehicle to the collision point. The delta time to collision point could be as small as 1 second, but the time for the ego vehicle to reach the point could be 6 seconds, with the actor at 7 seconds away. The initial risk level assigned could be high-risk, but the risk level could be reduced to very safe if the ego vehicle is far away from the collision point and has ample time to react.

6.2.4 Methodology implementation and experimental results

With the four configurations detailed in clause 6.2.2, the following section will show the methodology application to data from each of the respective configurations. With various sources of data, it is indicative how the methodology is based on the quality and realism of the data. A sample scenario will be detailed, followed by details of implementing the scenario in MathWorks' *driving scenario designer* and subsequently in simulation. The data collected from the vehicle-based demonstrator under WP1.3 Demonstrator 2 will also be analysed.

6.2.4.1 MathWorks and AV simulation

6.2.4.1.1 Description of sample scenario

In sample scenario A in Figure 30, an ego vehicle turns left at a non-signalised junction, with an oncoming vehicle going straight, and with a pedestrian crossing the target lane. The oncoming vehicle is sometimes referred to actor 1 or a TSV, while a pedestrian is termed a vulnerable road user (VRU) which is also applicable to users of PMDs or cyclists. A further variant B shown in Figure 30 shows two additional vehicles in the scenario.

---

[8] Threshold values shown for illustration of risk levels, these should be validated in scientific studies in future projects.

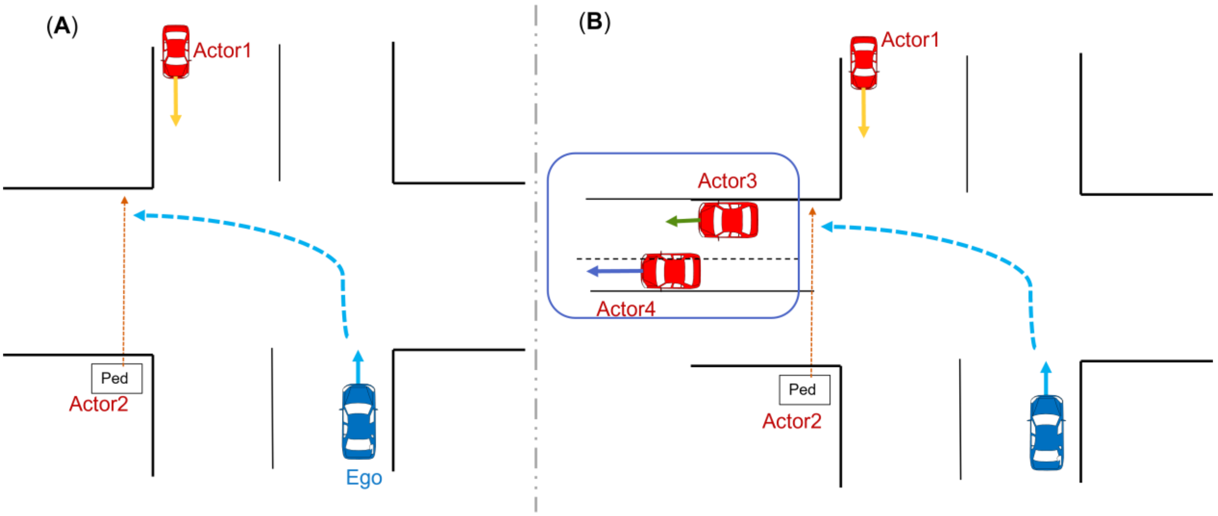


Figure 30 – Left (A): A sample scenario with an Ego vehicle making a left turn, interacting with two different actors
Right (B): two additional vehicles travelling to the left ahead of the ego vehicle after the left turn manoeuvre

6.2.4.1.1.1 Assumptions

In the above scenario, the TSVs and VRU will not yield to the ego vehicle.

6.2.4.1.1.2 Scenario description

Ego vehicle is travelling straight and making a left turn at the unsignalized intersection. An oncoming TSV goes straight, and a VRU travels in parallel to the ego vehicle, crossing the street simultaneously as the oncoming TSV. In a scenario variant (B), two other TSVs are in the lane which the ego vehicle turns into. The ego vehicle will be following actor 3 and in parallel to actor 4.

6.2.4.1.1.3 Implementation details

Parameters:

1) Ego vehicle (A and B)
   a) Vehicle type: Passenger vehicle.
   b) Manoeuvre: Ego vehicle turns left at the unsignalized intersection.
   c) Speed: MathWorks: 30 km/h, Simulation: 30 km/h then decided by ADS.
2) Actor 1(A and B): Oncoming TSV going straight
   a) Actor type: Sedan.
   b) Initial position: MathWorks: 60 metres, Simulation:50 metres away from ego vehicle.
   c) Manoeuvre: Moves forward towards ego.
   d) Speed: MathWorks: 30 km/h, Simulation: 30 km/h.
   e) Trigger: MathWorks: when run is started, Simulation:50 metres away from ego vehicle.
3) Actor 2 (A and B): VRU crossing path
   a) Actor type: Pedestrian.
   b) Initial position: MathWorks: 10 metres, Simulation:3 metres lateral from ego vehicle.
   c) Manoeuvre: Moves forward away from ego.
   d) Speed: MathWorks: 9.7 km/h, Simulation: 3 km/h.
   e) Trigger: MathWorks: when run is started, Simulation:10 metres away from ego vehicle.

4) Actor 3 (B): TSV going left in figure
    a) Actor type: Sedan.
    b) Initial position: MathWorks: 40 metres.
    c) Manoeuvre: Moves forward.
    d) Speed: MathWorks: 30 km/h.
    e) Trigger: MathWorks: 5.5 seconds after run is started.
5) Actor 4 (B): TSV going left in figure
    a) Actor type: Sedan.
    b) Initial position: MathWorks: 37 metres.
    c) Manoeuvre: Moves forward towards ego.
    d) Speed: MathWorks: 30 km/h.
    e) Trigger: MathWorks: 5.5 seconds after run is started.

6.2.4.1.2 MathWorks' *driving scenario designer*

With the scenario set-up B in MathWorks' *driving scenario designer* as detailed above, data of the actors and ego vehicle was obtained and analysed using the methodology detailed in clause 6.2.3.3. Figure 31 is a plot of the total risk of all interactions across the duration of the sample scenario. This is the result after each individual interaction is evaluated based on interaction situation type and its corresponding metrics, with severity applied from actor type and ego vehicle speed. Another way to evaluate the drive would be obtaining time spent in each risk zone as a percentage of the drive as seen in Figure 32. This helps an assessor breakdown the drive and analyse those different risk interactions detected, and the corresponding percentage of time spent in each risk zone during the drive. The bar chart in Figure 32 further highlights that in the sample scenario, the ego vehicle spent a large amount of time in the very safe (56.40%) and safe (14.29%) risk levels.

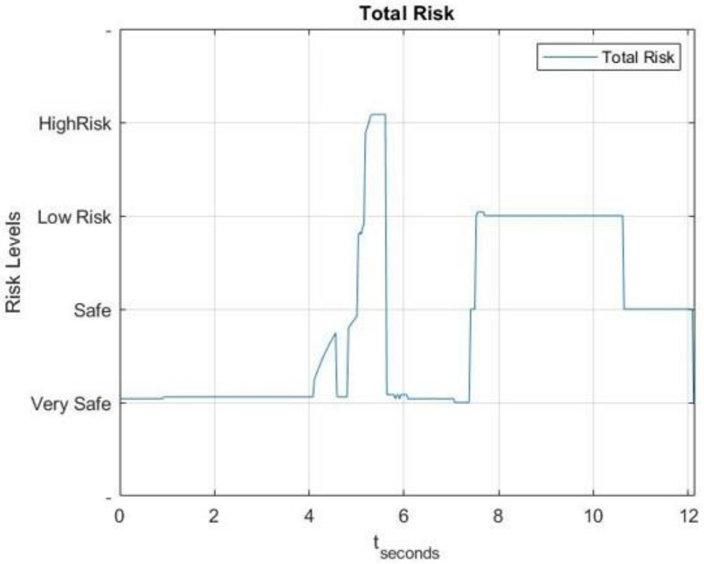


Figure 31 – Plot of total risk of all interactions across duration of the sample scenario B

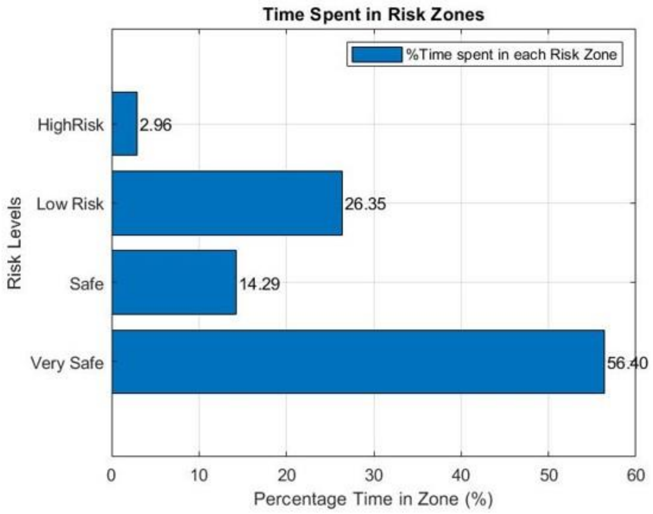


Figure 32 – Analysis of driving risk displayed in percentage of time spent in risk zone

Figure 33 shows a break down in the individual interaction risks with each actor during the sample scenario duration. The culmination of total risk at each time instant applies step four and five of the driving risk evaluator methodology. The plot in Figure 31 of the resultant total risk follows the maximum of individual risks in Figure 33. There is an increased risk with the pedestrian (VRU) as the ego vehicle approaches the left turn without slowing down whilst the VRU is crossing. This interaction is classified as a collision situation, and with the decreasing proximity, the risk increases till the pedestrian has passed the ego vehicle. Such escalation of risk is indicative that the ego vehicle had not made efforts to decrease a potentially dangerous situation. The sudden drop off in risk is also indicative that a near-miss has occurred. These are some ways to interpret the risk plots and possibly dangerous interaction moments in the drive evaluated.

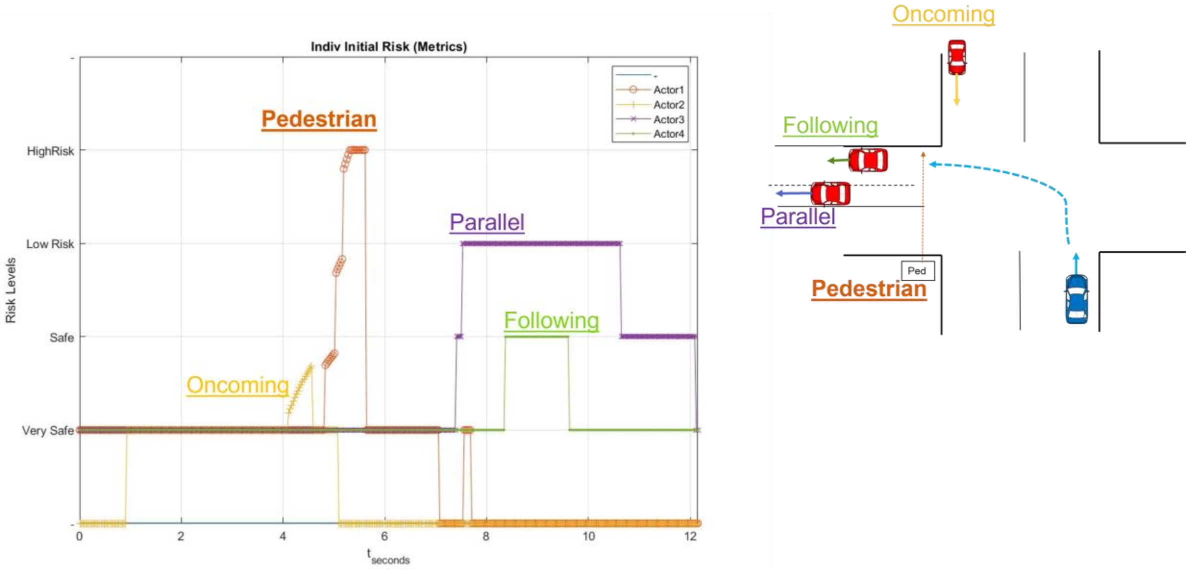


Figure 33 – Breakdown of individual interaction's risk in the sample scenario B

Further statistics listed in Table 19 would also enable a quick recap of the driving scenario, to get a maximum risk incurred during the drive, and the average risk of the drive. In the above scenario, there were moments of high-risk interactions, however the average risk of the drive is in a very safe level with a level of 1.83 to enable granularity if comparing with other drives. The values shown in this table should not be used in isolation when evaluating driving risk, but should be used with the total risk plots and percentage time in risk zone from Figure 31 and Figure 32 to provide a comprehensive evaluation of the drive.

Table 19 – Additional statistics of the driving risk evaluator methodology

Max risk	High risk (4.08)
Average risk	Very safe (1.83)

6.2.4.1.3 AV simulation

The scenario A seen on the left in Figure 30 is run with the combined simulation of ADS and environment in Apollo-SVL. The additional two actors are removed to simplify the scenario and focus on the interaction with the oncoming vehicle (Actor1) and the pedestrian crossing the road (Actor 2) when ADS is in the loop. The data is analysed using the methodology detailed above, resulting in a risk plot with time throughout the drive seen in Figure 34 which combines the risk of all actors at each time instant. This can also be presented as a percentage time spent in each risk zone as shown in Figure 35, where it is seen that 92.34% of the run time of the scenario is spent in a very safe zone with traffic interactions, and 2% in each of the other three risk level bands.

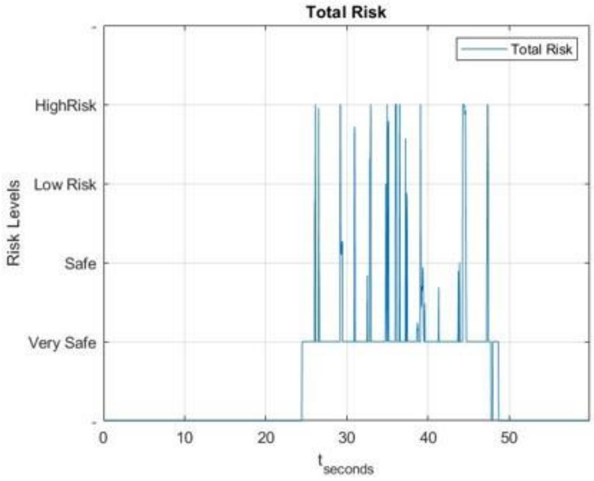


Figure 34 – Plot of total risk versus time of the drive through the scenario

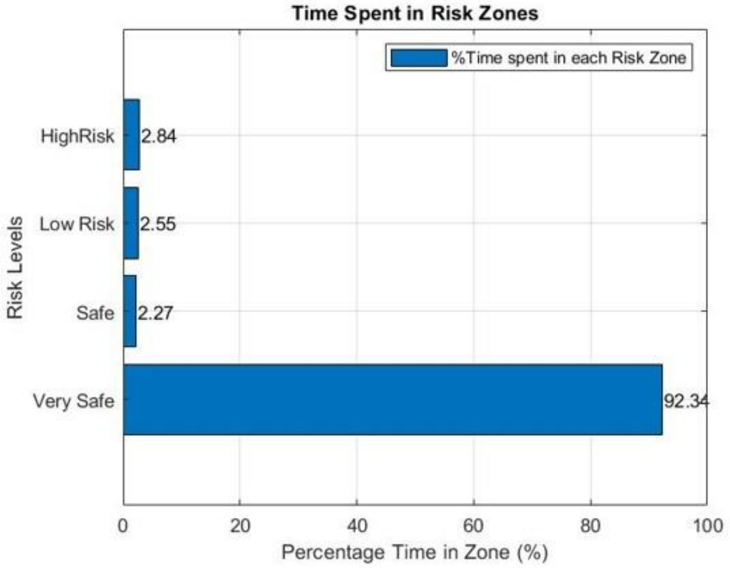


Figure 35 – Plot showing percentage of time spent in risk zone during drive

In Figure 34, there appears to be multiple jumps from very safe to high risk and some low risk levels, this can be further analysed from the risk from each individual actor in Figure 36.

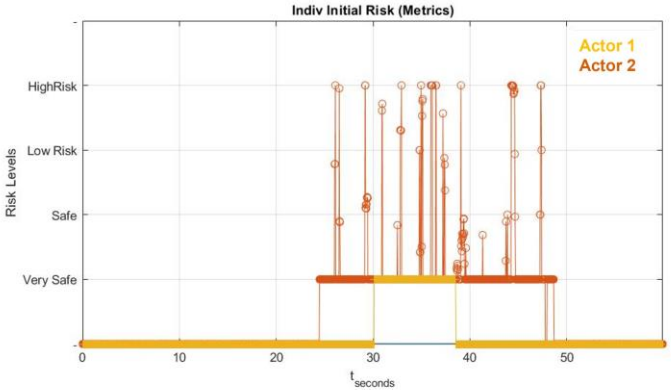


Figure 36 – Breakdown of individual actor's risk during the scenario

The levels of high risk are from the pedestrian (Actor 2 or VRU) interaction in the simulated scenario A. Further analysis of the simulation run shows the pedestrian alternating facing the left and right while crossing the road. The erratic change in movement direction would change the proximity of the collision point. This highlights the risk variation with actor movements, and the methodology picks up on this variation which could be potential events in real traffic interactions. Another difference from the AV simulation run with Apollo-SVL as compared to MathWorks, is that with Apollo ADS in the loop, the same scenario results in a larger time spent in a very safe risk level, due to the simulated ego vehicle slowing down when approaching these actors in the simulated environment. Further statistics are listed in Table 20, with a maximum risk and the average risk of the drive. For the AV simulation run of scenario A, the maximum risk is a high-risk level, and the average risk of the drive is very safe with a value of 1.20. Comparing these values to the MathWorks run would show that both drives incurred a high-risk interaction moment, but both drives also have an average risk in very safe risk level, but the AV simulation run has a lower average risk value, indicating a safer drive. However as mentioned earlier, the values from Table 20 should not be used in isolation, as this would seem to indicate both the AV simulation and MathWorks simulation performed similarly. By comparing the plot of risk with time, and time percentage spent in each risk level, it is clear that the AV simulation appears to be safer for the majority of the run, due to the reduction of time spent in high risk level.

Table 20 – Additional statistics of the driving risk evaluator methodology applied to the AV simulation

Max risk	High risk (4.0)
Average risk	Very safe (1.20)

Errors in data are a common occurrence, and in the data logged from Apollo-SVL AV simulation, discrepancies were noted primarily on the velocities of actors. It is seen that there are velocity spikes and missing data that appear non-periodically in the data output, however these actor vehicles appear to be moving smoothly during the simulation. These spikes and missing data occur for a very few time-steps in the data collected and have an impact on assessing the risk. Large velocities could indicate a shorter time to collision when such interaction scenarios occur, however it could be an error of the simulation detection at those time instances. Such seemingly non-realistic velocity spikes (could be values of 60 km/h while the neighbouring data points are at 25-30 km/h) are passed through a median filter. This is a common noise suppression filter that uses a non-linear method to reject signal samples that differ substantially from neighbouring signals [14] as is the use of such filter. This median filter adopted is only applicable for data coming from this specific simulation set up. This would not be applied to data received from external sources. The jumps in risk could be an indication that data quality could be improved from external sources. It should also be noted that the jumps in risk levels does not simply imply data improvement but could also indicates erratic actor behaviour as observed in this scenario earlier.

This methodology is shown to aid in the analysis of the safety of the interaction of the sample scenarios above, providing a continuous evaluation of the drive. This will also be applied to real traffic data collected by the vehicle demonstrator and will be detailed below.

6.2.4.2 Vehicle-based demonstrator

The vehicle-based demonstrator detailed in clause 6.2.2.1.3 is used to collect real traffic interactions data. The data output from Apollo's perception module is then processed using the methodology described previously. The vehicle was driven on CETRAN's test track and on public roads within the university compound and parts of the Jurong West area in Singapore gathering data from real traffic interactions. These are detailed below.

6.2.4.2.1 Drive and data collection on the CETRAN test track

The vehicle was driven within the CETRAN test track compound which allows for a simpler environment compared to public roads. A route with real traffic interactions within the CETRAN test track compound was driven as shown in Figure 37.

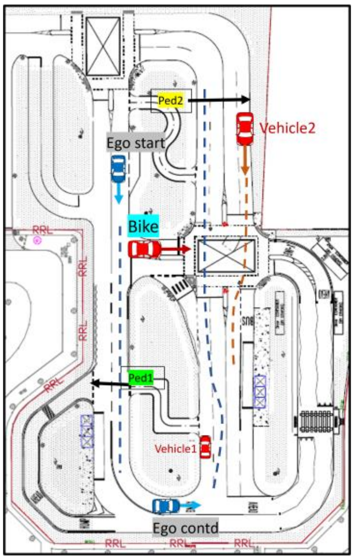
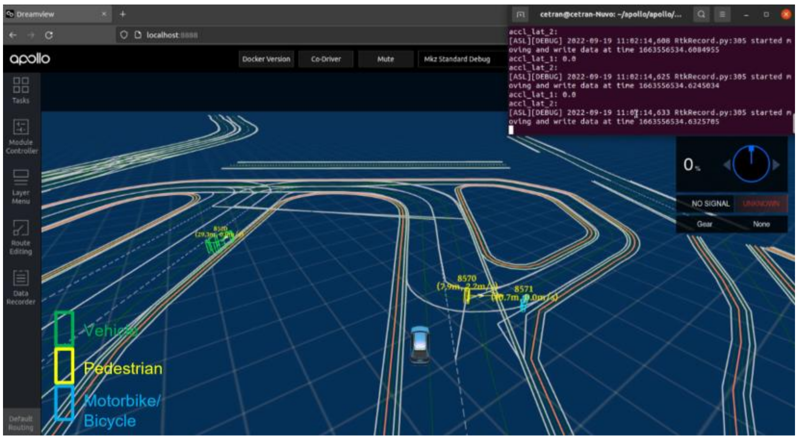


Figure 37 – Data collection from vehicle demonstrator on sample scenario at the CETRAN test track

The vehicle-based demonstrator is the ego vehicle driving on the track. The sequence of actor interactions is as follows:

- Pedestrian1: Crosses road
  - Manoeuvre: Ego vehicle will see actor approaching, slows down to a stop and allows pedestrian to cross.
- Vehicle 1: Parked at side of road
  - Manoeuvre: Ego vehicle slows down behind parked car and overtakes Vehicle 1.
- Vehicle 2: Driving in the opposite direction and side of road from Ego
  - Manoeuvre: Ego does not act from parallel drive from Vehicle2.

- Bicycle: Approaching from left of intersection and stops at stop line
  - Manoeuvre: Ego will see bicycle approaching but will have right of way and continues driving ahead.
- Pedestrian2: Cross road from left to right ahead of ego vehicle
  - Manoeuvre: Ego will see bicycle approaching but will have right of way and continues Manoeuvre: Ego will see pedestrian 2 and slows down while pedestrian continues to cross.
- Static pedestrians at traffic island



```
accl_lat_2:
[ASL][DEBUG] 2022-09-19 11:02:14,608 RtkRecord.py:305 started m
oving and write data at time 1663556534.6084955
accl_lat_1: 0.0
accl_lat_2:
[ASL][DEBUG] 2022-09-19 11:02:14,625 RtkRecord.py:305 started m
oving and write data at time 1663556534.6245034
accl_lat_1: 0.0
accl_lat_2:
[ASL][DEBUG] 2022-09-19 11:02:14,633 RtkRecord.py:305 started m
oving and write data at time 1663556534.6325705
```

Figure 38 – Sample view of detection of objects on CETRAN test track with actor type denoted

The vehicle-based demonstrator set-up allows for perception detection of a vehicle's surrounding traffic interactions through Apollo's software perception module. This is seen in Figure 38 which displays the detection in Apollo software – with the type of actors denoted. Data output from Apollo's perception module is then processed and analysed as was done with data from simulation, with the results of total risk with time shown in Figure 39 and the corresponding percentage time spent in each risk zone out of time with traffic interactions during the drive shown in Figure 40. The sample drive with real traffic interactions shows a large percentage of time spent in very safe risk levels for road interactions. Table 21 shows the summary risk values of the drive, with a maximum risk value of 4.1 which is within the high-risk band, and an average risk value of 1.6 which is in the very safe band.

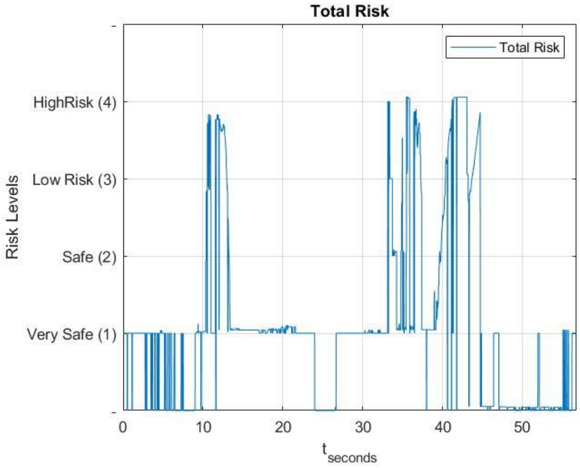


Figure 39 – Plot of total risk of all interactions during the duration of the sample drive on the CETRAN test track

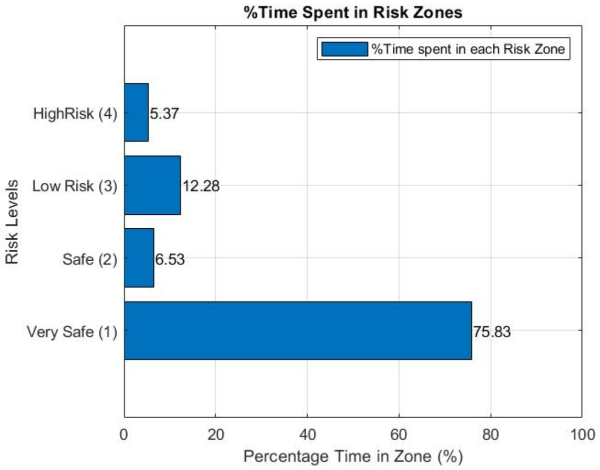


Figure 40 – Breakdown of risk of interactions displayed in percentage of interaction time spent in each risk zones

Table 21 – Statistics of sample drive on CETRAN test track

Max risk	High risk (4.1)
Average risk	Very safe (1.6)

The total risk plot shows more fluctuations of risk values as compared to previous simulation runs which could be due to missing detection data and sensitivity to change in values such as speed of actors that affects the risk calculations. Further analysis of the results then identified the interactions occurrence corresponding to each area of risk increment as seen in Figure 41. The first pedestrian crossing resulted in an initial increase in risk due to a sudden dash by the pedestrian, this peak then drops drastically as both the ego vehicle and pedestrian slowed down to a stop. There are also regions of no traffic interactions which results in no risk levels plotted. The next interaction then shows the sudden increase in risk when approaching a parked car and as the ego vehicle makes an overtaking manoeuvre, the risk drops drastically. The next rise in risk levels is due to the interaction with a parallel vehicle that is changing lanes on the opposite direction of the road while the ego vehicle is carrying out the overtaking manoeuvre, which results in the interaction seemingly in a collision course with this actor. As both vehicles straighten out in their respective lanes, the interaction would then be a parallel type and with sufficient lateral clearance, results in a very safe risk level. This is then immediately followed by a bicycle arriving from the left at the traffic intersection as the ego vehicle drives straight. This gradual increase in risk seen is due to the approaching path of the bicycle as it turns into the road on the left and heads towards the collision path. The approach speed and time left to the potential collision course also leads to the increasing risk as both bicycle and ego vehicle continues its trajectory. The risk does not drop to a very safe level even as the bicycle comes to a stop when the ego vehicle passes by as concurrently there is a pedestrian jaywalking ahead. As the pedestrian approaches the jaywalking path, the risk increases from a safe risk level to a high-risk risk level. The risk level then drops to very safe, which is indicative that a near-miss has just occurred after the pedestrian has crossed the path of the ego vehicle and is now safe. This drop from high risk level was also seen right after the interaction with the bicycle. The continuous evaluation of traffic interactions can be seen to pick up on the dynamic and continual risk of interactions in this sample drive, especially with the consecutive interactions of the bicycle and jaywalker. This first iteration of the methodology for continual risk evaluation allows for the tracking of consecutive interactions without being scenario prescriptive.

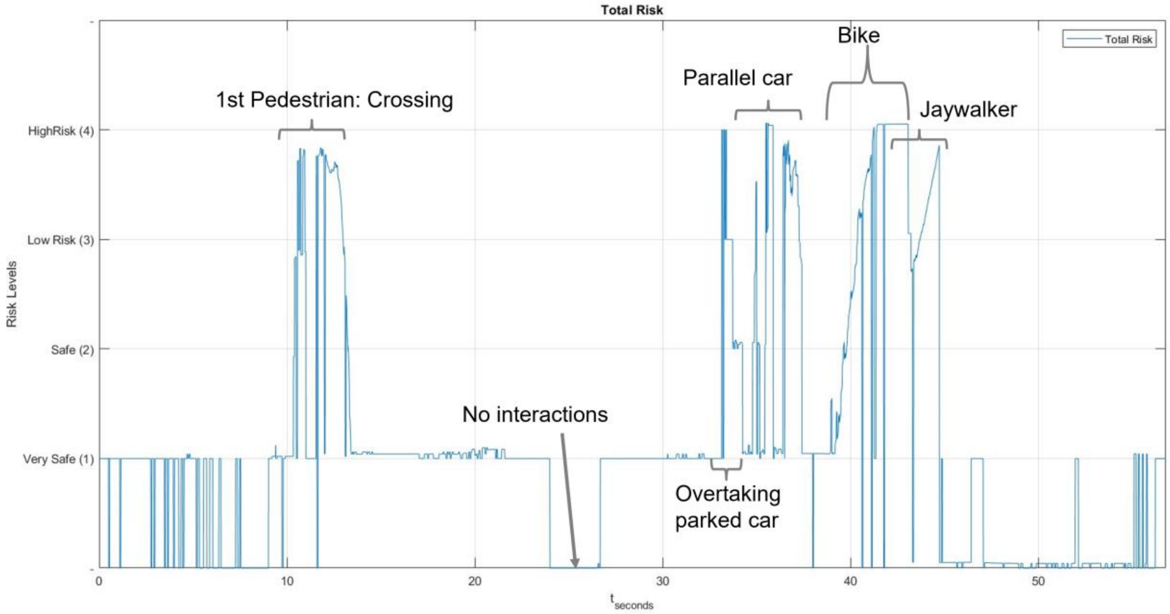


Figure 41 – Breakdown of risks during short sample drive

Analysis of the data collected by sensors from the vehicle demonstrator revealed that data from traffic interactions are not as smooth as that from simulation. Measurement of actor velocities could have noise from the sensors in addition to actual erratic movements from actors. Minor actual changes in velocities, compounded by the sensor noise and logging frequency of data could cause large changes in acceleration values in short time instances. The reality of movements by actors in real traffic interactions highlighted some sensitivities of the listed metric calculations which could cause momentary spikes in risk. This should be noted when applying the methodology of risk calculation to real-traffic data and filtering of data could be applied in the future.

6.2.4.2.2 Sample drive on public roads

The vehicle-based demonstrator was driven on a route within Nanyang Technological University (NTU) and towards the Pioneer Mass Rapid Transport (MRT) within the Jurong West region of Singapore. This is shown in Figure 42.

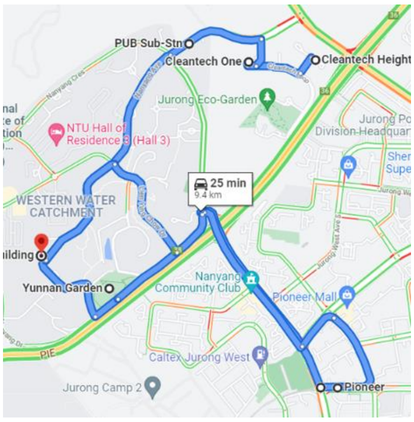


Figure 42 – Route driven on public roads in Jurong West region in Singapore

For this sample drive on public roads, there were a large number of traffic interactions detected. At various time instances, there were approximately 25 to 40 actors within proximity to the vehicle. As the route was within a university campus and residential communities in Jurong West, there were many dynamic actors on the road. These included pedestrians crossing with many pedestrians, vehicles travelling and parked on the roads. The total number of detections was possibly more than 10 000 at the end of the drive. A summary of the drive is shown in Table 22.

Table 22 – Statistics of sample drive of vehicle-based demonstrator on public roads in Jurong West region in Singapore

Length of drive	~ 25 minutes
Number of actors within proximity	At times 25-40
Number of detections	~ 10 000
Time of drive	12-1 + pm

The methodology is applied in a similar manner with the results shown below through a plot of total risk with time in Figure 43 and percentage of time in each risk band over total time of all interactions in Figure 44. The majority of the drive was in a very safe risk band with about 12-13% in low-risk and high-risk levels out of the time with interactions during the drive. This is reflected in the average risk score of 1.6 which is very safe seen in Table 23. The maximum risk of 4.9 within the very high-risk band is shown to have resulted within the duration of the drive.

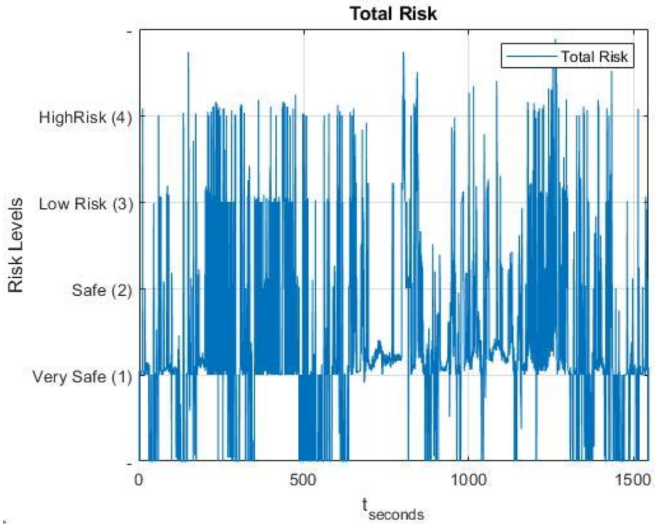


Figure 43 – Plot of total risk of all interactions during the duration of sample drive on public roads in Jurong West

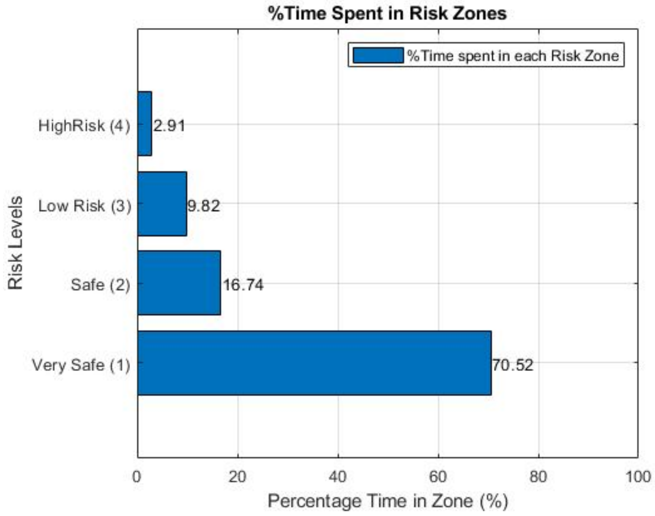


Figure 44 – Breakdown of time spent in each risk zone displayed in percentage of total time with interactions

Table 23 – Summary of risk for sample drive on public roads

Max risk	High risk (4.9)
Average risk	Very safe (1.6)

6.2.5 Observed limitations

- Matlab – perfect data; no perception issues
  - Good for checking of metrics thresholds
    - Easy comparison: parameters can be changed easily.
    - No ADS behaviour change to account for. One parameter change at a time.
  - Too simplistic – does not identify complexity from real-traffic data
    - Objects are perfectly detected, signals are smooth – no fluctuations, object list is ordered. No false detection.
- AV simulation – some realism of data issues from software (Apollo)
  - Data quality
    - Fluctuations of actor speed – not realistic.
    - Model movements not realistic – erratic heading when crossing street.
    - Order of object list would not be consistent – tracking of actors is difficult.
    - False detections within simulation.
- Vehicle-based demonstrator
  - Data quality
    - Order of object list would not be consistent – tracking of actors is difficult.
    - Perception
      - Changing object ID for the same actor.
      - False detections
        - Wrong classification: Bus-stop as vehicle, object as pedestrian.
        - Non detection: e.g., of dummies/pedestrians.

        - Accuracy (positions, velocities)
            - could be improved with more sensors
            - depends on perception module processing (software could have limitations: Apollo in this case)
    - Maps – only for visualization
        - HD maps: Observe some inaccuracies (missing/offset) of data which is likely from source: government curb lines.
        - Because the methodology is non-infrastructure dependent, the methodology picks up risks from oncoming traffic even though they are in their lanes e.g., Roundabout exits (seen as collision) → could incorporate infrastructure/rules of road and ego expected behaviour
            - Currently methodology is on interaction risk – does not consider ego behaviour → could be included in future risk assessment
- Current methodology limitations and possible future improvements
    - Threshold calibration
        - Continuity of risks between interaction type.
            - Variation of risk should be the result of change in hazard profile and not just change in situation. Difficult to evaluate.
        - Investigate the possibility of actor data to include size of vehicles/ bounding box to improve accuracy of clearances/metrics.
        - Clearcut thresholds now could be flexible with some parameters – e.g., Speed for parallel.
    - Metric refinement
        - For situations like parallel/static and aside – could consider the distance to the ego to differentiate risk level in more granularity.
        - Ego behaviour could be incorporated in assigning risk
            - Road rules
            - Execution of manoeuvres

In the CETRAN AV test centre based trial, there are 2 pedestrians crossing the road to simulate "the Molly problem". In the first instance, the driver acts in a courteous manner and stops and waits for the pedestrian to cross. In the second instance there is careless behaviour of both driver and pedestrian. This data shows that risk is reduced when the vehicle has stopped, but did not show enough differentiation in risk in the events overall. While this shows that metrics in general can be used to assess the situation, overall metrics need to be improved to better distinguish the perceived level of risk of these situations.

6.2.6 Link to data

All data for the trials can be downloaded from the following address:

https://doi.org/10.21979/N9/MEXRUP

The download location is DR-NTU (Data) is internationally certified as a trusted data repository by CoreTrustSeal.

In the download are five files:

1) CSV file with recorded data from trial at CETRAN AV test centre.
2) MP4 video file with overlay of driver view camera and Apollo perception log from trial at CETRAN AV test centre.

3) CSV file with recorded data from trial from trial on public roads on and near Nanyang Technological University.

4) MP4 video file with overlay of driver view camera and Apollo perception log (at 20x speed) from trial on public roads on and near Nanyang Technological University.

5) MP4 video file with overlay of driver view camera and Apollo perception log (at normal speed) from trial on public roads on and near Nanyang Technological University.

Total download size is over 2.1 GB.

References used in this section

[1] S.M. Sohel Mahmud a (2017), Application of proximal surrogate indicators for safety evaluation, IATSS Research, March.

[2] A. V. S. Consortium™, AVSC00006202103:AVSC Best Practice for Metrics and Methods for Assessing Safety Performance of Automated Driving Systems (ADS)., SAE, March 2021.

[3] S. S. Council (2021), TR 68 – *National Standards for the Safe Deployment of Autonomous – Part 1 Basic Behaviour*, Enterprise Singapore, September.

[4] T. Police (2020), *Basic theory of Driving Singapore*.

[5] T. Police, *Final Theory of Driving*.

[6] ISO, ISO 26262-1 (2018), *Road vehicles – Functional safety – Part 2* clause 5, Part 3 clauses 6,7, Part 4 clause 8, Part 9 clause 5.

[7] G. M. C. W. M. U. Rami Debouk, *Overview of the 2nd Edition of ISO 26262:Functional Safety – Road Vehicles*.

[8] C. Jurewicz, A. Sobhani, J. Woolley, J. Dutschke and B. Corbern (2016), *Exploration of vehicle impact speed – injury severity relationships*, Transport Research Procedia 14, pp. 4247-4256, December.

[9] H. Yang, K. Ozbay and B. Bartin (2010), A*pplication of simulation-based traffic conflict analysis for highway safety evaluation*, in World Conference on Transport Research, Lisbon, Portugal.

[10] I. The MathWorks, *Driving Scenario Designer*, Mathworks, 2022. [Online]. Available: https://www.mathworks.com/help/driving/ref/drivingscenariodesigner-app.html. [Accessed February 2022].

[11] Baidu, *Apollo Open Platform*, Baidu , 2020. [Online]. Available: https://developer.apollo.auto/developer.html. [Accessed March 2022].

[12] LG Electronics, *SVL Simulator*, SVL, 2022. [Online]. Available: https://www.svlsimulator.com/. [Accessed March 2022].

[13] Apollo, *Key Perception Components*, Baidu Cloud, 2022. [Online]. Available: https://developer.apollo.auto/platform/perception.html. [Accessed 2022].

[14] Practical Digital Signal Processing, *Median Filtering*, Elsevier B.V., 2003. [Online]. Available: https://www.sciencedirect.com/topics/engineering/median-filtering. [Accessed July 2022].

6.3 Smart Mobility Living Lab

The purpose of this section is to provide an overview of the Smart Mobility Living Lab (SMLL), a real-world testbed located on public roads in London, UK. This document covers:

- The origins and mission of the testbed.
- The set-up of the facility (covering the test environment, testbed capabilities, and the data gathering infrastructure used to support clients in their ambitions).

- Our approach to testing, relevant to the scope of the ITU FG AI4AD activities.
- Examples of how the SMLL testbed has been used to generate outcomes in specific projects.

6.3.1 SMLL: Origin and mission

The Smart Mobility Living Lab (SMLL) is a London-based real-world connected environment for testing and developing future transport and mobility solutions. It is the world's most advanced urban testbed of its kind with the sole purpose of accelerating the creation of mobility solutions that are clean, efficient, safe, reliable and convenient for everyone using public and private roads in London, to develop and validate new mobility and transport technologies.

The locations of the Royal Borough of Greenwich and Queen Elizabeth Olympic Park in Stratford provide a complex uncontrolled testing environment, interacting with live traffic and other road users. The testbed is designed to demonstrate and evaluate the use, performance, environmental impact, safety and benefits of connected and automated mobility technology and future transport services.

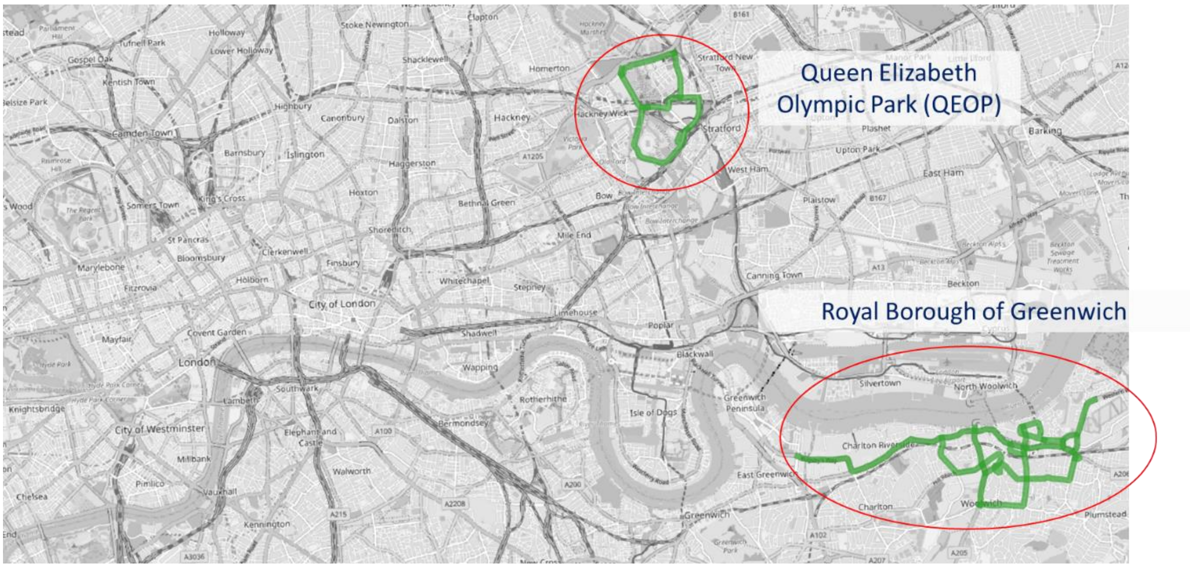


Figure 45 – Location of SMLL routes in London

The creation of the Smart Mobility Living Lab was jointing funded by the UK government and industry, via a consortium led by TRL, with partners including Cisco, Cubic, DG Cities, Loughborough University, London Legacy Development Corporation, and Transport for London. The build of the testbed took place between April 2018 and October 2020, since when it has been operated on a commercial basis by Smart Mobility Living Lab ltd., a TRL company. The Smart Mobility Living Lab testbed is a £25m facility.

Ultimately, the SMLL build project resulted in the delivery of an operational testbed for new mobility in east London (the Smart Mobility Living Lab), coming in on budget and within the project deadline. The testbed and the assets developed within the project were novated over to SMLL Ltd. for it's ongoing operation and exploitation with the mobility sector. The testbed consists of 24 km of test route, covered by 200+ monitoring points, instrumented with 300+ cameras, as well as a 20 km private fibre network and local mesh network, V2X infrastructure and LoRaWAN for IoT devices. This is combined with workshop facilities, control rooms, data centres, and open architecture vehicles. The successful delivery of the SMLL build project has enabled the trialling of numerous CAV projects in London, as well as the validation of systems and solutions across the mobility ecosystem.

The SMLL is a member of Connected Automated Mobility Testbed UK (CAM TBUK), a network of 7 test facilities in the UK. CAM TBUK is focused on enabling the development and validation of connected and automated mobility technologies and services. It does this by providing a range of connected and automated mobility testbed facilities, spanning controlled environments, semi-controlled settings (e.g., campus locations), and real-world locations. Each facility is operated as a standalone business, but efforts have been made to facilitate interoperability between the testbeds to enable clients to move between the testbed facilities seamlessly. For instance, an automated driving system technology developer could arrive at a controlled environment with a lower technology readiness level product, before progressing to more complex environments at another facility once they have validated their capabilities.

6.3.2 SMLL set-up

6.3.2.1 Testbed locations, environments and use cases

SMLL is used to demonstrate and evaluate the use, performance, safety and benefits of connected and automated mobility (CAM) technology, future transport services and mobility solutions. As such, the choice of locations for the testbed, and the placement of the monitoring infrastructure used to evaluate the performance of new mobility technologies and services has been critical in assessing the impact on people and places.

The locations of the Royal Borough of Greenwich and Queen Elizabeth Olympic Park in Stratford were chosen specifically to create a complex public testing environment, with 24 km of public roads. They provide a representative and technologically challenging cityscape which are ideal to evaluate the interaction of new technologies and services with their environment, and to understand the impact on people and place. Figure 46 and Figure 47 illustrate the SMLL route locations in London, as well as the location of the roadside monitoring infrastructure.

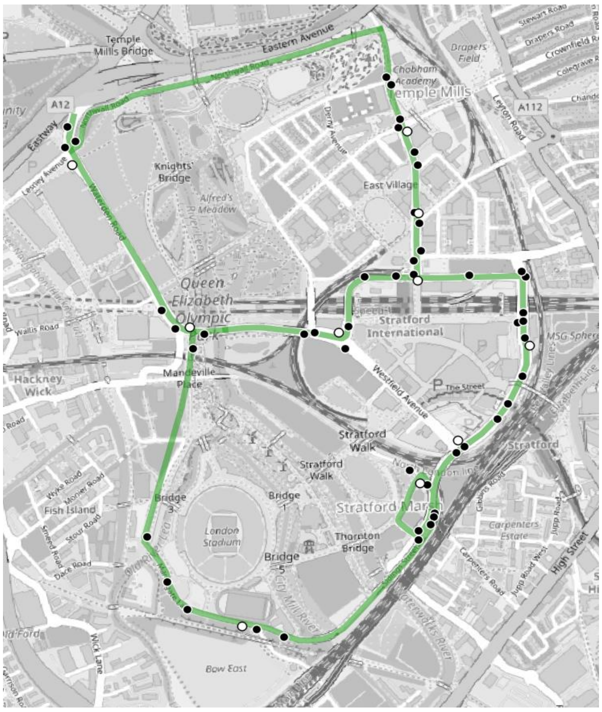


Figure 46 – SMLL routes in the Queen Elizabeth Olympic Park, Stratford, London (note, the black dots represent SMLL roadside monitoring infrastructure locations)

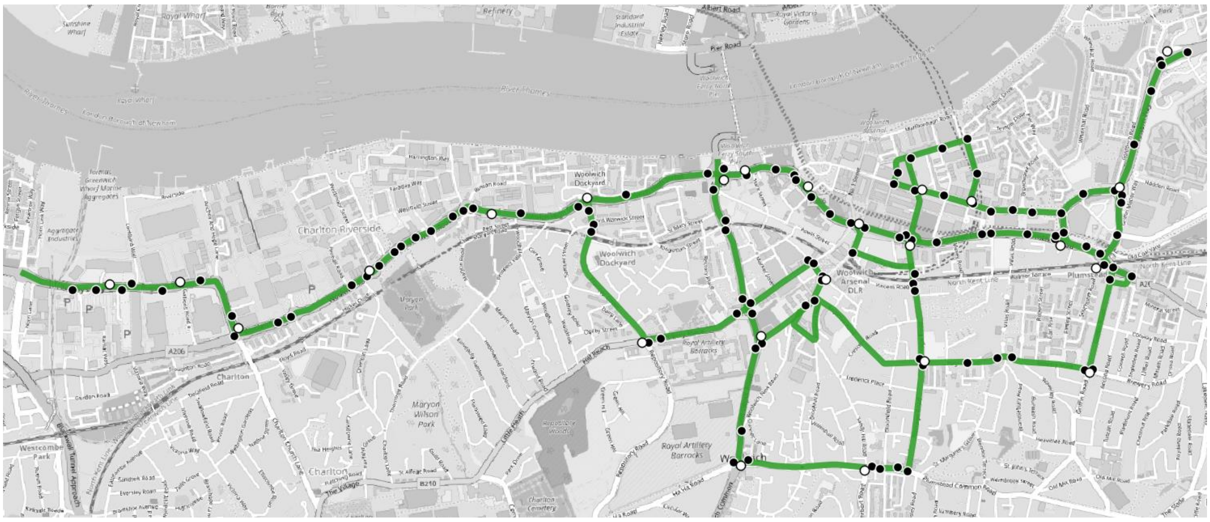


Figure 47 – SMLL routes and monitoring locations in the Royal Borough of Greenwich, London (note, the black dots represent SMLL roadside monitoring infrastructure locations)

The choice of SMLL testbed routes was a deliberate process. Consideration was given to a range of factors to ensure that there is an opportunity to expose the mobility technologies and services under test to as wide a range of operational scenarios as possible. The factors inputting to the choice of routes are displayed in Table 24.

Table 24 – SMLL route design considerations

Route choice factor	Description
Operational environments	E.g., residential, town centre, retail, commercial, sports and event venues, leisure environments
Road and junction types	E.g., Single track, dual carriageways, varying speed limits (5 mph to 50 mph), junction types (< 90% of the UKs road junction types feature on the SMLL routes), temporary road network changes (e.g., roadworks, event overlays)
Environmental features	1000+ unique road features including mini-roundabouts, bridges, underpasses, chicanes, cycle lanes, bus lanes, crossing types etc.
Dynamic behaviours	E.g., traffic densities, vehicle manoeuvres, road user behaviour
Interaction with other transport modes and road users	Includes consideration of interaction with other public transport modes and stations, taxi & private hire, active travel, emergent modes (e.g., e-scooters), and vulnerable road users
Historic collision record	E.g., taking account of previous incident records to identify trends linked to road geometries, environments and road user behaviours
Risk and hazard review	Focused on the complexity of the environments, and application to different use cases
New mobility technology and service use cases	E.g., Deliveries, passenger transport services, connectivity use cases (e.g., V2X applications, connected kerb), automation use cases (including ADAS applications), GNSS and sensor applications (e.g., areas with constraints e.g., multi-path for GNSS)

The elements of the highways code relevant to connected and automated driving have been mapped to an area of the SMLL route. This provides value both in planning particular operational design domains, and characteristics that a CAM services would need to abide by, but also in selecting the test cases to validate the prospective performance of a technology.

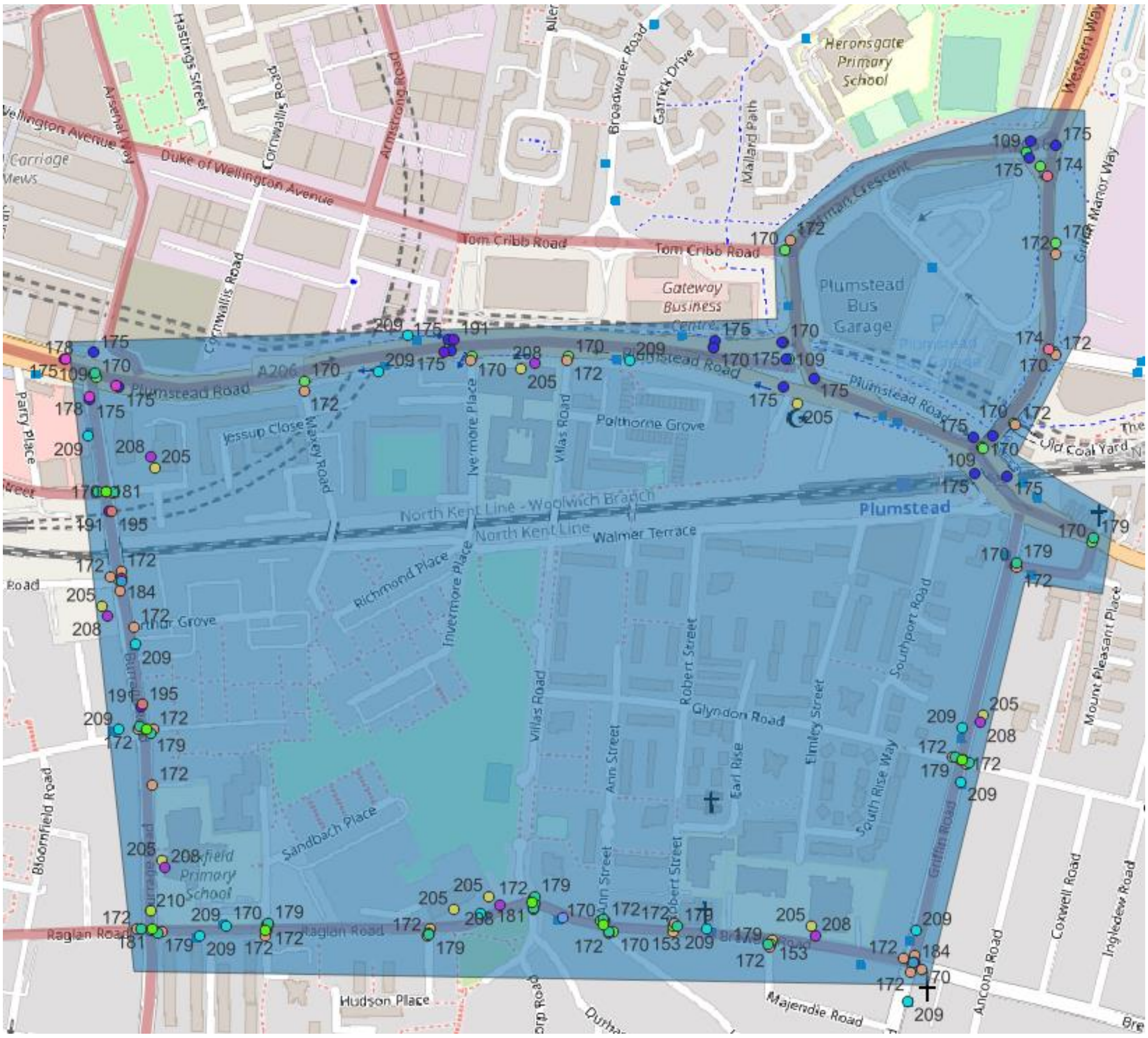


Figure 48 – Example of mapping the UK highway code rules to an element of the SMLL network

6.3.2.2 Physical data gathering and connectivity infrastructure

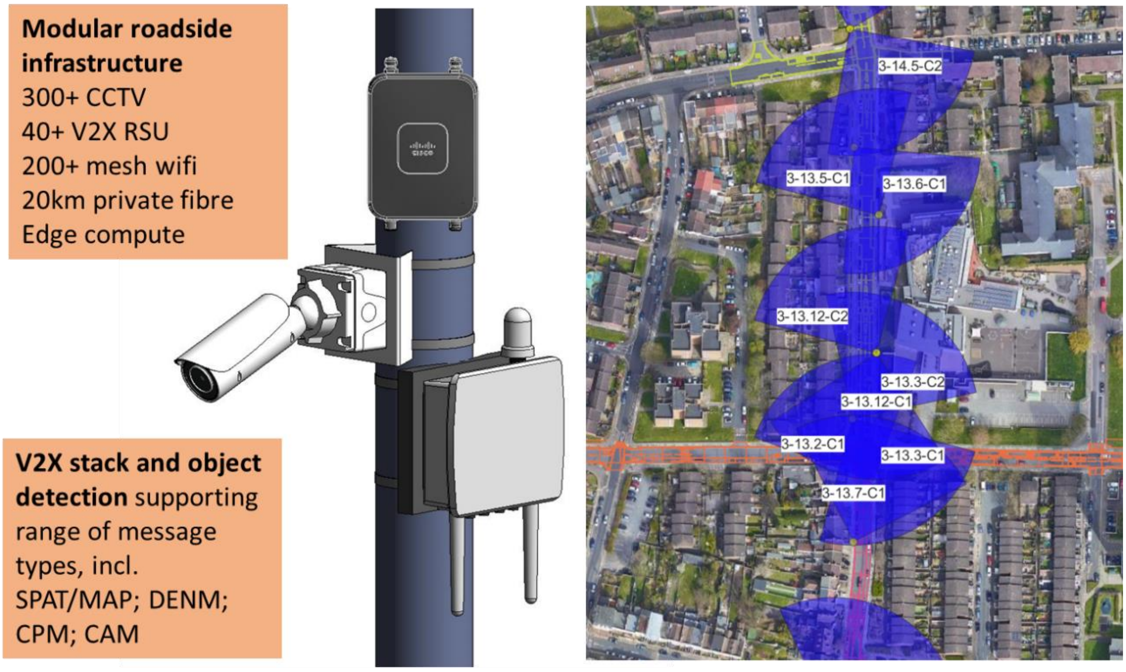
6.3.2.2.1 Roadside infrastructure and connectivity

The SMLL testbed is enabled by a unique and sophisticated infrastructure. Roadside monitoring equipment (fixed and flexible), technology connectivity using a range of protocols, an extensive data platform and a complex road environment create a complete and challenging real-world urban testbed (Figure 49). These facilities provide comprehensive and detailed test and trial evaluation and data for all kinds of CAMs, infrastructure, services and components, in London as well as in other cities in the UK.

A critical component of the SMLL is the ability to gather and provision different information assets as part of the process of developing independent evidence of the performance of different solutions and events. To this end, the system was design and built to deliver a range of different information assets within SMLL including:

- Raw and processed roadside CCTV data (e.g., object and event detection, classification, localisation) from 300+ cameras across a 24 km network.
- Vehicle data outputs (e.g., LiDAR, radar, stereo and mono camera outputs, GNSS IMU positional data, telemetry, CANbus data from SMLL's open architecture automated vehicles, plus customer data where relevant).

- Connectivity data (e.g., performance of connected devices, RF audits, V2X data) underpinned by a 20 km private fibre network, 200+ Wi-Fi access points, 40+ DSRC radios, and edge compute devices.
- Incident data (from STATS19, as well as incidents relating to trialling on our network, captured and managed using TRL's iMAAP software).
- Performance benchmarks of different solutions.
- User insights and behavioural.



Modular roadside infrastructure
300+ CCTV
40+ V2X RSU
200+ mesh wifi
20km private fibre
Edge compute

V2X stack and object detection supporting range of message types, incl.
SPAT/MAP; DENM;
CPM; CAM

Figure 49 – SMLL roadside monitoring infrastructure

6.3.2.2.2 SMLL open architecture vehicle development platforms

SMLL has invested in multiple open architecture vehicles (Figure 50). These use the Nissan eNV200 as the base vehicle platform, and have been fitted with a Streetdrone drive-by-wire platform. Additionally, these vehicles have been equipped with a range of sensors to facilitate automated driving (LiDAR, RADAR, stereo and mono cameras, ultrasonics, inertial navigation ground truth) and connectivity needs (e.g., DSRC OBUs, 4G/5G networking). The open architecture vehicles are also installed with LINUX PCs, and the open source Autoware automated driving system.

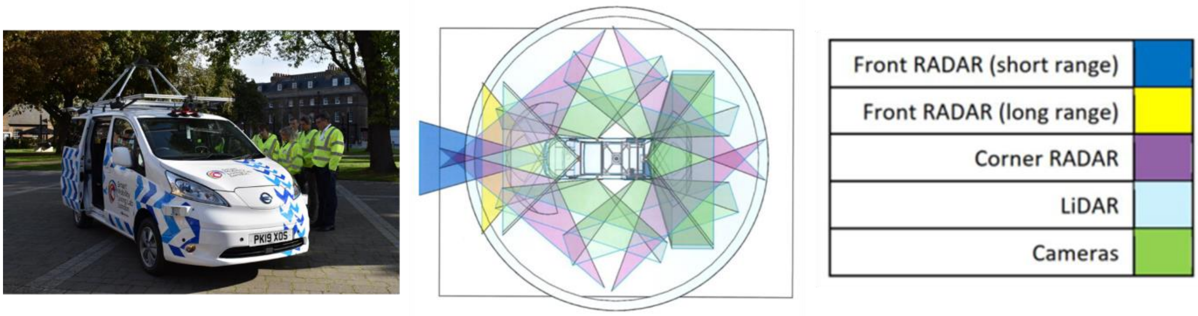


Figure 50 – SMLL open architecture vehicle platform

These vehicles have been used for not only CAV trials, but also to enable clients to test individual system components, integrating their solution with the wider vehicle stack. Additionally, the open architecture vehicles are used as data gathering platforms to validate performance.

A particularly effective use of the open architecture vehicles has been to use them to develop shared situational awareness of events by combining the on-vehicle data and the roadside data. This is illustrated in Figure 51.

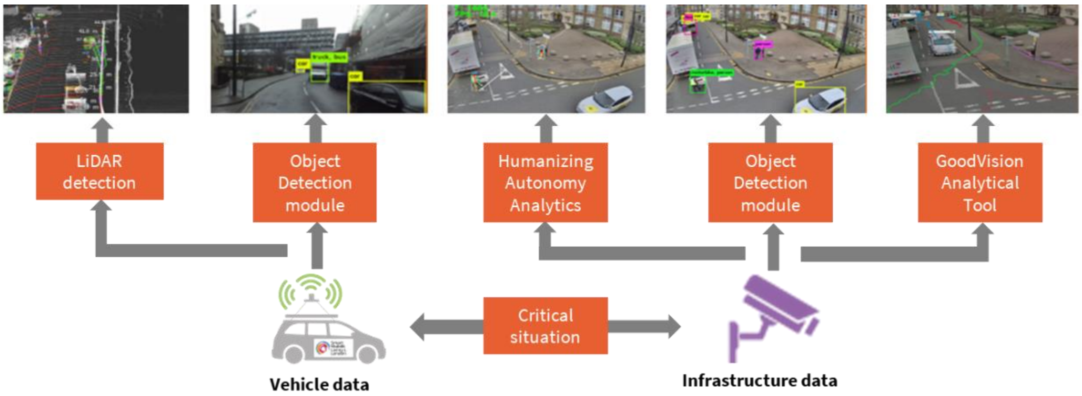


Figure 51 – Shared situational awareness combining roadside and on-vehicle detection (example from a 2020 CAV trial at SMLL)

6.3.2.3 Digital twin infrastructure

SMLL has a digital twin of its routes, and the data relating to movements and behaviours on that route. This is summarised in four areas:

- **SMLL environment model** (i.e., a representative digital model of the SMLL environment incorporating the geometry of the test bed route and associated attributes and layers).
- **Behavioural scenarios** (i.e., observed scenarios that have been derived from real-world driving events on the SMLL network, parameterised to run in simulation platforms**).**
- **Simulation platform** (i.e., the environment in which simulation scenarios can be deployed and executed, including facilitating ADS in the loop testing).
- **SMLL traffic model** (i.e., VISSIM based-model of traffic flows, with interface to the UTC SCOOT application).

6.3.2.3.1 SMLL environment model

SMLL has a digital model covering the full 24 km of route.

- High fidelity 3D models for all SMLL routes, featuring all road and roadside infrastructure assets, including surfaces, road markings, road and roadside furniture and vegetation – in Revit, 3DS max, FBX and other various 3D formats.
- Selected areas also include buildings as well as photographic textures for all surfaces for a more realistic look.
- LiDAR scanned point clouds for all 24 km of SMLL routes.

Figure 52 shows a sample of digital model of SMLL routes.

Note, these models interface with the CARLA open-source simulator for autonomous driving research or a simulation host platform to run the model within. Models can be exported in the necessary format to load into CARLA.

Figure 52 – Sample of digital model of SMLL routes

6.3.2.3.2 Simulation scenarios

SMLL have developed a scenario generation pipeline that can be used to detect, create and execute real-world scenarios in simulation. Currently this method creates the individual test cases or concrete scenarios that are then run in the same 3D representation of the environment.

Figure 53 depicts the high level workflow for SMLL's scenario generation pipeline:

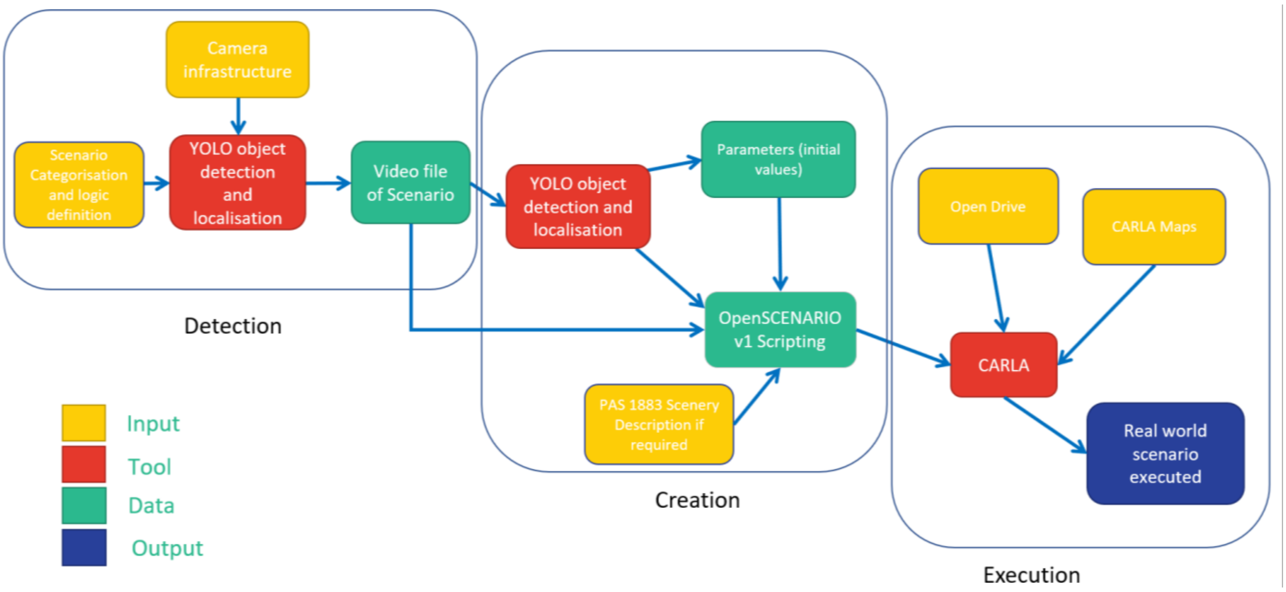


Figure 53 – High level data flow of scenario generation

Scenarios are captured based on what is occurring in the scene. This stores 1 minute video files that are then used for processing in the creation stage. The logic can be dependent on the vehicle information:

- Speed in approach, acceleration or deceleration to a location.
- Proximity to points in the road.
- Class type i.e., pedestrian.

Using combinations of this logic and the boundaries of acceptable values a library of scenarios can be collected.

An example of this was in a recent project with Zenzic and a consortium of partners, we sought to detect driving event scenarios that were relevant to the autonomous emergency braking – pedestrian (AEB-P) system. To capture this scenario, basic logic about the vehicle movements and classification were used. Firstly, it was necessary for a pedestrian to be in the scene along with a vehicle, this was applied with a hard deceleration of > 5 m/s^2 to capture a car braking hard for a pedestrian stepping out at pedestrian crossing. See Figure 54.

Figure 54 – AEB-P scenario

Once the scenario has been detected the video can be run through the object detection algorithm once more to generate information about each object at each frame along with an annotated video. The information is stored in CSV file containing the speeds, positions and crucially time.

Through a combination of using the data in the CSV file and the annotated video, the following was identified:

- The Ego vehicle.
- Key agents in the scene.
- Manoeuvres.

This information was then used to construct an openscenario V1.0 using actions only. A library of manoeuvres was called upon to build the scenario with the initial conditions and triggers. The timing data and positional data allows for triggers in the openscenario file to be scripted appropriately so that it can be executed with a vehicle under test. Parts of the scenario can then be parameterised to further extend the testing of the system for example varying the speed of the Ego vehicle on approach or different environmental conditions.

6.3.2.3.3 Simulation platform

SMLL has utilised CARLA, and incorporated our environmental models into this. To enable CARLA to be used to incorporate the representative data captured in the real world, SMLL has adjusted the source code of CARLA to support left-hand side driving.

In the case of the AEB-P scenario referred to above, the developed scenarios were executed in SMLL's CARLA simulator to test and debug. Scenario runner was used to execute the scenario, the automatic_driving.py was also used as the ego vehicle in this instance. This operates the car with basic CARLA driving rules as per the opendrive file. Each scenario is executed in the exact same location in the simulation as it was detected, this can become extremely powerful tool to extend testing, de-risk real world trials and potentially to virtually certify vehicles in the future.

Figure 55 shows the AEB-P scenario running in CARLA vs. the real scenario. These images showcase the strong correlation between the simulation and the real-world video captured.

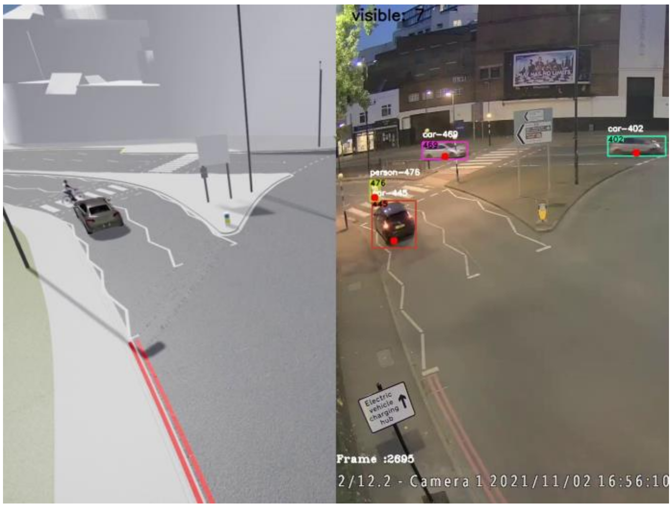


Figure 55 – AEB-P scenario: CARLA (left) - Real world (right)

6.3.2.3.4 SMLL traffic model

Urban traffic control (UTC), powered by SCOOT 7 is a specialist form of traffic management that, by coordinating traffic signals in a centralised location, minimises the impact of stop times on the road user. Integrating and coordinating traffic signal controls over a wide area allows great control of the traffic flow.

Combined with VISSIM traffic simulation software, UTC provides a modelling environment which brings in real world traffic control systems which can be used to model and test control strategies, incident responses and provide the data enabling a green light optimal speed advisory (GLOSA) service, along with future connected vehicle data exchange use cases.

SMLL has developed sample traffic in models VISSIM for key parts of the route in Greenwich. This enables us to develop insights regarding the network impact that introducing certain new mobility modes and services would have. These microsimulation models can also be integrated with the simulation environments for ADS development to generate agents as part of scenarios.

These are available as a cloud based product via an open data API.

6.3.3 SMLL approaches to testing

This clause comments on SMLL's approach to testing for those areas relevant to the ITU FG AI4AD work.

6.3.3.1 Perception and driving behaviours

SMLL can support the development of client's mobility solutions from concept to commercialisation. A key element that we foresee in the future relates to supporting the in-service monitoring of deployed connected and automated mobility solutions and services.

In the context of the ITU FG AI4AD work, this particularly relates to SMLL's capacity to analyse the performance of the ADS' perception modules, and the resulting driving behaviours. This is summarised in Figure 56.

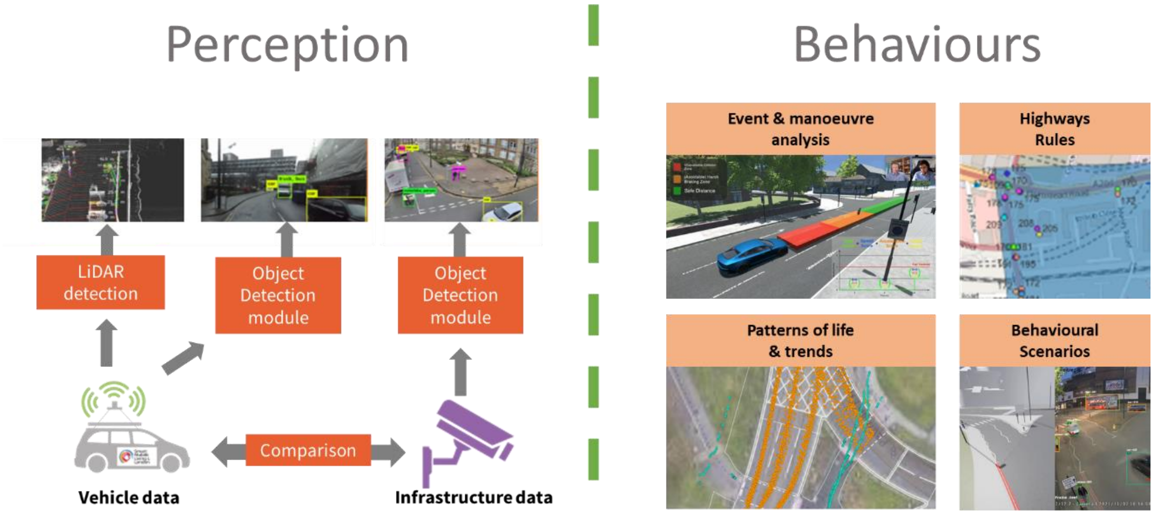


Figure 56 – Analysis of perception modules vs resulting driving behaviours

6.3.3.2 Using SMLL infrastructure and test methods to collate, analyse and appraise data

Key data to be captured to enable behavioural scenarios to be identified, categorised and appraised included information relating to the scenery, environmental, and dynamic elements of an event.

SMLL has used its roadside monitoring infrastructure to capture this data, such that behavioural scenarios can be analysed.

One such recent exercise focused on the analysis of collisions, parameterising behavioural events such as speed events, and lateral jerk. Figure 57 shows the max speed location of speeding events at a junction. An area of interest to validate this approach was selected on the basis of historic collision record, trend analysis of the speed profiles, and instance of high speed events, and lateral jerk events. The purpose of this was to develop a local risk profile and use the SMLL roadside infrastructure to evaluate current driving behaviours.

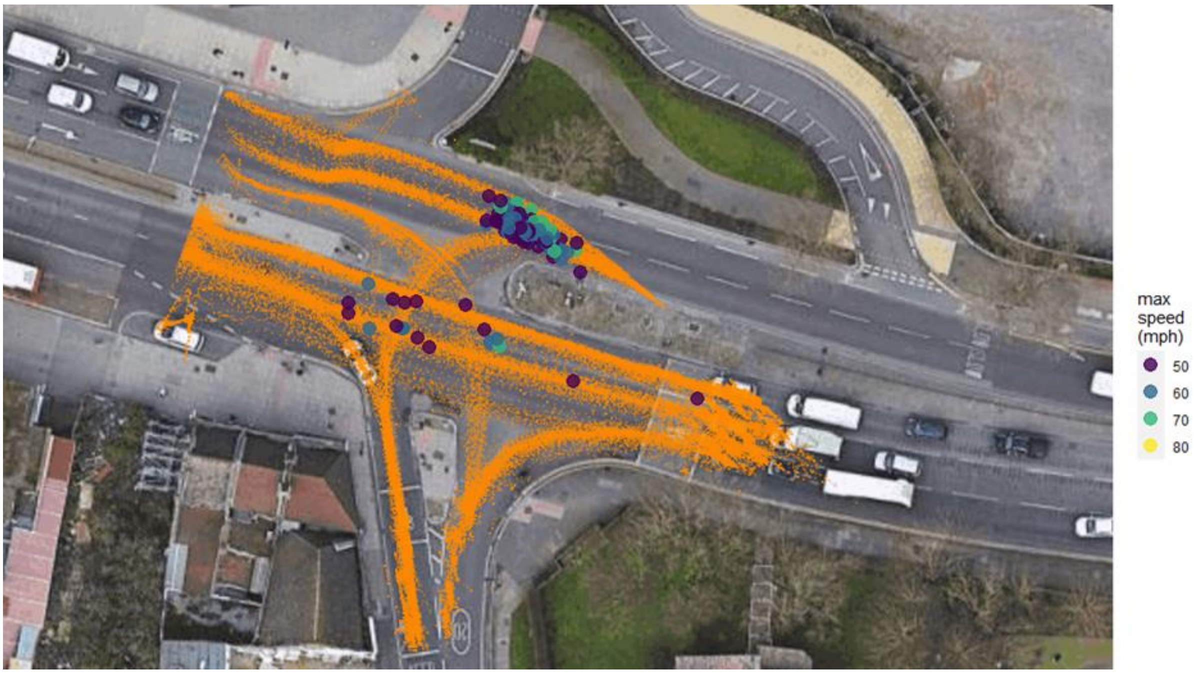


Figure 57 – Max speed location of speeding events at the junction

One of the outcomes of the research project was a conclusion that the data points, shown in Figure 58, represented an ideal set to aid the reconstruction of scenarios, and evaluation of the behaviours:

Id
Date of Detection (dd/mm/yy)
Time of Detection (hh:mm:ss)
Camera ID (Scene Location)
Detected Location (lat, long)
No. of Actors in Scene
No. of vehicles in Scene
No. Pedestrians in Scene
Time to Collision (s)
Primary Actor Track ID
Secondary Actor Track ID
Primary Actor Class
Secondary Actor Class
Max Jerk Primary Actor
Max Jerk Secondary Actor
Max Speed of Primary Actor
Max Speed of Secondary Actor
Primary Actor Manoeuvre
Secondary Actor Manoeuvre
Event type
Traffic density

Flow rate
Presence of parked vehicles
Presence of special (emergency) vehicles
Traffic agent type
Rain (precipitation)
Visibility
Snow
Precipitation type
Streetlights
Cloud cover
Daylight Conditions (match event time vs sunrise/set time)
Wind Speed
Road surface condition
Risk priority (0-3) (0 = non event, 3 = max risk)

Figure 58 – Full list of ideal data fields based to evaluate scenario-based risk metrics

7 Discussion

This document describes the set-up of monitoring and assessment of AVs in three demonstrators; COLUMBUSS in the Netherlands, CETRAN in Singapore and Smart Mobility Living Lab in England. Although the set-ups are different, all demonstrators acknowledge that objectivation of the ODD and the metrics and thresholds are required for monitoring and assessing AVs. Also, they agree there is no validated method yet. All three demonstrators are starting with the practical implementation of objectivation of the ODD, metrics and thresholds. So far, they have no proof of their concepts.

The COLUMBUSS demonstrator uses metrics and thresholds developed by Green Dino to train and test human drivers on driving simulators. The driver safety score (DSS) proved to be a reliable and valid metric for predicting accident involvement. Green Dino adjusted this metric for the monitoring and assessment of AVs. A prototype is implemented in a professional truck and bus driving simulator for human-in-the-loop testing and in a multi-purpose vehicle that drives at low speed.

The practical demonstration of monitoring and assessment of AVs is in a very early stage of implementation without good results to support validation. The complexity and costs of the projects are enormous. The set-up of the demonstrators differs, but the ideas behind the research and development do not. Working together and making a joint effort will lower complexity and costs and speed up licensing. Therefore, a standard approach as targeted by the Focus Group on AI for autonomous and assisted driving is very important.

Automated driving offers excellent potential for solving big social issues like traffic injuries and deaths and accessibility of transportation of people and goods. Open source and open access are necessary for the transparency and affordability of self-driving technology.

8 Conclusions and next steps

robotTUNER, CETRAN and TRL/SMLL, who acted as editors of this Technical Report under the FG-AI4AD, agreed to continue the monthly meetings and collaborate on developing and testing an open-source platform for monitoring and assessing AVs. The COLUMBUSS project office offered to host these activities. Any FG-AI4AD participants who wish to contribute to this effort are welcome to join.

It is recommended that the safety state framework (SSF) prototype be improved because it objectivates the individual safety performance of all traffic participants. This enables the assessment of driver safety skills, which is impossible with safety envelopes/cushions. White box AI and human peer performance reference models also improve human acceptance of automated driving. The methodology is comparable with the assessment of novice human drivers and, therefore, relatively easy to implement. Making it a standard, the commercial use of AVs will speed up.

The CETRAN metrics knowledge, and thresholds and scenarios collected in the Smart Mobility Living Lab in London, will be used by Green Dino to improve the SSF. The improved SSF should be implemented and tested in all three demonstrators to support thorough validation. A performance benchmark with alternative software will give insight into the value of the SSF serving as a standard for monitoring and assessing the behaviour of AVs in simulation and natural traffic. robotTUNER, the Metropole Region Rotterdam, The Haque/ MRDH and the Province of Groningen are committed to supporting these activities financially.