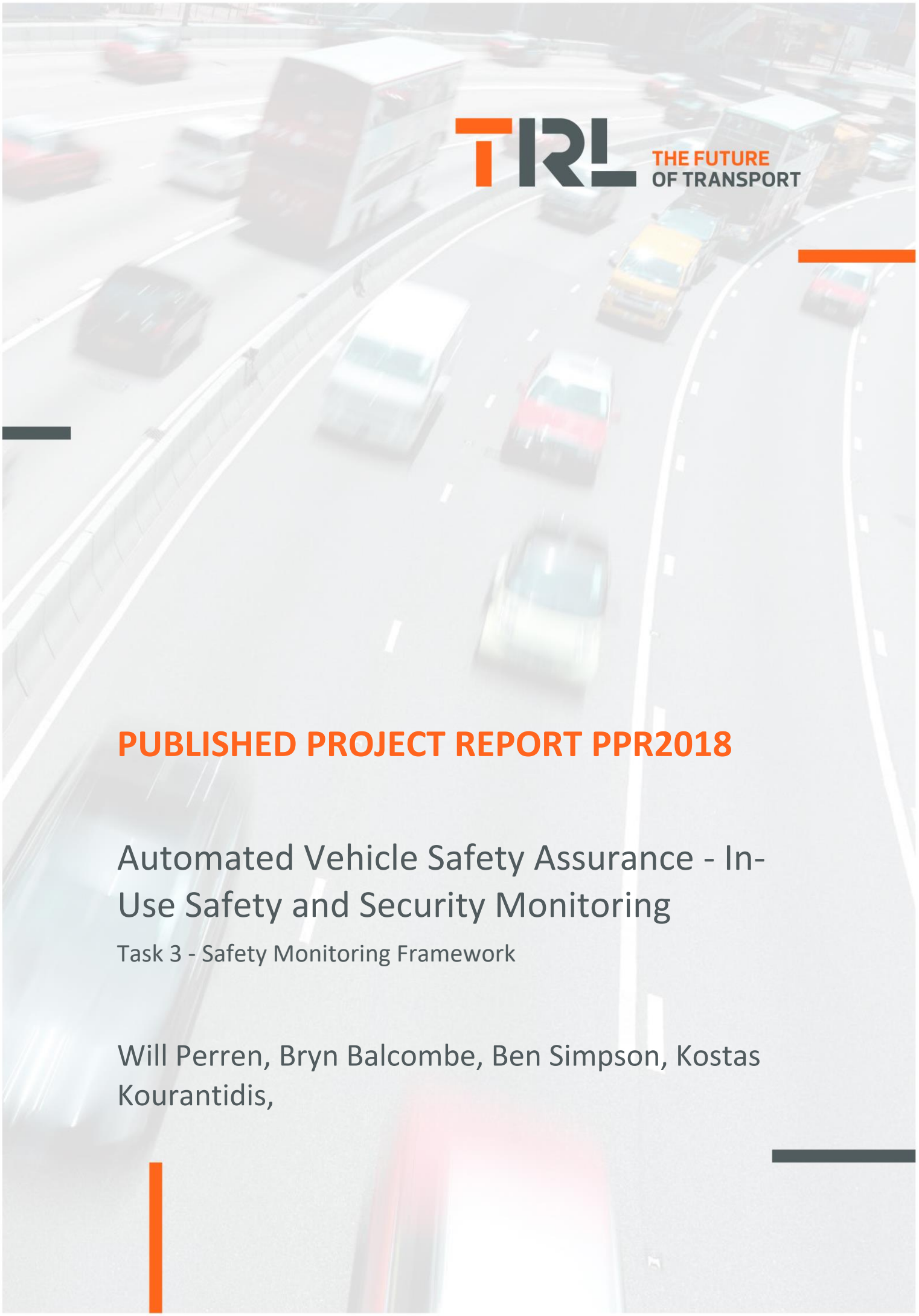


THE FUTURE OF TRANSPORT

PUBLISHED PROJECT REPORT PPR2018

Automated Vehicle Safety Assurance - In-Use Safety and Security Monitoring

Task 3 - Safety Monitoring Framework

Will Perren, Bryn Balcombe, Ben Simpson, Kostas Kourantidis,

Report details

Report prepared for:	Department for Transport		
Project/customer reference:	TETI0042		
Copyright:	© TRL Limited		
Report date:	30/06/2022		
Report status/version:	1.0		
Quality approval:			
Gareth Slocombe (Project Manager)	G. Slocombe	David Hynd (Technical Reviewer)	D. Hynd

Disclaimer

Executive Summary

The Department for Transport (DfT) has commissioned this work to contribute to ensuring that Automated Vehicles (AVs) coming to market are both safe and secure. This is an opportunity not only to set national standards to facilitate deployment of AVs and attract home investment, but also to lead the way for regulatory approaches internationally.

In this project the scope is limited to slow moving (up to 20 mph), light-duty (up to 3,500 kg GVW) fully electric pods and shuttles without a driver (or, in the context of the Law Commissions' project, a user-in-charge) who is replaced by an Automated Driving System (ADS), that will have real-world applications for automated vehicles operating on public roads in a mixed-traffic environment. The vehicles may be passenger-carrying (maximum of 16 occupants, may include standing passengers) or goods-carrying. This will initially apply to vehicles operating on a fixed route or within a fixed geographical area. It should be noted however that the proposed framework is intended to be scalable to other AV deployments.

This document has aimed to collate all potential requirements of an in-use monitoring framework and then assess and propose possible approaches to implement it. This report outlines the various elements of a monitoring framework and provides recommendations for how each element may be addressed in practice. Bringing all elements together a framework for in-use monitoring has been proposed. A high-level summary of the framework is given below:

Pre-deployment:

1. The In-Use Regulator sets out the minimum requirements for monitoring that the Manufacturers and Operators must ensure compliance against prior to deployment (as part of approval requirements). A set of minimum in-use monitoring requirements has been proposed in the Minimum Dataset Specification report (Chapman and Perren, 2021).
2. The Manufacturer develops their Vehicle Safety Case (VSC) for approval. In the VSC, the Manufacturer should demonstrate compliance with the minimum monitoring requirements and also specify additional monitoring required to ensure compliance with the safety arguments stated within the VSC (See Section 7.1.1)
3. The Manufacturer's Safety Management System (SMS) determines their ability to conduct in-use monitoring and develop their VSC robustly. In order to ensure compliance, the Approval Authority (possibly supported by the In-Use Regulator) should conduct an audit of the Manufacturers' SMS prior to approval and periodically during deployment. Key recommendations regarding SMS requirements for in-use monitoring and Audits are given throughout this report where relevant and collated in the following section (Section 12).
4. The vehicle safety and security monitoring capability is assessed at type approval as part of approval, the Manufacturer develops an operational manual that outlines safe use of the AV. It should also set out operational monitoring requirements specific to the AV. This document will be used by the AV operator to establish their monitoring responsibilities and define them within the deployment approval (and authorisation)

step, in a Deployment Safety Case (DSC). Key monitoring responsibilities proposed to be overseen by the Operator are outlined in Section 7.1.3.

In-Service

5. When in-use, the Manufacturer and operator enact their responsibilities for in-use safety monitoring. The primary mechanism for monitoring is through the identification and subsequent data collection of unsafe events. The Road Incident Taxonomy Report for this project (Reed, N., 2022) defines and classifies events within scope of the scheme. The Minimum Dataset Specification report (Chapman and Perren, 2021) outlines the minimum set of leading and lagging measures used to identify events of interest.
6. In addition to the minimum data set, there are a number of additional mechanisms for event detection to account for the limitations in the ability of vehicle's event-based data capture. These are outlined in Section 0.
7. Upon identification of an event of interest, it is necessary for the Manufacturer to investigate the event. Data recall requirements to enable investigation are outlined in Section 0. This involves identifying an actual risk event or an infraction occurred, classifying it (in line with the Road Incident Taxonomy Definitions (Reed, N., 2022)) and investigating it sufficiently to identify causal factors.
8. If the event was a severe event requiring immediate action (i.e. a collision involving injury), this should be immediately reported to the In-Use Regulator. It will also trigger the post-incident response to ensure police, regulator and independent investigating authority response is coordinated effectively as required. This is defined in the Post-Incident Response Framework report (C. Arnold, 2022) for this project. In addition to this, if the event show indicates that VSC/DSC has been invalidated, this should also be reported to the regulator immediately. Requirements for immediate reporting are outlined in Section 9.1.1.
9. For all other monitoring data, reports are expected to be made to the regulator periodically. This is to establish trends over time which could further indicate issues with the AV deployment and identify any actions required by the regulator to investigate further. The proposed requirements for periodic reporting and aggregated data analysis by the Manufacturer are provided in Section 9.1.4 and 9.1.5. The process for In-Use Regulator Analysis and monitoring of AV safety performance is provided in the Outcome Reporting report for this project (Reed *et al.*, 2022)
10. If the In-Use Regulator establishes non-compliance, they can take action against the Manufacturer or the operator (as required). For non-ADS issues, we propose this is handled by existing the Market Surveillance capacity of the DVSA. For ADS issues, the In-Use Regulator should establish the most appropriate course of action. Where the regulator was misled by the Manufacturer or operator (e.g. through misrepresentation or withholding of data), then criminal prosecution may be sought. For all other cases, the range of regulatory sanctions proposed by the Law Commissions (Law Commission & Scottish Law Commission, 2022) can be used. Considerations for the proportionate and fair application of sanctions are discussed in Section 0. The aim of the sanction is to initiate remedial or restorative action to

improve AV safety and maintain compliance with AV approval and authorisation requirements.

11. Where there is non-compliance, but as a result of a potential deficiency in approval and authorisation requirements, this should instigate a review of potential changes to the requirements to close the gap and improve public safety. The Change Control Process report for this project outlines a recommended method for enacting considered, positive changes to the AV safety assurance scheme (Perren, 2022).

Table of Contents

List of Figures and Tables

Tables

Figures

List of Abbreviations

ADS:	Automated Driving System
ADSE:	Automated Driving System Entity
AEVA:	Automated and Electric Vehicle Act 2018
ALKS:	Automatic Lane Keep Assistance
ASDE:	Authorised Self Driving Entitity
ATG:	Advanced Technologies Group (Uber)
AV:	Automated Vehicles
AVS:	Autonomous Vehicle Visualisation
BSI:	British Standards Institute
CAA:	Civil Aviation Authority
CCAV:	Centre for Connected and Autonomous Vehicles
CLOCS:	Construction Logistics and Community Safety
DDT:	Dynamic Driving Task
DfT:	Department for Transport
DMS:	Driver Monitoring Systems
DSC:	Deployment Safety Case
DSSAD:	Data Storage System for Automated Driving
DVSA:	Driver and Vehicle Standards Agency
EDR:	Event Data Recorder
FNOL:	First Notification of Loss
FORS:	Fleet Operator Recognition Scheme
FRAV:	Functional Requirements for Automated Vehicles (UNECE)
GB:	Great Britain
GPS:	Global Positioning System
GPSR:	General Product Safety Regulations
GVW:	Gross Vehicle Weight
HSE:	Health and Safety Executive
ISMR:	In-Service Monitoring and Reporting
ITU:	International Telecommunication Union
ITU-FGAI4AD:	ITU Focus Group on AI for Autonomous and Assisted Driving
LSAV:	Low-Speed Automated Vehicles

MRC:	Minimal Risk Condition
MRM:	Minimum Risk Manoeuvre
MSU:	Market Surveillance Unit
NATM:	New Assessment and Test Methods
NHTSA:	National Highway Traffic Safety Administration
NTSB:	National Transportation Safety Board
NUIC:	Non-User in Charge
ODD:	Operational Design Domain
OEDR:	Object Event Detection and Response
OEM:	Original Equipment Manufacturer
PAS:	Publicly Available Specification
PAYD:	Pay As You Drive
PHYD:	Pay How You Drive
PPH:	Pay Per Hour
PPM:	Pay Per Mile
SAE:	Society of Automotive Engineers
SDV:	Self-Driving Vehicle
SMLL:	Smart Mobility Living Lab: London (TRL)
SMS:	Safety Management System
ToR:	Terms of Reference
UNECE:	United Nations Economic Commission for Europe
VCA:	Vehicle Certification Agency
VMAD:	Validation Methods for Automated Driving (UNECE)
VRU:	Vulnerable Road User
VSC:	Vehicle Safety Case
WRRR:	Work Related Road Risk

2 Introduction

The Department for Transport (DfT) has commissioned this work to contribute to ensuring that Automated Vehicles (AVs) coming to market are both safe and secure. This is an opportunity not only to set national standards to facilitate deployment of AVs and attract home investment, but also to lead the way for regulatory approaches internationally.

In this project the scope is limited to slow moving (up to 20 mph), light-duty (up to 3,500 kg GVW) fully electric pods and shuttles without a driver (or, in the context of the Law Commissions' project, a user-in-charge) who is replaced by an Automated Driving System (ADS), that will have real-world applications for automated vehicles operating on public roads in a mixed-traffic environment. The vehicles may be passenger-carrying (maximum of 16 occupants, may include standing passengers) or goods-carrying. This will initially apply to vehicles operating on a fixed route or within a fixed geographical area. It should be noted however that the proposed framework is intended to be scalable to other AV deployments.

The pre-use approval and authorisation requirements of a possible safety assurance scheme will ultimately establish the expectations of the safety of the AV within its operating environment via establishing safety goals, performance requirements and behavioural rules. The primary role of in-use monitoring is to validate that the performance of AVs continues to meet, or exceed, the desired safety performance throughout its operational life.

Collecting data in the field to monitor automated vehicles and validate their performance is not a novel idea. This data is collected by AV developers to improve their system capabilities. However, there are no requirements currently to share this information, even though the information is vital to validate regulatory compliance, generating safety learnings for continuous improvement and ensuring public confidence.

In-use monitoring interfaces with insurance telematics, market surveillance, pre-deployment approval, fleet monitoring and ADS developer safety and quality management systems. In-use monitoring requirements are also being considered nationally by the Law Commissions and internationally by the UNECE working group Validation Methods for Automated Driving (VMAD) and the European Commission. This reports reviews how in-use monitoring of AVs, with a specific focus on Low-Speed Automated Vehicles (LSAV), relates to existing processes and requirements. Based on this, key elements of in-use monitoring are discussed, and recommendations are made for an in-use monitoring framework to support the deployment of AVs, that validates regulatory compliance, delivers safety learnings and provides public confidence.

3 Requirements of in-use monitoring

In-use monitoring is the collection, reporting and analysis of data regarding the safety of a vehicle(s) while in operation. The purpose of in-use monitoring is to deliver the safest possible transport solutions as quickly as possible to the public. To achieve this in-use monitoring must ensure that the initial safety assessment prior to market introduction is confirmed in the field, and to share learnings around safety events to allow the whole community to learn from operational feedback, fostering continuous improvement of both technology and regulation.

As part of the project brief, in-use monitoring of AVs needs to deliver:

- processes to identify events that are relevant to safety and collect data to understand AV behavioural safety performance;
- the ability to assess safety and allow regulatory intervention both proactively - prior to hazards arising, and reactively - following a hazard (e.g a collision);
- the ability to provide evidence as to whether the safety in the field aligns with what was stated during type approval;
- processes to allow developers and in-use regulators to proactively manage the risk posed by AVs; and
- methods for gathering data to highlight any regulatory gaps, generate training scenarios and evaluate the safety performance of AVs compared to human driven vehicles.

In addition to the requirements set out by this project, the in-use monitoring framework should take into account the in-service requirements that currently exist in the UK. In-service risk management of vehicles exists in different systems in the UK, involving different stakeholders. A summary of existing systems and processes for gathering operational feedback for safety management is given below. How these systems would interact with the UK in-use monitoring scheme is also discussed.

While relatively immature for AVs, operational feedback is a well-established practice in vehicle safety regimes in the UK and internationally. Some key regimes are also discussed to ensure that learning from current best practice are taken into account.

3.1 Law Commission Framework

In 2018, the Centre for Connected and Autonomous Vehicles (CCAV) asked the Law Commission of England and Wales and the Scottish Law Commission to examine options for regulating automated road vehicles. Chapter 6 of the final report published on the 25 January 2022 specifically addresses in-use safety assurance (Law Commission & Scottish Law Commission, 2022) and made the following recommendations;

> ***Recommendation 18**:*
>
> *The new Act should establish an in-use safety assurance scheme which gives an in- use regulator responsibilities to monitor the safety of authorised AVs and investigate infractions involving AVs, and powers to enforce its decisions.*

Recommendation 19.

The in-use regulator should be under a statutory obligation:

(1) to collect and analyse data to measure the safety of automated driving against the Secretary of State's published safety standard;

(2) to publish their findings; and

(3) to explore a range of possible measures to assess automated driving safety.

Recommendation 20.

The in-use regulator should be given powers to collect relevant data from ASDEs and NUIC Operators so as to allow the regulator to compare the safety of automated and conventional vehicles.

A number of topics relate specifically to requirements of the In-Use Safety Monitoring Framework.

3.1.1 How Safe is Safe Enough? As safe as the existing fleet?

The "how safe is safe enough" question for automated vehicles often considers the value of comparisons against existing human driven vehicles. Three standards for assessing automated vehicle safety were considered;

(a) as safe as a competent and careful human driver;

(b) as safe as a human driver who does not cause a fault accident;

(c) overall, safer than the average human driver.

No one standard received majority support from the Consultation respondents (Law Commission & Scottish Law Commission, 2021).

The lack of data for the existing human driven vehicles is often cited as a major limitation for making meaningful comparisons of automated vehicle performance.

However, this report highlights that those existing frameworks for collision and near miss reporting could provide and suitable datasets for making performance comparisons between existing humans driven licensed fleet Operators and those operating automated vehicles. In particular, the reporting requirements for the Fleet Operator Recognition Scheme (FORS), Construction Logistics and Community Safety (CLOCS) standard the expectations of TFL's Work Related Road Risk (WRRR) initiative.

This direct comparison between fleet operations should be more meaningful as it compares automated vehicle performance to a specific cohort of professional drivers which these services are looking to replace or augment.

The need to make such a comparison was recognised in the Law Commissions final report (Law Commission & Scottish Law Commission, 2022) while the Secretary of State is tasked with defining the overall level of safety;

Recommendation 6.

The new Act should require the Secretary of State for Transport to publish a safety standard against which the safety of automated driving can be measured. This should include a comparison with harm caused by human drivers in Great Britain.

3.1.2 Leading and Lagging Measures, Infractions & Sanctions

Consultation Paper 3 received the following responses in relation to safety measures, infractions and sanctions;

> *Respondents **widely supported the collection of data on AV performance for in-use monitoring** (Q18). This extended to both **leading measures (instances of bad driving which could have led to harm)** and **lagging measures (outcomes which led to actual harm)**. Many emphasised the importance of choosing appropriate measures, which would not necessarily mirror some of the classic leading measures used to indicate "bad driving" in humans. Instead new measures would be needed, like the **frequency of emergency manoeuvres** or **unstable lateral positioning within lane**. We understand that Government is considering these issues as part of the CAVPASS programme.*
>
> *In Chapter 11 we considered two challenges (Q22 & Q23). The first was how to deal with breaches of traffic rules. The second was how to learn from collisions so as to promote a safety culture. In both cases we proposed a move away from the current emphasis on the criminal prosecution of human drivers. Instead, we proposed that the **in-use safety assurance scheme** should **investigate breaches of traffic rules by AVs** driving themselves and **apply a flexible range of regulatory sanctions on ADSEs**. **There was broad agreement on all the policies we put forward in this chapter**.*

The Automated Driving System Entity (ADSE) term was introduced by the Law Commission, in Consultation Paper 3. It was superseded in the final report by Authorised Self-Driving Entity (ASDE) which is;

> "*the vehicle Manufacturer or software developer who puts an AV forward for authorisation as having self-driving features*".

In this report we refer to the more generic term of "manufacturer" when referring to the entity applying for self-driving authorisation (unless specifically referencing the Law Commissions work).

A comprehensive breakdown of the responses is included in 13, however, it is apparent that the in-use safety assurance scheme must be able to investigate;

- Near miss events (leading measures)
- Collisions events (lagging measures)
- **safety-related traffic infractions** (such as exceeding the speed limit; running red lights; or careless or dangerous driving);
- other t**raffic infractions**, including those subject to penalty charge notices;

While gathering sufficient information on each event such that the most appropriate sanction can be imposed including;

(1) informal and formal warnings;

(2) fines;

(3) redress orders;

(4) compliance orders;

(5) suspension of authorisation;

(6) withdrawal of authorisation; and

(7) recommendation of attendance at a restorative conference.

A more detailed discussion on the implications of these topics occurs in Section 0 and Section 11.

3.2 Automated and Electric Vehicle Act 2018

Through discussions with DfT, CCAV and the Law Commissions it was established that in-use monitoring should not only validate that a AV is driving safely, but also confirm that it is operating legally. While operating legally does mean driving in such a way to not cause traffic infractions, it also means that it is legally permitted to drive on the road in certain circumstances and situations. It is necessary then that in-use monitoring must confirm that it is operating within the scope of the approval and authorisation it was awarded. At the highest level, this means confirmation that the vehicle is "self-driving" and can continue to be registered as one.

The legal definition of self-driving in the UK is provided in the Automated and Electric Vehicles Act (AEVA) 2018. The Secretary of State must prepare, and keep up to date, a list of all motor vehicles that are in their opinion, able to safely and lawfully drive themselves. Under AEVA a vehicle is "driving itself" if it is operating in a mode in which it is not being controlled, and does not need to be monitored, by an individual An in-use monitoring scheme may be able to confirm that this condition is met throughout a vehicle's operation. To meet this requirement in-use monitoring must be able to detect instances where the AV required monitoring or control. This includes:

- Instances of remote monitoring or control outside of specified functionality.
- Instances of monitoring and control of the vehicle by the passenger to ensure safety (i.e. an emergency stop button).
- Interventions by other road users or the police to control the vehicle to maintain safety.

These requirements are taken forward into Section 0.

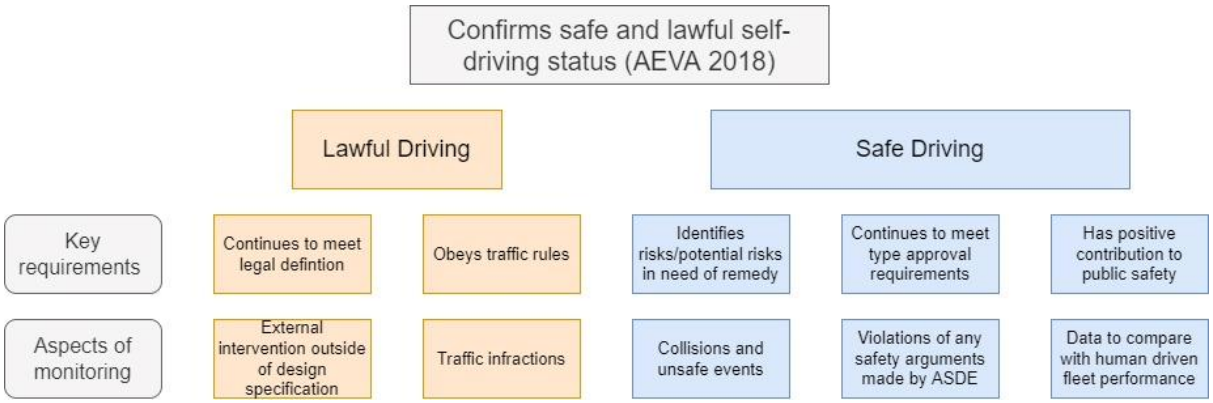


Figure 1: Hierarchy of monitoring requirements

Note: not all traffic rules are legal requirements, however, many define behaviours that constitute the competent and careful expectation against which careless and dangerous driving is assessed.

4 Public Expectation (The Molly Problem)

In October 2020, the International Telecommunication Union (ITU) Focus Group on AI for Autonomous and Assisted Driving (FG-AI4AD) conducted a public survey to establish the post-collision ethical expectations for automated vehicles (ITU-FGAI4AD, 2020).

To elicit public response a very simple situation described as The Molly Problem was posed as follows;

> *A young girl called Molly is crossing the road alone and is hit by unoccupied self-driving vehicle. There are no eye-witnesses. What should happen next?*

As series of specific questions were then posed which covered post-collision behavioural expectations, levels of explainability and relevance to near-miss collision events.

The preliminary survey results are shown in the table below. They represent the response from 296 individuals (70% male 25% female, 5% not specified), aged between 18-73 years old (mean 41yrs), living in rural, city, suburban and mainly urban environments. Three quarters of respondents were both willing and excited to travel in an automated vehicle and 95% held a driving licence.

Table 1: The Molly Problem Public Consultation Responses (11-20th October 2020)

Question	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 of the *speed* at point of the collision	98%	1%	1%
8	the software to recall *when* the collision risk was identified	93%	6%	1%
9	the software to recall *if* Molly was detected	96%	3%	1%
10	the software to recall *when* Molly was detected	96%	2%	2%
11	the software to recall *if* Molly was detected as a *human*	91%	6%	3%
12	the software to recall *when* Molly was detected as a *human*	90%	7%	3%
13	the software to recall *whether* mitigating action was taken	98%	1%	1%
14	the software to recall *when* mitigating action was taken	97%	2%	1%
15	the software to recall *what* mitigating action was taken	96%	3%	1%
16	similar recall abilities for *near-miss events*	88%	5%	7%

17	expect *driving* to be *prohibited* for software *without recall capability*	72%	15%	12%

The responses to Q1-4 indicate the public expectation that AVs will remain compliant with Section 170 (Duty of driver to stop, report accident and give information or documents) of the Road Traffic Act 1988.

The responses to Q5-15 indicate the public expectation that AV software is capable of providing an explanation for the road traffic situation, the level of risk awareness and the decisions taken for mitigating actions.

These explainability requirements align closely with the BSI proposal for Digital Commentary Driving as a technique for CAV safety benchmarking during real-world driving (BSI, 2021).

The Q16 response indicates the public expectation that explainability applies to near-miss events as well as collisions.

Finally the Q17 response indicates that without recall/explainability capability an AV should be prohibited from driving. This would be applicable if the recall capability could not be verified and validated during type approval (pre-deployment) or if the capability failed during in-service operations (post-deployment).

4.1 Fleet Monitoring

In addition to defining the ASDE the Law Commission recommended an additional legal entity; the Non-User In Charge (NUIC) Operator (Law Commission & Scottish Law Commission, 2022);

> *Every NUIC vehicle should be overseen by a licensed NUIC operator, with responsibilities for dealing with incidents and (in most cases) for insuring and maintaining the vehicle.*
>
> ***Recommendation 54.***
>
> *To obtain a NUIC operator licence, the applicant should submit a safety case, showing how safety will be assured. Among other things, the applicant's safety case should set out:*
>
> 1. *How oversight will be provided to vehicles, including suitable connectivity, equipment, staff training and rest breaks;*
> 2. *Incident management, including communication with passengers, road users and the emergency services, together with measures to remove vehicles causing an obstruction;*
> 3. *Systems, expertise and equipment to maintain vehicles, install updates and ensure cybersecurity;*
> 4. *Data management;*
> 5. *Whether safety relies on any element of remote driving, and (if so) how this will be done safely; and*

6. *Ways to learn from mistakes, including links with local authorities, highway authorities and the police.*

Where an ASDE and the NUIC operator are the same entity, the entity may submit a joint safety case covering both roles, to be assessed by the authorisation authority.

In other cases, the safety case should address the Manufacturer's written specifications for what must be done to ensure safe operation.

In the context of this report, we refer to the more generic term of Operator, rather than NUIC Operator, unless specifically referencing the work of the Law Commission.

The requirement for fleet Operators to report incidents and provide First Notification of Loss (FNOL) to insurers is in line with existing practice for human driven fleets.

In accordance with the Health and Safety at Work Act 1975 it's accepted that, for human driven vehicle fleets, the fleet operator, as an employer, has a duty of care for the safety of their employees (drivers) and others affected by their business activities (all other road users).

There are a number of frameworks designed to assist fleet Operators achieve these expectations including the voluntary Fleet Operator Recognition Scheme (FORS, 2021), TFL's Work Related Road Risk (WRRR) initiative and the Construction Logistics and Community Safety (CLOCS) Standard.

As an example of the similarities, the FORS Operators Requirement O3 (FORS, 2021) specifies the need to; *"document and investigate road traffic collisions, incidents and near-misses"; "determine the contributory and root causes of road traffic collisions, incidents and near-misses to prevent recurrence and minimise road risk".* FORS Operators shall demonstrate they have *"a policy and supporting procedures in place to record and investigate road traffic collisions, incidents and near-misses".*

The data collected from these in-use safety monitoring frameworks could be a good source of data for meeting Law Commissions expectation of automated driving performance comparison to existing humans driven fleets;

Recommendation 6.

The new Act should require the Secretary of State for Transport to publish a safety standard against which the safety of automated driving can be measured. This should include a comparison with harm caused by human drivers in Great Britain.

Recommendation 20.

The in-use regulator should be given powers to collect relevant data from Manufacturers and Operators so as to allow the regulator to compare the safety of automated and conventional vehicles.

Appendix C covers the expected data that should be captured and reported as part of the FORS Collision Manager process and the CLOCS Collision management and reporting requirements.

4.1.1 Collision Detection and Automated Vehicle Inspection

Section 170 of the Road Traffic Act 1988 defines the *"Duty of driver to stop, report accident and give information or documents".*

The section applies to mechanically propelled vehicles operated on a road or other public space in which an accident occurs that causes personal injury (other than the driver) or damage to another vehicles, animal (horse, cattle, ass, mule, sheep, pig, goat or dog) or property constructed on, fixed to, growing in or otherwise forming part of the road or place.

Collision Detection is a prerequisite for executing the "stop" condition required by Section 170. While Collision Avoidance is the objective for all drivers it's clear that Collision Detection and reporting act as a safety net for public safety.

To maintain the same safety standards for post-collision care automated vehicles would need to replicate the Collision Detection capabilities of existing human drivers. Technical standards for Collision Detection will be required and validation should occur during automated vehicle type approval. The regulatory scheme should consider extending the current functionality of the eCall emergency call system to include use in automated vehicles and the detection of vulnerable road user collision detection.

What happens if an automated vehicle is involved in a collision which is not immediately detected by on-board systems? To identify these missed collisions, it's proposed that the fleet operator has operational processes and procedures for inspection of the automated vehicle. For example, this may include a requirement to inspect the vehicle each time it returns to the licensed fleet operator depot or periodically within each 24hr period.

For small fleet sizes manual inspection may be feasible. For larger fleet sizes it may be more appropriate, and more accurate, to use automated inspection solutions which are able to detect bodywork dents and scratches, wheel rim damage and underbody scratches (DeGould, 2022). These automated solutions include the ability to record high-resolution images of the vehicle condition and can process up to three vehicles per minute.

If collision damage is detected, then data from the automated driving system should be stored and analysed to identify the time, location, and cause of the incident. Even if no collision damage is detected, the inspection event should be recorded and the associated images stored to provide evidence for any future claims that may arise.

4.2 Market Surveillance

Market surveillance is the process for ensuring that products and services conform to applicable laws and regulations following their market introduction.

The existing legal basis for monitoring product safety in the UK and the EU is primarily the General Product Safety Regulations (GPSR) 2005 and Regulation (EU) 2018/858 which recognises the need to introduce market surveillance to complement type approval requirements.

The GPSR puts general obligations on producers and distributors to ensure that products are safe. Regulation 36 of the GPSR also empowers authorities to conduct market surveillance of products.

In-use monitoring can be considered as complementary to existing market surveillance activities, although there are some key differences between traditional market surveillance and the Law Commission recommendation for in-use regulation.

In summer 2016, the Department for Transport established the Market Surveillance Unit (MSU) to check that vehicles on the UK market comply with type approval and emissions standards. The MSU is based within the Driver and Vehicle Standards Agency (DVSA) and works closely with the Vehicle Certification Agency (VCA) (DfT, 2022).

The Vehicle Market Surveillance Unit will carry out investigations and tests based on:

- annual programme of tests and inspections
- reports from industry
- reports from general public

The purpose of these investigations is to identify confirmed or potential safety related defects and raise these issues to the vehicle producer for them to investigate further and propose a rectification plan. A safety-related defect is a failure due to design and/or construction, which is likely to affect the safe operation of the product – and pose a significant risk to the driver, occupants and others. A safety defect can be of a physical component or software and could occur at any point in the life of the product. The definition of a safety defect is intentionally left broad in order to account for individual circumstances. In general, any vehicle that conforms with type approval requirements is assumed to be safe unless evidence shows otherwise. Any product issue that breaks this presumption of conformity would then be considered a safety defect. (DVSA, 2021)

In their final report, the Law Commissions noted the current market surveillance powers available to regulators and proposed that the in-use regulator be given additional powers and responsibilities. (Law Commission & Scottish Law Commission, 2022). There Consultation Paper 3 explored several options for how this may be implemented in practice:

1. Extend the existing market surveillance scheme to include safety defects arising from ADS performance
2. Supplant existing market surveillance with a new scheme specifically for AVs
3. Continue existing market surveillance for non-ADS elements and develop an enhanced scheme for ADS elements only.

The first option is not recommended by the Law Commissions. The current law is designed to deal with conventional vehicles. It puts considerable emphasis on the mechanical test to ensure that specific vehicle systems adhere to technical regulations. The annual programme of tests and inspection may apply to any products on the UK market. Products will generally be selected based on their actual or expected market share, with other products added to make sure a full range of Manufacturers is included. Some other products will also be added at random to make sure that the products tested cannot be predicted. This form of testing is unlikely to detect issues relating to AV behavioural performance when completing its dynamic driving task.

Conversely however there are consequences of creating a completely new in-use safety assurance scheme that supplants existing market surveillance entirely. The non-ADS elements

of the vehicle will still require assurance that they continue to conform with type approval requirements. The annual programme of tests and inspections will still be a useful tool to handle these type of safety defects. In-use monitoring may not detect, for example, an air bag defect until a sufficiently severe event has occurred that would normally trigger airbag deployment. Current market surveillance is also a more established process, and therefore more familiar and well understood by existing vehicle producers. This includes the process for developing rectification plans and working with existing regulatory sanctions.

When DVSA finds evidence of unsafe or illegal vehicles, products or parts it can:

- offer advice and guidance
- issue warnings
- issue recall notices
- issue civil penalties (fines); and/or
- prosecute the individual or business (DVSA, 2021)

The Law Commission proposes that AVs be subject to a broad range of different regulatory sanctions, they are:

- informal and formal warnings;
- fines;
- redress orders
- compliance orders;
- suspension of authorisation;
- withdrawal of authorisation; and
- recommendation of attendance at a restorative conference. (Law Commission & Scottish Law Commission, 2022)

These sanctions are discussed in further detail in Section 0, however for Low Speed Automated Vehicles it may be problematic to have two sets of sanctions that could potentially apply to AVs. Having the single set of sanctions proposed by the Law commissions that would apply to AVs would require that the historical market surveillance sanctions not apply at all. This would likely increase the burden on Manufacturers of components or parts that are used in both AVs and conventional vehicles as they may have to undergo both schemes and face different sanctions for the same defect.

For this reason, it expected that retaining the existing market surveillance processes for non-ADS safety issues would be desirable, and the in-use monitoring scheme discussed in this document should apply to safety issues with the ADS only. However, the presence of the enhanced in-use monitoring scheme for ADS issues should also act as a data source to existing market surveillance frameworks. As such safety events that are identified through the scheme are as a result of a defect with the non-ADS elements would be notified to the MSU and would be subject to existing market surveillance processes.

4.3 Telematics based Insurance

Telematics based insurance provides a useful reference when considering the value in-use monitoring can have for assuring automated vehicle safety during operation.

In-use monitoring of human driving behaviour using onboard telematics systems was first introduced as a commercial insurance solution in 2007 (Automotive World, 2022). The market solutions have continued to evolve, as does has technology used to capture the driving performance data.

Telematics insurance, or Usage Based Insurance (UBI) is often spilt into two different models; Pay As You Drive (PAYD) and Pay How You Drive (PHYD) (Amodo, 2022).

PAYD uses distance (Pay Per Mile PPM) or time (Pay Per Hour PPH) to estimate exposure to dynamic collision risk.

PHYD uses additional driving behaviour data to create a unique driver Safety Core - which may be considered as a combination leading metrics. Typical source data collected can be related to;

- acceleration
- braking
- speeding
- cornering
- lane changing manoeuvrers

Additional PHYD data points can include;

- driving area (e.g. urban or rural),
- trip duration,
- road types,
- route choices,
- possible collision data,
- airbag deployments,
- trip location and time.

There are many similarities with the data points above and those that might be used to describe an automated vehicles operational design domain (ODD) and it's execution of the dynamic driving task (DDT). However, insights into the DDT performance are often limited to vehicle dynamics (speeds, accelerations, braking, cornering g-forces) and do not include the context of the road traffic situation in which these events occur e.g. hard braking to avoid a collision with a child that has run into the road at the last minute may be treated equally to a distracted driver braking late for a vehicle that stops in the road ahead.

A complete DDT performance assessment should include an evaluation of the Driver's Object and Event Detection and Response (OEDR) capability (which requires capture of the 3D world

in which the action occurs). This would apply equally to both human driven and automated vehicles.

Telematics based insurance has typically been targeted at young and inexperienced drivers or those classed as highest premium drivers (those returning after a ban, older drivers with a significant claims history, younger drivers with high-powered cars) (LexisNexis, 2018). In 2018 this was estimated to represent 800,000 policies (approx. 2% of the total insured UK vehicles).

The main motivation for consumer adoption of UBI has been reduced insurance premiums, while the main barrier has been the perceived loss of privacy.

While adoption of UBI for personally owned vehicles remains low the situation is different for Fleet Operators. In 2018, the BT Fleet Survey report estimated that; "*51% of all major fleets in the UK are already using telematics, including 74% of fleets containing more than 100 vehicles*" (Natwest, 2020).

Insurance is not the only reason fleet Operators adopt telematics. Other reasons include; *reduce unnecessary mileage (50%), driver monitoring (sleep detection, distraction) and driver behaviour (acceleration and braking) (45%), improve fuel efficiency (42%) and provide evidence to dispute legal and insurance claims when there are accidents or other problems (39%).*

According to GlobalData's *2021 UK SME Insurance Survey* (Life Insurance International, 2021) "*33% of UK SMEs with commercial vehicle insurance have some form of usage-based insurance (UBI), and a further 18.6% are considering purchasing this type of product*".

The broad adoption of telematics-based solutions within human driven fleet operations may be expected to continue with automated vehicles and is likely to be extended as increased data from onboard vehicle sensors and software becomes accessible.

4.3.1 Example – Driver Safety Score Metrics

A number of insurers create individual driver safety scores based upon data collected using in-use monitoring.

For example, Tesla's Driver Safety Score metrics are used for it's own car insurance scheme as well as being used as a selection criteria for drivers participating in Full Self-Driving Beta testing (Tesla, 2022).

The Safety Score is;

> *…an assessment of your driving behavior based on five metrics called Safety Factors. These are combined to estimate the likelihood that your driving could result in a future collision. We combine your daily Safety Scores (up to 30 days) to calculate the aggregated Safety Score, displayed on the main 'Safety Score' screen of the Tesla app.*
>
> *The Safety Score Beta is intended to provide drivers transparency and feedback of their driving behaviors. The Safety Score is a value between 0 and 100, where a higher score indicates safer driving. Most drivers are expected to have a Safety Score of 80 or above.*

The five Safety Factors, which, are measured directly using the vehicles sensors and Autopilot software are listed below;

1. ***Forward Collision Warnings per 1,000 Miles*** *- events where a possible collision due to an object in front of the vehicle is considered likely*
2. ***Hard Braking*** *- a backward acceleration in excess of 0.3g*
3. ***Aggressive Turning*** *- left/right acceleration in excess of 0.4g*
4. ***Unsafe Following*** *- This measurement is called "headway." Unsafe following is the proportion of time where your vehicle's headway is less than 1.0 seconds relative to the time that your vehicle's headway is less than 3.0 seconds.*
5. ***Forced Autopilot Disengagement*** *- vehicle has determined that you have removed your hands from the steering wheel and have become inattentive.*

Measures such as "Forward Collision Warnings" and "Unsafe Following" both related to the drivers OEDR capability and require knowledge of the external 3D world in which the DDT is executed to be derived from vehicle sensors.

The collected dataset of human driving behaviour also provides a baseline for comparing the performance of vehicle operation when the Full Self-Driving system is engaged[1]. The quarterly Tesla Vehicle Safety Report (Tesla, 2022) provides comparison of safety performance, one example from Q2 2021;

> *In the 2nd quarter, we recorded one crash for every 4.41 million miles driven in which drivers were using Autopilot technology (Autosteer and active safety features). For drivers who were not using Autopilot technology (no Autosteer and active safety features), we recorded one crash for every 1.2 million miles driven. By comparison, NHTSA's most recent data shows that in the United States there is an automobile crash every 484,000 miles.*

Lagging measure comparisons, such as these, should be treated with caution as it's highly likely that Autopilot is engaged more often on US freeways where it's much easier to accumulate a high number of incident free miles compared to more complex urban or rural roads. This caution serves as an important reminder that safety data requires context and that ultimately safety assurance is likely to be based upon the specific context of each safety relevant event.

While valuable these simple metrics do not provide a comprehensive evaluation of OEDR performance. For example, they do not include measures such as "safe passing" distances and speeds appropriate for vulnerable road user (VRU) interaction or 'near-miss" metrics such as post-encroachment times (PET).

However, Telematics based insurance does provide an indication of how driver behaviour may be assessed in the future – for both humans and automated drivers.

4.3.2 Example - Dash Cam use for First Notice of Loss & Collision Reconstruction

Another example of in-use monitoring for human driven fleets that has some relevance to automated driving is the use of aftermarket dash cams.

[1] It should be noted that while Tesla's Full Self-driving system is operating the human driver retains full responsibility for supervision and intervention in the execution of the dynamic driving task.

Aftermarket dash cams have become increasingly popular way of recording video to be used as evidence in collision or conflict events. Dash cams offer fully integrated devices that can combine recording of video, audio, gyroscopes, accelerometers and GPS.

Dash cam data that is recorded on-board the vehicle can also be transmitted wirelessly, using WIFI and Bluetooth interfaces, to enable integration with mobile phone applications. Higher end devices may also include integrated cellular communications for real-time remote monitoring and tracking.

The use of AI algorithms within these dash cam solutions provides a useful indication on improvements to First Notification of Loss (FNOL) reporting that might be expected as standard for automated vehicles.

For example, in 2021 dash cam Manufacturer Nexar introduced an AI-based First Notification of Loss (FNOL) and collision reconstruction solution (Nexar, 2021);

> ***AI-Enabled First Notice of Loss -*** *…uses a combination of accident detection algorithms and artificial intelligence to detect and report accidents to insurers within minutes of an event occurring. By utilizing ~5 years of data …to train and deploy robust detection models that can detect an incident as small as, and unfortunately as common as, a fender bender.*
>
> ***Collision reconstruction -*** *Collision Insights utilizes sensor data and camera video to reconstruct accidents …a detailed reconstruction, …provides drivers and claim adjusters much needed context and a source-of-truth to resolve claims quickly and efficiently.*

As well as static document FNOL reporting it is also possible to visualise collision event data within a cloud hosted interactive dashboard within minutes of the event[2].

In the context of in-use regulation these innovative FNOL solutions are a good example of the tools that may be required for in-service monitoring and reporting which could be adopted by the Manufacturer, operator or regulator.

The source data required should be readily available from the automated driving system. In fact, while dash cams are only able to provide information on frontal impacts and automated driving system should be able to provide full 360 degree data.

The use of similar cloud hosted analysis would also enable the in-use regulator to meet the scale and demand for investigation of near-miss, safety critical infractions or collision events automated vehicles expected by the Law Commissions in their final report;

> ***Recommendation 21.***
>
> *The in-use regulator should be under a statutory obligation:*
>
> *(1) to investigate traffic infractions referred to it; and*
>
> *(2) if the infraction has been caused by the ADS feature/s, apply a flexible range of regulatory sanctions.*

[2] https://youtu.be/SBp8YsAaBFk

"Traffic infraction" refers to an action (or inaction) which forms part of the dynamic driving task and which (if conducted by a human driver) would make the human driver liable for a criminal offence or civil penalty.

Aggregated and anonymised data from such reports could be integrated with or used to augment existing tools, such as the CrashMap website, which provide fleet Operators with insights into the collision risks within a given operational location[3]. For example, the image below centred around the Smart Mobility Living Lab in London shows the location of 124 incidents that occurred just within 2020.

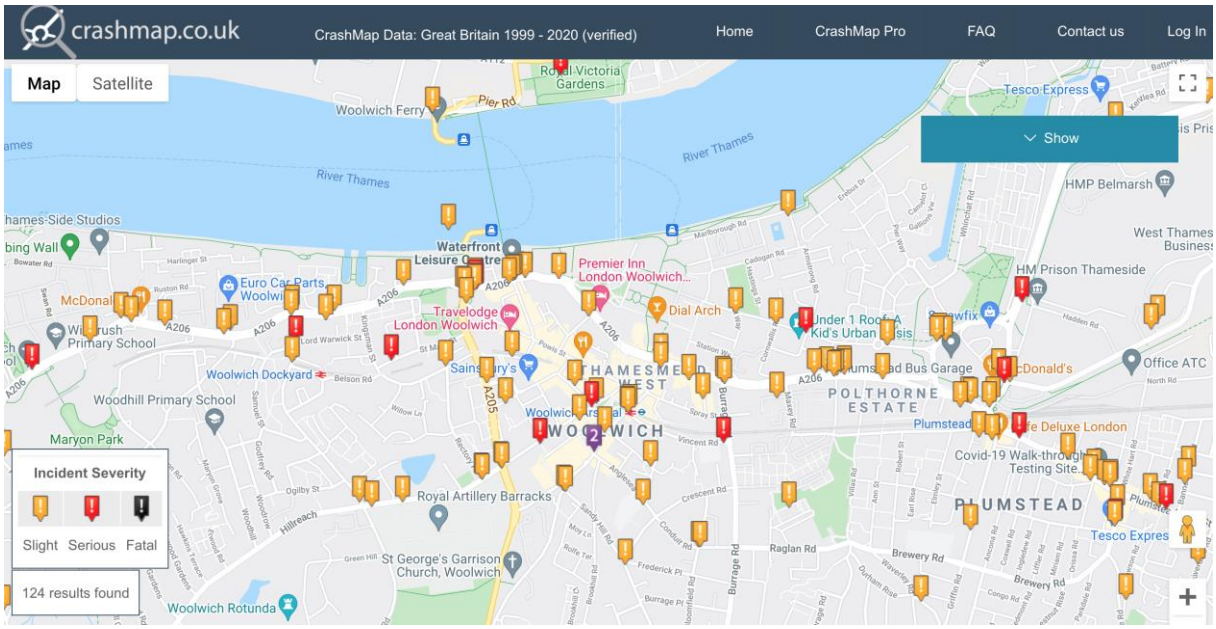


Figure 2: Collision Data for areas surrounding Smart Mobility Living Lab (SMLL) London

The interface allows selection of location and filtering based upon; year (1999-2020); casualty type (Pedal Cycle, child, motorcycle, pedestrian); vehicle type (pedal cycle, motorcycle, car, goods vehicle, bus, young driver). Full incident details can be retrieved by selecting the Incident Severity !"!" pins rated as slight, serious and fatal.

[3] https://www.crashmap.co.uk/Search

Figure 3: Collision heatmap for area surrounding Smart Mobility Living Lab (SMLL) London (Image copyright: https://www.plumplot.co.uk/uk-car-crashes-heat-map.html)

Heatmaps are an alternative visualisations of the same government Road Safety Data (DfT, 2021) which can provide additional insights into the collision hot spots which may indicate contributary factors from the road environment[4].

It would be beneficial for the in-use safety assurance regulator to have access to similar visualisation tools with the ability to compare human driven fleet performance against automated vehicle fleet performance. Whilst also being able to look at performance variations between different automated fleet Operators. Data could include collisions but also a more granular list of non-injury reported collisions, near-miss events as well as safety and non-safety relevant traffic infractions.

4.4 Automated Vehicle Operations – Supervision vs Monitoring

SAE J3016 divides driving into three types of driver effort; strategic, tactical and operational.

> *Strategic effort involves trip planning, such as deciding whether, when and where to go, how to travel, best routes to take, etc.*
>
> *Tactical effort involves maneuvering the vehicle in traffic during a trip, including deciding whether and when to overtake another vehicle or change lanes, selecting an appropriate speed, checking mirrors, etc.*
>
> *Operational effort involves split-second reactions that can be considered pre-cognitive or innate, such as making micro- corrections to steering, braking and accelerating to maintain lane position in traffic or to avoid a sudden obstacle or hazardous event in the vehicle's pathway.* (SAE, 2021)

[4] https://www.plumplot.co.uk/uk-car-crashes-heat-map.html

Tactical and Operational functions are defined as those required to “operate a vehicle in road traffic” within J3016;

1. Lateral vehicle motion control via steering (operational);
2. Longitudinal vehicle motion control via acceleration and deceleration (operational);
3. Monitoring the driving environment via object and event detection, recognition, classification, and response preparation (operational and tactical)
4. Object and event response execution (operational and tactical);
5. Maneuver planning (tactical); and
6. Enhancing conspicuity via lighting, signaling and gesturing, etc. (tactical).

Tactical and Operational functions are those required to execute the Dynamic Driving Task (DDT). In order to be listed as self-driving an automated vehicle must be able to execute the DDT without monitoring or control by a human driver.

For clarity the use of the term “monitoring and control” is a threshold test required to establish if a vehicle is self-driving e.g. if the vehicle requires a human to intervene in the DDT to ensure it’s safety it is not self-driving. “Monitoring and Control” influences the decision making and behaviour of the vehicle.

The term “in-use monitoring”, as used in this report, has a very different meaning. It is used specifically to refer to post-hoc monitoring of behaviour e.g. after the decision making and actions to control the vehicle has occurred.

As a function, in-use monitoring has a similar function to a driving test examiner, it only monitors whether the driver’s outcomes were safe. That means it can be used to independently monitor the DDT as well as the DDT subtask of Object and Event Detection and Response (OEDR).

For a Fleet Operator, responsible it may be better to use the term supervision to avoid confusion. Supervision, or remote oversight, would normally occur at the “strategic layer” focused on safe and efficient operation of the entire service but not the DDT.

Fleet operator supervision may make use of traffic updated maps combined with real-time location data from vehicles is essential for planning the shortest, fastest, cheapest, most efficient, or safest routes. This real-time location data and map interface is also central to the user experience of booking journeys or checking on their progress.

Fleet operator supervision is likely to be common solution for both human driven vehicles and automated vehicles within the same fleet.

The Law Commissions focused specifically upon the in-use regulators responsibility in-relation to safety. However, it is foreseeable that the in-use regulator would also be best placed to gather data required for assessing the environmental impact of AVs and their impact on road network traffic efficiency.

4.4.1 Human Driven Vehicles – examples of in-use monitoring

As briefly mentioned before the concept of real-time in-use monitoring of drivers is not new. These existing human driver monitoring approaches are well known to the public and will set their expectations for in-use monitoring of automated vehicle driving. For example, if a human driver is required to have their execution of the DDT monitored then it is likely that they will expect the same is true for automated vehicle.

For human driven vehicles Tactical and Operational in-use monitoring currently takes place in the form of telematics-based insurance schemes (see Section 4.3).

Safety assessment of Tactical and Operational driving performance is usually based upon leading metrics, which act as indicators of future collision risk.

For human driven fleets ego vehicle centric data provides insights into operational performance (e.g. throttle, brake, steering and resultant accelerations) while dash cam solutions provide the external situational context for these actions. Both the ego vehicle and external situational context are required to assess tactical performance of the object and event detection and response (OEDR) task).

Additionally, the in-use monitoring of human driven fleets often includes interior dash cams, or Driver Monitoring Systems (DMS) which monitor for drowsiness, attention, gaze and distraction e.g. factors the imply the driver is not completing the OEDR task with the level of safety expected from a competent and careful driver.

It is estimated that almost 40% of the 35 000 fatal car accidents in Europe each year can be attributed to an inattentive driver (Smart eye, 2020;Aptiv, 2021).

It's important to note that Driver Monitoring Systems do not directly monitor OEDR performance or the Tactical or Operational functions. Instead, they are designed to provide some level of "*explainability*" when performance of these tasks falls below expected standards. For example, in the event of a collision they can be used as evidence that the driver was inattentive at the time.

4.4.2 Automated Vehicle – 3D world model for OEDR monitoring

Tactical and Operational in-use monitoring of automated vehicles can also provide "*explainability*" insights required for safety assurance of the automated driving software.

While DMS act as a proxy for a human driver's "state of mind" during execution of the OEDR and DDT, for automated vehicles the software driver's "state of mind" is directly accessible through digital data.

With access to data from the automated driving software, it is feasible for in-use monitoring to assess the performance of the object and event detection and response (OEDR) task which is required for assuring safe execution of the Dynamic Driving Task (DDT).

From an explainability perspective, direct OEDR monitoring can be used to provide assurance that the software is correctly identifying hazardous situations and planning the most appropriate vehicle response to mitigate the risk. The level explainability is dependent upon the data that is recorded and the ease of it's interpretability. The data topic is addressed within the BSI Digital Commentary Driving work and that of the ITU FG-AI4AD in which

interpretability is linked to answering the very simply questions posed by The Molly Problem discussed at the beginning of Section 3.

Directly monitoring the OEDR and recording key events is also fundamental to a no-blame safety culture of continual learning and is essential for capturing novel scenarios to further improve to the verification and validation process as well as improve ADS safety performance. This "*scenario generation*" based upon real-world experience provides data, to support the Work Package 3 activity in the creation of new scenarios and identification of new behavioural rules. Th process also aligns with UNECE WP29 VMAD work described in the New Assessment and Test Methods (NATM) document. The work of that group is discussed below in Section 4.5.

Capturing scenarios and assessing the driving behaviour of the automated vehicle both require access to data that represents the circumstances and situations present during DDT operation as well as the software's execution of the ODER task.

The minimum data requirement is that the ADS share the digital world model representation it used for decision making and path planning. Described simply this is the information about where the ego vehicle is located relative to the road infrastructure and where other road users are located relative to the ego vehicle. It also includes detected events, such as traffic light status and predictions of risks for how the spatial relationships change over time.

It's worth noting that this type of data is already being used to build public trust in the OEDR capabilities of automated driving systems.

By extracting this OEDR data in real-time from the ADS developers are able to create 3D visualisations of the ego vehicle location in the road, the surrounding objects and the status of traffic lights[5][6].

The same type of 3D tools are used by developers for offline analysis of datasets collected during on-road operations. Open-source initiatives, such as the Autonomous Vehicle Visualisation (AVS), aim to harmonise the tooling enabling technology companies, research institutions, original equipment Manufacturers (OEMs), and start-ups.

In their research paper (Chen *et al.*, 2019;AVS, 2022) describe AVS as;

> *…a new standard for describing and visualizing autonomous vehicle perception, motion, and planning data, offering a powerful web-based toolkit to build applications for exploring, interacting and, most critically, making important development decisions with that data.*

[5] Aurora Tech 3D Visualisation https://aurora.tech/aurora-driver

[6] Waymo 3D visualisations; https://blog.waymo.com/2019/09/driven-by-waymo-designed-with-trust.html

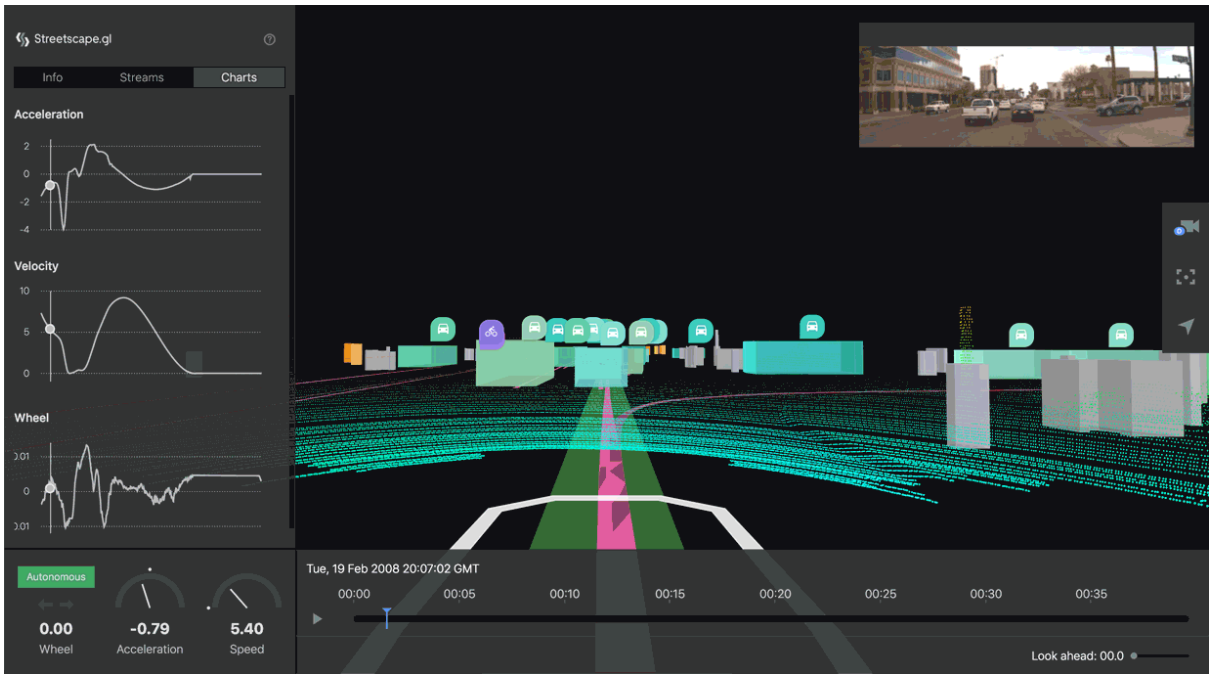


Figure 4: Uber ATG's Autonomous Vehicle Visualisation (AVS) showing real-world performance (Chen, Lisee, Wojtaszek, & Gupta, 2019)

As automated vehicles move from development to deployment tools such as AVS will become of increasing value to in-use safety assurance regulators, insurers and collision investigators.

4.4.3 Automated Vehicle – 3D World Model for remote assistance

As further indication as to the value in capturing 3D world model information and it's practical availability from the ADS software it's useful to consider their existing application in AV remote assistance solutions.

When an AV encounters an edge-case not covered by the ODD it may stop and request remote assistance from a human operator. In-use monitoring should cover these remote interventions given their safety critical nature in influencing the DDT.

The purpose of these "tele-assist" solutions is to enable a human in a remote control centre to visualise the 3D world in which the AV is driving, evaluate the quality of the OEDR performed and provide guidance on DDT execution.

Aurora describe their teleassist solution for the Aurora Driver in the following way;

> *We're not giving teleassist specialists direct access to the vehicle's steering, brake, or throttle. Instead, we're building an API layer that allows them to collaborate with — rather than replace — the Aurora Driver*[7]*.*

Teleassist solutions focus upon both the Strategic and Tactical driving tasks, while Operational Control remains the responsibility of the self-driving software.

[7] https://aurora.tech/blog/teleassist-how-humans-collaborate-with-the-aurora

Both 3D World Model visualisations and direct camera feeds may be used to provide the remote teleassist specialist with the context that is required to assist the automated vehicle complete the OEDR task.

Capturing the 3D world model data exchanged between the AV and the remote operator would provide valuable in-use monitoring insights into their safety and their frequency of use.

4.4.4 Automated Vehicle – Continual Learning Frameworks

In-use monitoring is a key process in creating a process and framework to support the no-blame safety culture of continual learning. The objective is that lessons learned from real-world driving experiences provide feedback to improve DDT performance.

For human drivers continually learning occurs instantly during the execution of the dynamic driving task (DDT). Another road user sounding their horn can often be a good trigger that indicated the previous action had a negative safety outcome. The feedback and learning are immediate.

In contrast automated driving software learning is not an online process occurring during the DDT. In fact, “Machine Learning”, contrary to the public misconception, is an offline line process requiring a dedicated compute infrastructure. When operating onboard the vehicle in much more constrained compute environments the software is just making inferences and executing instructions during the DDT– there are no learnings.

In-use monitoring is essential to capture these real-world learnings by evaluating the DDT performance and using safety triggers to capture data required for offline learning. These triggers act like the virtual horns of other drivers.

In-use monitoring must be able to capture safety critical situations, events and incidents in which the AV is involved and may have caused. It’s essential that the data captured provides sufficient evidence for investigation of collisions, near-misses and infractions.

For example, if during real-world operation the AV runs a stop light, passes too close to a cyclist or is involved in a collision firstly these events need to be detected and secondly the data collect must enable the continual learning required to improve the software.

The in-use monitoring process must define the safety critical events that should be captured and the specify the evidence required to be recorded for investigating the event.

The data, metrics and thresholds required for detecting events must be defined. So too does the data required as evidence of OEDR performance and incident reconstruction which are the foundation to continual learning.

On-road trials are being used by many AV developers to gather real-world data to feed into their own proprietary continual learning frameworks. Often this involves equipping test vehicles with dedicated data logging capabilities which can record all data and all times. The search for relevant safety critical events can then be conducted as an offline process.

Data recording and offline learning often has two distinct purposes: improving perception or improving prediction/planning. The data required for “perception” depends upon the proprietary set of sensors used by each individual AV. The data required for

"prediction/planning" a more abstract representation of the 3D world and can be considered common across all AVs.

The 3D world model data can be used for generating new scenarios for testing prediction and planning, as well as incident investigation. This 3D world model data is not just common between different AVs but also between AVs and the smart roadside infrastructure through which they driver and may communicate with using V2X.

Storing raw data for lidar, radar and camera sensors required for "perception" training is orders of magnitudes larger than the minimal object list data required for "prediction/planning

The proprietary nature of the sensors data for each AV makes it difficult for an in-use regulator to create a technical specification for data recording. While the storage requirements make it impractical for an in-use regulatory to require continual sensor data recording.

However, for the in-use regulator, it would be possible to define a common technical model for the 3D world model data. The data requires significantly less data storage, recorded continually and used as input to real-time safety metrics that generate triggers for event based recording of sensors data.

As an example of the relevance of this world model data consider the following inputs to Toyota Woven Planet's SafetyNet" ML-based planner (*input data is encoded in an ego-centric frame of reference where the self-driving vehicle (SDV) is always at a fixed location relative to a frame*) (Matt Vitelli, 2022);

> *1) SDV: the current and past poses of the SDV and its size.*
>
> *2) Agents: the current and past poses of perceived agents, their sizes, and object type (e.g. vehicle, pedestrian, cyclist) produced by the SDV's perception system.*
>
> *3) Static map elements: road network from High Definition (HD) maps including lanes, cross-walks, stop lines, localized using the SDV's localization system.*
>
> *4) Dynamic map elements: traffic light states, and static obstacles detected by the perception system (e.g. construction zones).*
>
> *5) Route: the intended global route that the SDV should follow.*

The list above helps set reasonable expectations of the 3D world model data in-use monitoring could access from a self-driving software system.

4.4.5 Automated Vehicle – Offline enhancement of captured real-world data

As described above the amount of processing power available onboard the vehicle is restricted in comparison to cloud data centre scale compute infrastructure.

Th other major on-board data processing constraint is that time moving forward. Analysing datasets offline removes this constraint and being able to run data backwards in time can significantly improve the quality of 3D world models that can be generated from the same source data.

For example, HAV software developer Motional states the following in relation to their auto-labelling offline perception system (Caesar, Holger, 2021);

> *…the "online perception" system employed by the AV to detect the traffic participants (vehicles, cyclists, pedestrians, traffic cones, etc.) in its environment. However, online perception systems are heavily constrained by multiple factors. First, the primary role of the AV stack is to detect every object in real time in a matter of milliseconds. Second, the computational power and memory of onboard computer systems are limited by cost and energy consumption. The computer system should cost a fraction of what the car itself costs and should not drastically reduce the battery time of an electric vehicle.*

For an in-use safety assurance regulator, or collision investigator, it may be important to have access to data generated using this "offline perception system" approach. The "offline perception system" output get's closer to providing a ground truth for comparison with the "online perception system" output. It may also enable clearer detection of the complex scenarios such as;

> *…merges, lane changes, protected or unprotected left or right turns, interaction with cyclists, interaction with pedestrians at crosswalks or elsewhere, interactions with close proximity or high acceleration, double parked vehicles, stop controlled intersections and driving in construction zones.*

4.4.6 Automated Vehicle – In-use monitoring metrics used by developers

To aid the identification of in-use monitoring safety metrics it's useful to consider how developers are assessing the safety performance of their models within simulation.

For example, Motional split the metrics used into two separate groups. Common metrics which are applicable to the entire DDT and Scenario-based metrics specific to a single circumstance/situation;

Common metrics.

- *Traffic rule violation is used to measure compliance with common traffic rules. We compute the rate of collisions with other agents, rate of off-road trajectories, the time gap to lead agents, time to collision and the relative velocity while passing an agents as a function of the passing distance.*
- *Human driving similarity is used to quantify a manoeuvre satisfaction in comparison to a human, e.g. longitudinal velocity error, longitudinal stop position error and lateral position error. In addition, the resulting jerk/acceleration is compared to the human-level jerk/acceleration.*
- *Vehicle dynamics quantify rider comfort and feasibility of a trajectory. Rider comfort is measured by jerk, acceleration, steering rate and vehicle oscillation. Feasibility is measured by violation of predefined limits of the same criteria.*
- *Goal achievement measures the route progress towards a goal waypoint on the map using L2 distance (an internal ML model measure).*

Scenario-based metrics.

- *For lane change, time to collision and time gap to lead/rear agent on the target lane is measured and scored.*

- *For pedestrian/cyclist interaction, we quantify the passing relative velocity while differentiating their location. Furthermore, we compare the agreement between decisions made by a planner and human for crosswalks and unprotected turns (right of way).*

Toyota Woven Planet adopt a different set of performance evaluation metrics based on the following binary events;

1) Collisions: the simulated SDV is <5cm from the road boundaries, static obstacles, or agents.

2) Close-calls: the simulated SDV has no collision, but either gets within 25cm of another agent, has a time- to-collision <1.5s, or has a time headway to another agent <1s.

3) Discomfort braking: the simulated SDV's jerk drops below $-5m/s^3$.

4) Passiveness: the simulated SDV travels slower than its behaviour in the dataset by < −5m/s and it is spatially behind its dataset position.

5) Off road events: the simulated SDV deviates from the dataset route centre line by >10m.

It will be important for the in-use safety regulator to understand the types of source data, metrics and thresholds being used by developers for safety assessment and establish the most appropriate balance between offline and online evaluations.

It would be reasonable to expect that once the in-use monitoring metrics have been defined by the regulator these will also be used by developers during the development process.

4.5 UNECE Validation Methods for Automated Driving (VMAD) – New Assessment/Test Methods (NATM)

During the 178th session of the United Nations Economic Commission for Europe (UNECE)'s World Forum for Harmonization of Vehicle Regulations (WP.29), the Framework document on automated/autonomous vehicles was adopted and the Terms of Reference (ToR) for the Informal Working Group on Validation Methods for Automated Driving (VMAD) were developed (WP.29, 2019). The ToR outlines that VMAD's mandate is to develop New Assessment/Test Methods to validate the safety of automated systems based on a multi-pillar approach including audit, simulation/virtual testing, test-track, and real-world testing, and In-Service Monitoring and Reporting (ISMR) (VMAD SG3-14-14, 2021).

ISMR requires Manufacturers to collect, analyse, and report information relevant to the safety of their ADS vehicles operation in the field. The three main purposes of in-service monitoring and reporting is to use retrospective analysis of data from Manufacturers and other relevant sources to:

- demonstrate that the initial safety assessment (residual risk) in the audit phase before the market introduction is confirmed in the field overtime ("safety confirmation").
- to fuel the common scenario database with important new scenarios that may happen with automated vehicles in the field ("scenario generation") and

- to derive safety recommendations for the whole community by sharing learnings derived from key safety accidents/ incidents to allow the whole community to learn from operational feedback, fostering continuous improvement of both technology and legislation ("safety recommendations").

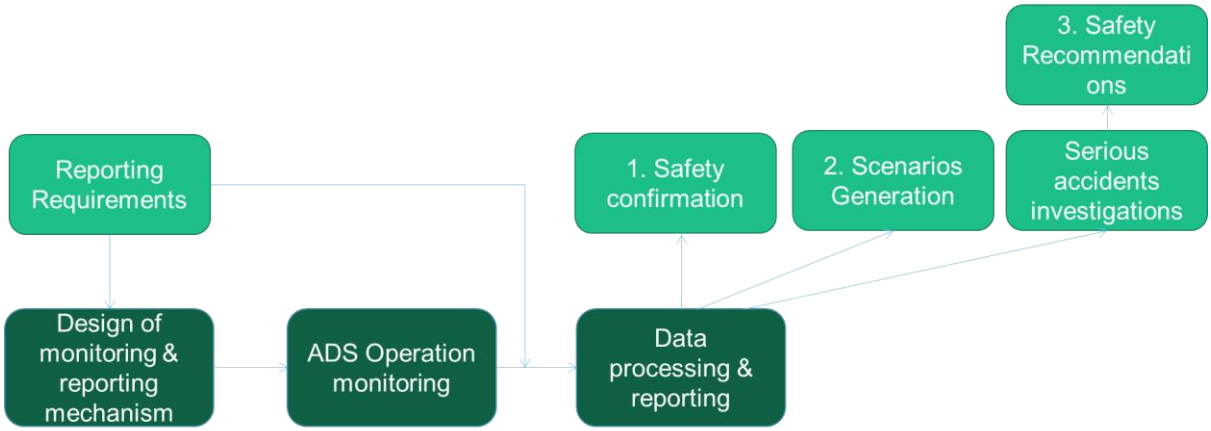


Figure 5: Draft in-use monitoring framework proposed by VMAD

ISMR applies to events occurrences which endanger or which, if not corrected, would endanger a vehicle, its occupants or any other person, and more generally, to all occurrences relevant to the safety performance of the ADS.

At the time of writing, the VMAD group are still developing the proposals for ISMR. The most recently developed proposals are summarised below to consider what learnings can be taken from the UNECE's approach and applied to this in-use monitoring framework.

4.5.1 In-service monitoring and reporting

For in-service monitoring, the Manufacturer should set up a monitoring program aimed at collecting and analysing vehicle data and other sources data to get evidence on the in-service safety performance of the ADS. This is supported by the Audit Pillar where Safety Management System Requirements are being set. The focus of the audit pillar is to ensure acceptable monitoring systems are in place. The focus of the ISMR pillar is to set data capture requirements.

The sole purpose of ISMR is the prevention of accidents and incidents and not to attribute blame or liability. It is not stated at this stage how this would interact with sanctions schemes as this would be the remit of national authorities. However, as detailed in sections 3.1 and 0, the framework proposed by the law commission expects that in-use monitoring to provide the regulator with evidence to support a civil sanctions programme.

NATM proposes that the Manufacturer should make available, a report, which provides information on the ADS performance during operation on public roads. The documentation should provide evidence of the ADS performance in safety relevant occurrences. In particular, the documentation package should demonstrate that:

- There are no inconsistencies with the assessed safety performance prior to market introduction. Specifically that the level of approved residual risk is not exceeded during real-world operation;

- the ADS respects the performance requirements set by UNECE (specifically the FRAV and VMAD groups)

Two documents are expected to be produced:

1. The In-service Data Report, which is periodically submitted to the Authority and shall contain information relevant to the requirements set above;
2. Supporting data used to elaborate the information provided into the In-service Data Report, exchanged with the Authority by means of an agreed data exchange file. The recommended data elements that form this supporting data is yet to be agreed.

The In-service Data Report is to be submitted as required by the national authority. No firm reporting frequency is recommended although it is suggested that there should be both periodic and reactive reporting. Periodic reporting would be at a fixed interval e.g. every six months. Reactive reporting would take place as soon as collected data provides evidence of an inconsistent ADS behaviour compared to the safety level declared prior to market introduction, or when collected data provide evidence of degradation of the safety margin).

The following safety-relevant occurrences are recommended to be reported by the Manufacturer:

- Interventions by the technical supervisor,
- In conflict scenarios, especially in accidents and near-accident scenarios,
- In the event of unplanned lane changes or swerving,
- In the event of malfunctions in the operating process.
- In the event that the driver (if any) does not respond on time to transfer of control requests

5 Detection of events

An AV should have the ability to detect collisions the same as, or better than a human driver. Current system designs for AVs put emphasis on safety requirements and test criteria for collision avoidance, but there is limited consideration for collision detection. Collision avoidance is also crucial, but it is inevitable that there will be instances where that collision avoidance fails and other situations where a collision is truly unavoidable. In these instances, the residual risk for not detecting a collision is high. If the AV fails to detect a collision it will not initiate the appropriate response (emergency stop, or MRM etc.) which could result in increased consequence severity and potential for secondary collisions before intervention. As such it is our recommendation that collision detection itself becomes a safety goal which much be argued in order to meet the definition of safe driving. This section considers how collisions and other unsafe events can be detected.

Task 2 of this project has developed a dataset specification which included data elements used to create triggers to identify safety relevant events for both lagging (an actualised risk event, e.g collision) and leading (an event indicative of an increased potential for risk, e.g. a near-miss) measures of risk. The Task 2 dataset specification (Chapman and Perren, 2021) outlined what data as well as potential value thresholds to be used to identify events and trigger data capture for risk evaluation. These dataset proposals were based on the data known to be collected on both human driven vehicles and existing automated vehicles as well as data known to be useful in risk evaluation across various transport domains. The suggested trigger measures are summarised in Table 2 below.

Table 2: Currently proposed leading and lagging measures as part of the dataset specification (Chapman and Perren, 2021)

Lagging measures	Leading measures
Vulnerable Road User (VRU) impact detection system activation triggers	Infraction Measurement – excess speed (Limit)
“Wake up” of occupant roll over protection systems	Infraction Measurement – excess speed (Safe)
Minimum Risk Manoeuvre (MRM) activation	Safety Envelope – proximity
System triggers of “wake-up” occupant protection systems	Driving style – longitudinal jerk
Battery / under vehicle impact protection	Driving style – lateral jerk
Vehicle door release when in motion	ODD exit
Safety Envelope close proximity detected	Hazard Identification, reaction and risk perception
Passenger emergency or remote operator control override mechanism	Safety pre trigger events – e.g. ABS pre-charge, Forward Collision warning

Vehicle dynamics beyond expected ranges (e.g. over max speed, or harsh events beyond design range)	
Unavailable or disabled autonomous sensor or control, fault triggers	

In this section a top-down approach is taken, starting with the key framework requirements and assessing to what extent the proposed dataset can adequately address these requirements, and what other procedural, administrative, or technical options are available for event detection.

5.1 Hazard Analysis

As discussed in Section 3.2, the framework must provide assurance that the AV follows the two pillars of safe and lawful operation. For the safety pillar, monitoring must ensure that the level of safety determined at type approval is being validated in use. This level of safety is dependent on:

- The safety goals and performance criteria assessed at type approval. At the highest level, the safety goals are:
  - Do not cause collisions
  - Avoid preventable collisions
  - Protect occupants
- The safety arguments stated in the vehicle and deployment safety cases

In-use monitoring must therefore identify any instances where these safety goals and arguments cease, or potentially cease to be met. It is widely accepted within in-use monitoring activities that monitoring collisions alone is not sufficient. It is necessary to capture any safety risks that endanger or which, if not corrected, would endanger a vehicle, its occupants or any other person. As such it is necessary for trigger for events to capture[8]:

- Collisions and actualised risk outcomes
- Other safety relevant events
  - Near collisions
  - Safety critical events
  - Proximity conflicts
  - Non-conflict critical incidents
- Traffic violations
  - Safety relevant rule violations

[8] This is defined in the Road incident Taxonomy for this project (Reed, N., 2022)

- Non-safety relevant road rule violations

In order to understand whether the current data set specification can adequately identify both collisions and safety relevant events, a hazard analysis exercise was conducted.

5.1.1 Methodology

Hazard Analysis was conducted to identify all hazards events that AV might encounter or cause during operation in automated mode. The considered list of hazards was constrained by the scope of the LSAV for this project (see Section 2). Hazards were identified via:

- Literature review and stakeholder consultation
- Using TRL's past knowledge and experience with past trials of AVs
- The safety goals and hazard analysis conducted by Horiba-Mira and York University in parallel to this work.

Hazards were analysed and structured using a deductive logic approach starting with a top-level hazard and then broken down into the sequential causes. This structure is visualised in Figure 6 below.

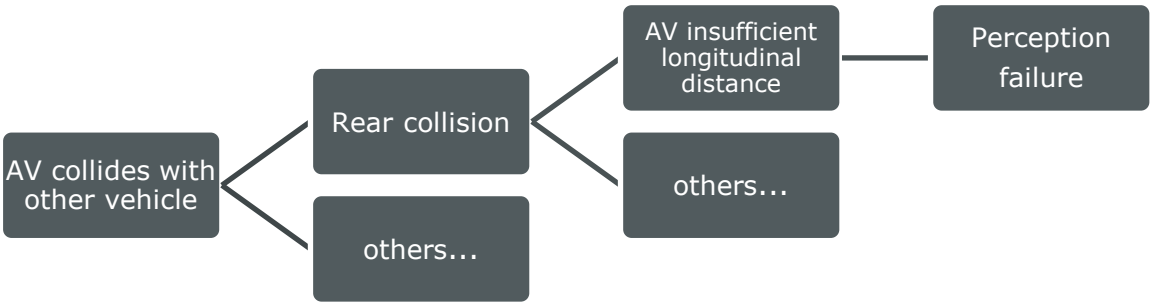


Figure 6: Structure of hazard analysis methodology

Hazard analysis was conducted separately for collision events and safety relevant events to determine the coverage of lagging and leading measures respectively. The currently proposed data triggers for lagging measures and leading measures were mapped against the identified hazards.

For each hazard event, an assessment was made as to whether each measure would activate and trigger data capture. It is recognised that the reliability of trigger activation for any hazard would be threshold and severity dependent. For example, a rear collision would only trigger via the "wake-up occupant protection systems" flag if the collision was severe enough to activate this safety system. The aim of this hazard analysis is to determine whether there are gaps in the existing dataset where it is impossible to detect some hazards. As such, data triggers were assigned to each event where there is a credible scenario where that trigger would flag the event of interest. The effectiveness of each flag is discussed in Task 2 dataset report (Chapman and Perren, 2021).

5.1.2 Results – Lagging measures

The hazard analysis exercise identified hazardous events involving an AV:

- Collisions,
- Passenger injury,
- Secondary collisions (i.e. not involving the AV but induced by the AVs behaviour),
- Post-collision hazard escalation, and
- Thermal events and noxious releases

Each affected party was identified for each hazardous event. Each collision type was further split by the event partner involved, in order to assess the differences in how different objects and collision partners could be detected. Finally, the causes of each event were assessed in order to understand whether the cause of the event and its context played a part in whether it could be detected by the proposed lagging measures.

In total, 366 hazardous events split by cause and affected party were identified. The coverage of lagging measures is summarised below in Table 3. Again, this assessment only considers whether it is credible that the measure could identify the event. It does not consider reliability of triggering the lagging measure.

Table 3: Lagging measure hazard coverage

	VRU impact detection system activation triggers	Wake up of occupant roll over protection systems	Minimum Risk Manoeuvre activation	System triggers of wake-up occupant protection systems	Battery / under vehicle impact protection	Vehicle door release when in motion	Safety Envelope close proximity detected	Vehicle dynamics beyond expected ranges	Unavailable or disabled autonomous sensor or control, fault triggers
Number of events	34	22	215	297	63	1	243	87	207
% coverage	9%	6%	59%	81%	17%	0%	66%	24%	57%

The assessment found that of the lagging measures currently proposed, one or more of them could credibly detect every hazardous event identified.

5.1.3 Results – Leading measures

The hazard analysis exercise identified hazardous events involving an AV:

- Near collisions and conflicts
- Unsafe vehicle dynamics – Harsh acceleration, deceleration, turning
- Rapid evasive manoeuvres

In total, 330 hazardous events split by cause and affected party were identified. The coverage of leading measures is summarised below in Table 4.

Table 4: Leading measure hazard coverage

	Infraction Measurement – excess speed (Limit)	Infraction Measurement – excess speed (Safe)	Safety Envelope – proximity	Driving style – longitudinal jerk	Driving style – lateral jerk	ODD exit	Hazard Identification, reaction and risk perception	Safety pre trigger events
Number of events	24	30	222	155	141	221	126	190
% coverage	7%	9%	67%	47%	43%	67%	38%	58%

The assessment found that there are 19 events that could not be feasibly identified with the current leading measures. All relate to near collision scenarios with various potential collision partners where there is a perception-based failure.

5.1.4 Hazard Analysis Discussion

The hazard analysis approach attempted to identify all safety relevant events that the AV would be expected to be able to detect and identified how effective the currently proposed set of leading and lagging measures would be in detecting them. While considerable effort has been devoted to ensure the completeness of the hazard analysis for the scope of this scheme, the hazards associated with an AV deployment will be highly dependent on the design of the AV, the deployment environment and the Manufacturer and Operators safety management systems and operational processes.

However, the hazard analysis approach undertaken was found to be a useful way of identifying events that require monitoring and ensuring the monitoring approach adequately addressed the identified events. This approach could be undertaken by Manufacturers when developing their monitoring approach. This is in line with current continuous improvement and monitoring approaches in safety management systems where monitoring requirements are derived from the risk assessment (BSI, 2020;Underwriter Laboratories, 2020).

When conducting this analysis, it was found that the context behind the event is crucial in determining whether a proposed measure would detect the event. This finding agrees with previous work (RAND, 2018;Peng, 2019;Automated Vehicle Safety Consortium, 2021). Some flags may be dependent on the severity of the hazardous event. For example, a high-speed differential collision may trigger the wake-up of occupant protection systems, whereas a lower speed differential collision may not. Detection of other events, based upon proximity or acceleration, may be dependent on the value of the detection threshold.

Even basic proximity measures require adjustment to the circumstances and situations of usage. For example, simple Time-To-Collision (TTC) proximity measures have been well studied (Sayed et al. 2013, Shariat-Mohaymany et al. 2011, Van der Horst 1990a, Vogel 2003, Sayed et al. 1994, Van der Horst 1990b) but across these investigations' threshold recommendations vary to match the needs for differing environments for instance giving differing thresholds for operation in differing operating domains, e.g. approaches to junctions vs. wider road usage.

Within the lower speed highly contested urban environments that LSAVs are likely to be deployed into, road incidents typically occur at relatively low operating speeds when in a contested environment with mixed road users. Recommended thresholds can therefore be found matching the requirements for risk detection in the employed environment. For instance, if given a TTC based approach, assuming the approach used, it would require thresholds matching the expected risk. To support intersections or areas with potential mid-level driver conflict and allow vehicle operation a TTC of 1.5-1.6 seconds could be used (supported by Van der Horst 1990a, Sayed et al. 2013, Huang et al. 2013).

Perception failures are a key area of residual risk as it proximity-based measures break down if the object has not been identified in the first place.

Failure to identify an object that has been involved in a collision or near collision also indicates a significant failure of the ADS Further assessment on the causes of perception failures was done to explore other potential measures. Cause of perception failures were broken down into:

- Detection of object too late
- Failure to detect object
- Incorrect classification of object
- Rapid changing of classification of object
- Sensor failure
- Failure to identify ODD exit
- Failure to predict object trajectory
- Detects object which does not exist

There is some coverage of these with existing measures proposed. For instance, perception issues caused by sensor failures would likely be identified by fault codes from unavailable or disabled autonomous sensor or controls. Other triggers may be used to account for these perception failures, such as:

- Rapid change in classification
- Impossible trajectory predictions
- Instances of objects appearing and disappearing

These triggers could provide greater coverage of perception-based failures but need to be discussed in terms of reliability and effectiveness. It is recommended that Manufacturers develop methods of monitoring these types of failure and define them within the Safety Case. Regardless, there is still the residual risk of complete non-detection of an object. It is widely accepted that this issue would require reporting mechanisms that are external to system under test which are explored in Section 7.1.3

The proposed dataset should be seen as absolute minimum in terms of in-use monitoring. As part of their risk management processes, Manufacturers should assess how hazards can be monitored in the field, with a focus on:

- The occurrence of unmitigated hazards and partially mitigated hazards
- The occurrence of hazards that have been accepted without mitigation
- Violations of assumptions, design goals, and conclusions made based on an evaluation of evidence made in the safety case

Manufacturers should be encouraged to identify further measures and other monitoring approaches to ensure tolerable coverage and evidence this as part of their type approval safety case. Monitoring options that can be considered are discussed in Section 7.1.3.

Safety envelope proximity triggers were found to be key for both leading and lagging measures due to its high coverage of safety relevant events. By selecting different thresholds, it could also serve as a measure for near miss scenarios and actualised risk events, serving as both a leading and a lagging measure.

6 Road Rule Compliance using ODD, OEDR and 3D World Model

Throughout the Law Commissions consultation there was strong public concern about road rules. However, idea of Government creating a "digital road rules book" that can be programmed into automated driving systems to cover every situation, circumstance or future scenario was rejected as infeasible.

Instead in the final report Law Commission noted 95% approval for the idea that the "*UK Government should establish a forum to collaborate with developers on how road rules apply to automated driving*" (Law Commission & Scottish Law Commission, 2022).

> *Recommendation 31*
>
> *The UK Government should establish a forum for collaboration on how road rules, traffic laws and guidance such as the Highway Code should apply to automated driving*

Chapter 6 of the final report recommends that breaches of traffic rules by a AV would no longer involve criminal prosecution, instead the in-use regulator will be given powers to apply a wide range of regulatory sanctions.

It was noted that frequent changes to road rules would mean AVs will require much greater monitoring while in-use to ensure they can deal with (changing) road rules.

The recommendation from the Law Commission was for the Secretary of State for Transport to publish a safety standard for AVs;

> ***Recommendation 6.***
>
> *4.66 The new Act should require the Secretary of State for Transport to publish a safety standard against which the safety of automated driving can be measured. This should include a comparison with harm caused by human drivers in Great Britain.*
>
> ***Recommendation 7.***
>
> *4.67 In exercising their functions, the authorisation authority and in-use regulator should have regard to the published safety standard.*

The safety standard would help reassure the public that AVs are safe overall, especially in the event of adverse incidents. However, the Law Commission made very clear statements regarding compliance with road rules;

> *4.53 AVs will be required to comply with road rules: if a road rule has been breached, our recommended scheme gives the in- use regulator power to issue civil penalties. More fundamentally, incidents should be seen as learning opportunities, leading to continuous improvements. Following an adverse incident, the in-use regulator would have power to issue an improvement notice. However, the authorities should not withdraw authorisation for incidents within the tolerance, unless there was evidence to suggest that the incidents would be repeated in a way that would fail the overall standard.*

There is therefore a clear need for the in-use regulator to be able to establish when road rules have been breached and gather evidence on these incidents in order to apply the appropriate level of sanctions.

A detailed analysis of the Highway Code rules relevant to LSAVs. For each rule the analysis identifies the applicable ODD elements, the expected OEDR performance requirements and the need for in-use monitoring of the 3D world model to provide evidence of rule compliance.

It also provides an analysis of the new Highway Code updates covering the safe passing of vulnerable road users and how in-use monitoring data, metrics and thresholds can be used to detect breaches. A summary of this analysis is given in Table 5.

Table 5: Summary results of traffic rules analysis for 165 LSAV relevant UK Highway Code rules identifying which rules require DDT elements, ODD attributes and performance metrics to assess rule compliance.

Rule attributes	No of rules	Percentage %
Total LSAV relevant rules	165	100 %
Specifies DDT elements only	41	24.8 %
Specifies DDT and ODD attributes	119	72.2 %
Does not specify any DDT or ODD elements	5	3 %
Monitored via OEDR performance metric and threshold	149	90.3 %

Finally, the appendix provides the results of research conducted by TRL analysing the applicability of Highway Code rules within the Smart Mobility Living Lab (SMLL) London which indicates the on-road testbed provides 85% coverage of LSAV applicable rules.

7 Additional monitoring

The Minimum Dataset Specification report (Chapman and Perren, 2021) found that event-based data capture was the primary recommended method for collecting in-use monitoring data, highlighting the issue that persisting large volumes of continuous vehicle data across a vehicle fleet for long periods of time would be largely unfeasible.

However, event-based monitoring at this stage is unlikely to reliably detect all events of interest. As such there it is expected that there will be false negatives – instances where an event occurs but the AV is unaware and so data capture is not triggered. There will also be situations (unknown risk) that were not previously considered and thus not captured through event-based monitoring. Several external mechanisms are suggested in Section 0 in order to reduce this residual risk. This is highlighted in Figure 7.

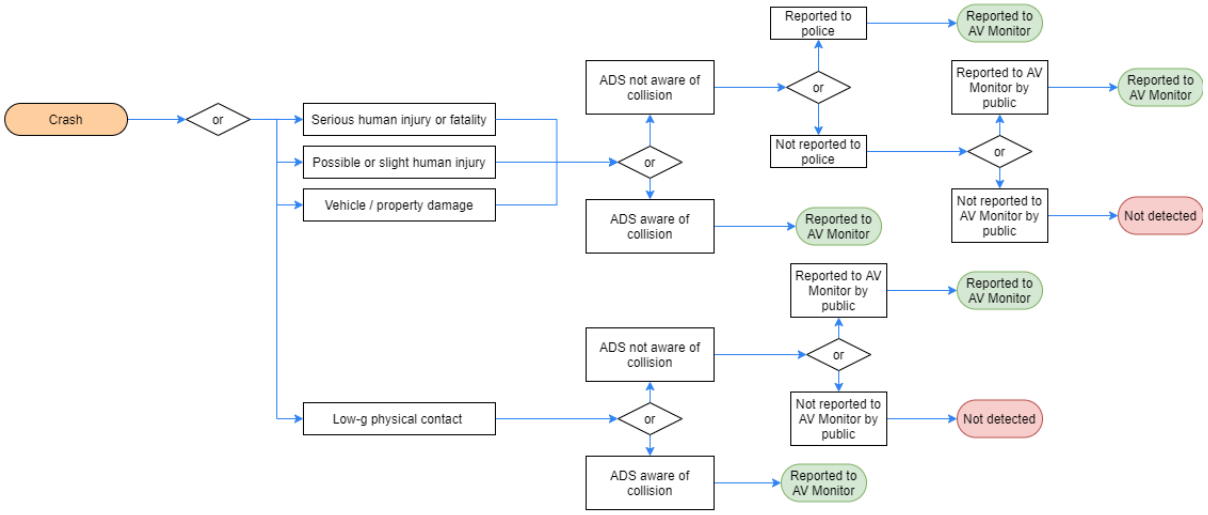


Figure 7: Logic flow diagram for collision detection

Logically the situation for difficult traffic infractions is even more difficult, as an AV should never knowingly break the law. As such, any traffic infraction committed by an AV will be one that the vehicle itself cannot detect directly.

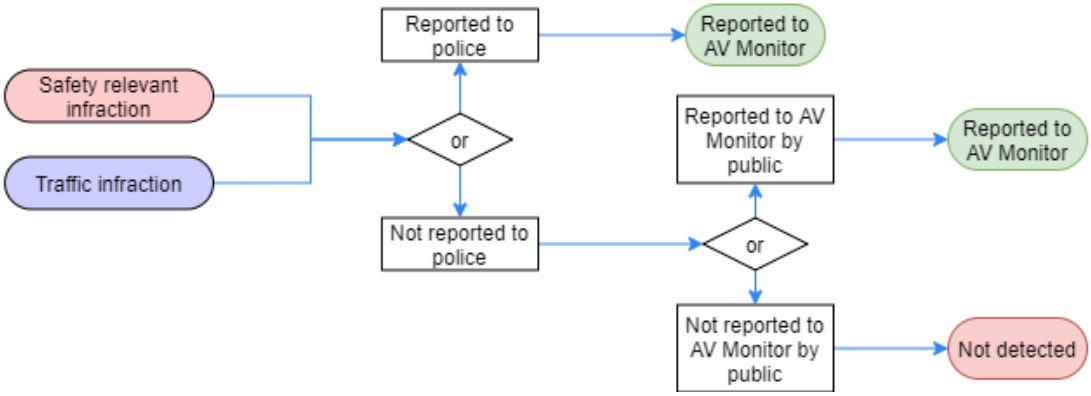


Figure 8: Logic flow diagram for detecting infractions

It is clear that external mechanisms for identifying risk events will be extremely important. These are described below.

7.1.1 Manufacturer defined monitoring requirements

The WP5 task 2 report (Chapman and Perren, 2021) detailed data monitoring requirements, defined a minimum set of performance metrics and data required to be stored. Collectively these data ensure safety risks requiring further mitigations can be identified and resolved. Specific focus was placed in data to identify collisions and unsafe events as defined by the Task 1 in-use monitoring taxonomy (Reed, 2021).

In Section 3.2, a hierarchy of monitoring requirements was specified (Figure 1) which highlighted that in-use monitoring should also ensure continuing compliance with type-approval requirements.

The proposed in-use monitoring dataset will provide evidence of this continuing type-approval compliance in a manner common to all Manufacturers. However, type-approval may also consider additional specific data and metrics submitted within an individual Manufacturer's safety case to support their specific safety arguments and assumptions.

It is common good practice with safety case development to ensure a process by which the arguments and assumptions made in the safety case are monitored. Safety Cases for automated vehicle trialling in accordance with BSI PAS 1881 shall include monitoring being conducted during operations to ensure enable the continued safety performance of the vehicle during operations and to validate risk decisions and assumptions made in the safety case (BSI, 2020).

Similarly, the UL 4600 standard requires that, in addition to collisions and practicably detectable incidents (i.e. near miss and other risks), a set of metrics and safety performance indicators (SPIs) to monitor for violations of assumptions, design goals, and the conclusions specified within the safety case (Underwriter Laboratories, 2020).

In line with current best practice, it is recommended that the Manufacturer develops a in-use monitoring plan that is evidenced within the safety case. The in-use monitoring plan shall evidence that there are processes in place to collect and investigate data in line with the minimum dataset specification as well as any additional processes required to monitor continued compliance with the safety case. Safety violations detected by thus in-use monitoring plan would be expected to be shared with the in-use regulator.

7.1.2 Continuous monitoring of driver behaviour (not the DDT)

Continual monitoring of the Dynamic Driving Task (DDT) to ensure it's safe execution would mean that the AV would not meet the self-driving definition. This type of continuous monitoring is performed by a human and intended to be prevent collisions e.g. the human monitors the OEDR performance of the ADS and intervenes when errors need correcting e.g. predicting that without intervention the vehicle may pass too close and too fast to a cyclist. Conceptually the human performs a similar role to a driving instructor in a dual-controls vehicle in which the ADS is the learner driver.

Continuous monitoring of driving behaviour is more like the function of a driving test examiner. The examiner also monitors the driver's skill in performing the OEDR task but the focus is upon assessment of the outcomes e.g. how close did the vehicle pass the cyclists rather.

There have been numerous efforts to develop methods for continuous in-use monitoring for the purposed of driver behaviour risk evaluation. The MOVE_UK project looked at methods for evaluating continuous in-use monitoring for longitudinal risk analysis. Following this, BSI produced proposals for CAV safety benchmarking termed Digital Commentary Driving.

Commentary driving is a technique used to train and assess human expert drivers, in which they are required to verbalize relevant information in the driving scene. This is used to determine that they can perceive, prioritize and act effectively when driving. Digital commentary driving (DCD) is proposed as an objective measure of CAV safety performance. DCD involves the continuous collection of data from an AV on its perceptions, decisions, reactions and feedback whilst driving. This data effectively probes the "understanding" the CAV has of its environment (BSI, 2021).

In task 2 of this project, methods of continuous monitoring of AVs were examined and a proposal for a set of continuous data was made (Chapman and Perren, 2021):

- Vehicle telemetry - GPS, speed, gyroscopes, accelerometers, telemetry accuracy and quality measurement (as undertaken in commercial telematics vehicle tracking)
- Proximity data for nearby objects - data derived from Object detection, distance, object classification, object direction using in vehicle camera/radar/ultrasonic/lidar etc.

With the addition of further data captured during transition events (i.e. wiper status on, ADS status engaged, etc.).

Continuous monitoring data would be an essential data source for events where the AV was unaware of the event occurring. Continuous data could provide information in terms of the state of autonomous systems and vehicle operating telemetry at time of an incident as well as factors contributing to safety.

Continual data collection is one of the few reliable and objective data sources available in situations where the AV was unaware of the event. This offers several benefits to Manufacturers. Firstly, it can help liability determination where the only other witness is the claimant. It can also help against false claims. Finally, continual data can help provide basic risk evaluation to help determine the causes of an event. This would allow regulators to have a greater understanding of an event and then take proportionate measures (i.e fair and just sanctions; see Section 0). Without this data, a cautious approach may be taken by the regulator who may impose more severe sanctions.

Ultimately the regulator will require evidence to support in-use monitoring investigations. This data may be made available through continual monitoring, through reliable event -based capture (though this is not expected to be fully reliable, at least initially), or through other operational means such as a steward in the vehicle. The data collected may also vary. While continual monitoring is recommended, it is thought that the best approach would be for the Manufacturer to define the appropriate level of continual data collection in order to control residual risk. This decision would be based on the deployment context, the Manufacturers Safety Management System (SMS), and their own cost-benefit analysis; balancing data storage and their risk tolerance for the exposure to claims and sanctions. This would be expected to be evidenced as part of their in-use monitoring plan within the SMS and the Safety Case.

7.1.3 External mechanisms for event detection

There are several operational processes that could interface with in-use monitoring to manage the residual risk of not detecting collision and unsafe events. The options identified are outlined below.

7.1.3.1 Vehicle checks

Processes for daily vehicle checks are a requirement for compliance with the FORS standard (FORS, 2021). The purpose of daily vehicle checks is to ensure that vehicles are inspected for safety by competent persons and defects rectified (if necessary) by competent persons prior to usage on each shift to ensure the safety of operations.

The walkaround check procedure include an inspection of the whole vehicle, trailer and any specialist equipment. In particular, the walkaround check covers the serviceability of:

- Wheels and tyres
- Brakes and steering
- Lights and markers
- Mirrors and window glass
- Obstructions to driver vision
- Bodywork condition
- Fluid levels and any leakages
- Vehicle safety equipment

For AVs additional items may be included such as sensor condition and cleanliness, antenna condition, any fault statuses with the ADS, etc.

For a fleet of AVs, any damage identified by this process may indicate a collision has occurred that the AV was unaware of or didn't flag or stop for. This may then be used as a trigger to look back at the data from the operating interval between checks to investigate the situation that led to the damage.

There is a residual risk that events which are not detected by the software or through fleet operator visual inspection, such as running over a pedestrian's foot, might go unreported. This point is addressed through public report as discussed below.

7.1.3.2 Public reporting

Public and societal confidence in the technology is key to promoting user acceptance of the technology. As such, a mechanism for public feedback may be appropriate, such as the creation of telephone hotline and / or online platform for the public to be able to report unsafe events directly to the in-use regulator, the Manufacturer or fleet operator.

For conventional fleet operator schemes such as FORS, there is no requirement to have a means for public reporting of collisions and unsafe events as the driver takes on this responsibility (FORS, 2016). Without a driver in the vehicle, public reporting may be an

effective redundancy to detect collisions and unsafe events. This may be valuable for passenger carrying AVs as it would allow passengers to act as witnesses to an event.

However, there is a risk that provision of a mechanism for public feedback may introduce significant amounts of irrelevant reports by the public. This may result in excessive burden on the Operator to investigate these events. Since there are no comparable processes in place today, it isn't possible to determine how effective a public feedback mechanism may be and how reliable the data will be that's provided by the public. It could however be an option that Operators choose to employ to act as a trigger to review operational monitoring data as another means of identifying events. If employed, it would be expected that the processes for handling public reports, disregarding false reports and investigating genuine ones be defined in the deployment safety case.

7.1.3.3 Police reporting

As with conventional vehicles, traffic offences and collisions may be reported to the police.

- Witness or victim testimony
- Outputs from traffic infrastructure – such as speed cameras and traffic light cameras.
- Reports from witnessing police officers
- Collision reports from attending police investigators

These reports may serve as triggers for investigation by the Manufacturer. The Law Commission framework recommends that traffic infractions be handled by the in-use regulator rather than the police, so that civil rather than criminal penalties can be applied (Law Commission & Scottish Law Commission, 2022).

As such there will need to be an interface between the regulator and police forces where AVs are deployed. For AVs without a driver in the vehicle it is envisaged that the police, upon receiving the report, would use the DVLA to identify the AV is approved and authorised under the scheme, and then forward the case to the in-use regulator for them to conduct further investigation determine the most appropriate course of action. In parallel, there is an expectation that the Manufacturer will have also identified and reported the issue to the in-use regulator. If the incident has not been reported by the Manufacturer this would likely lead to regulatory sanctions being imposed. This would likely involve the regulator referring the incident back to the Manufacturer for them to provide available information to provide an explanation of the event.

Police reports would be expected to contain the following:

- Time, date, and location of the event
- The committed or alleged offence by or involving the AV
- Vehicle identification- primarily registration but other information such as make, model and livery may also show the vehicle in questions is suspected to be an AV
- Any evidence collected by the police

This evidence is already requested or collected by the police through their various reporting mechanisms such as STATS19 forms (DfT, 2011), and online public reporting forms (Police.UK, 2022). The most significant operational change would be for the police to identify the vehicle as an AV and refer the case to the in-use regulator. Therefore, it must be possible for the police to be able to triage the event and determine whether the AV was involved. This may be achieved by requiring specific identifiers on the AV such as clearly distinct registration plates or livery. Operators could consider notifying local police forces prior AV deployment to ensure that they are aware of the AV operating schedule and any emergency response requirements. This is good practice for the operational safety management of AV trials currently (BSI, 2020)

8 Data Recall

In-use monitoring is a valuable tool to assure the safety of automated vehicle fleets but the prerequisite is having data to monitored to detect safety critical events and data that can stored and recalled when conducting an investigation.

The type of data will determine the type of safety assurance that can be offered by an in-use monitoring framework. Or conversely the type of safety assurance required to meet the regulator's objective, or public expectations, will determine the data required.

In accordance with the Automated and Electric Vehicles Act (AEVA) 2018;

> *The Secretary of State must prepare, and keep up to date, a list of all motor vehicles that—*
>
> *(a) are in the Secretary of State's opinion designed or adapted to be capable, in at least some circumstances or situations, of safely driving themselves, and*
>
> *(b) may lawfully be used when driving themselves, in at least some circumstances or situations, on roads or other public places in Great Britain.*

In which Section 8 Interpretation states;

> *…a vehicle is "driving itself" if it is operating in a mode in which it is not being controlled, and does not need to be monitored, by an individual;*

A pre-deployment type-approval scheme has value in establishing which automated vehicles should be included upon the list (the "a" requirement). A post-deployment in-use monitoring scheme has value in determining lawful use and which automated vehicles should remain on the list (the "b' requirement).

To meet the AEVA 2018 threshold both schemes need evidence to prove that the listed automated vehicles are capable of "safely driving themselves" while "operating in a mode" in which they are not "controlled" or "monitored" by an individual in at least "*some circumstances or situations".*

For in-use monitoring it appears that the following evidence is required as a minimum;

1. What *mode* was the automated vehicle operating?
2. Was the automated vehicle being *controlled* by an individual while operating in an automated mode?
3. Was the automated vehicle being *monitored* by an individual while operating in an automated mode?
4. Under what *circumstances or situations was the automated mode active*?
5. Was the *driving* task executed *safely*?

This minimum evidence needs to be supported by data as will be discussed below.

8.1 Data Recall – Operating Mode, Control and Monitoring

UN Regulation No. 157 covering the "Uniform provisions concerning the approval of vehicles with regard to Automated Lane Keeping Systems" specifies that vehicles should be fitted with a Data Storage System for Automated Driving (DSSAD).

The DSSAD is designed to record occurrences as a flag with an associated date, timestamp to within one second and reason for the occurrence. The specified occurrences include;

(a) Activation of the system

(b) Deactivation of the system, due to:

(i) Use of dedicated means for the driver to deactivate the system;

(ii) Override on steering control;

(iii) Override by accelerator control while holding steering control;

(iv) Override by braking control while holding steering control.

(c) Transition Demand by the system, due to:

(i) Planned event;

(ii) Unplanned event;

(iii) Driver unavailability;

(iv) Driver not present or unbuckled;

(v) System failure;

(vi) System override by braking input;

(vii) System override by accelerator input.

(d) Reduction or suppression of driver input;

(e) Start of Emergency Manoeuvre;

(f) End of Emergency Manoeuvre;

(g) Event Data Recorder (EDR) trigger input;

(h) Involved in a detected collision;

(i) Minimum Risk Manoeuvre engagement by the system;

(j) Severe ALKS failure;

(k) Severe vehicle failure

Implicit in the specification is the idea that the automated driving system only has a single mode of operation which is either active or inactive e.g. under ALKS control or under human control.

The ALKS transition demand would not be relevant for Low Speed Automated Vehicles (LSAVs) being considered for the initial phase of the GB scheme. However, the automated driving system may still be considered to be active or inactive e.g. executing the Dynamic Driving Task (DDT) or not executing the DDT.

So it is still necessary to capture more than one mode of operation. While the ADS is in it's active mode it would also be important to detect and record any human OEDR monitoring or control the DDT execution because doing so would violate the definition of self-driving.

For example, the Aurora Teleassist system, described in Section 4.4.2, considers a hybrid mode of operation in which the automated driving software remains in sole control of the vehicle actuators but a remote-control station with a teleassist human operator may be monitoring the execution of the DDT while assisting in the OEDR task required for safe vehicle control.

To meet the requirements of AEVA 2018 it must be possible to differentiate between an automated driving mode that executes the DDT without control or monitoring and a teleassist mode – which would not be considered self-driving. While out of scope for AVEA 2018, it would be useful for the in-use regulator to be aware of when a teleassist mode is active, the periods in which monitoring is active and the periods when an individual is actively controlling the vehicle as part of the OEDR execution.

In addition to teleassist modes of operation some Operators may to implement a full teleoperation mode. In teleoperation a remote individual may have full control of vehicle steering, throttle and brake actuation. In this situation the automated vehicle remains responsible for capturing and communicating the sensor data necessary for a remote individual's situational awareness, while the OEDR and vehicle control tasks may become the sole responsibility of the teleoperator.

The modes of operation for the automated vehicle should be made clear during type approval and in-use monitoring should continually track their activation state.

8.2 Data Recall – Safe Driving, Circumstances and Situations

Recording the operating mode, monitoring and control requires simple internal system state identification. In comparison, safe driving and the circumstances and situations in which the automated driving system is active requires a more complex external world state reference.

AEVA 2018 does not define "*circumstances or situations*" in it's text, however, it does state in Section 1 that the Secretary of State's self-driving list must include all motor vehicles that;

> *(b) - may lawfully be used when driving themselves, in at least some circumstances or situations, on roads or other public places in Great Britain.*

In the Cambridge Dictionary *"circumstances and situations"* are defined respectively, as "*an event or condition connected with what is happening or has happened*" and "*the set of things that are happening and the conditions that exist at a particular time and place*".

There is a close alignment between the meaning of "*circumstances and situations*" and the SAE J3016 definition of the Operational Design Domain (ODD) for an automated driving system which are the;

> *Operating conditions under which a given driving automation system, or feature thereof, is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics.* (SAE, 2021)

For additional context BSI PAS 1883 specifies a taxonomy for ODD attributes which includes three high-level attributes:

1. Scenery: non-movable elements of the operating environment
2. Environment: weather and other atmospheric conditions
3. Dynamic elements: all movable objects and actors in the operating environment (BSI, 2020)

Which implies that *circumstances and situations* are the unique combinations of scenery, environment and dynamic elements that define the events and conditions which exist at the particular time and place that the automated vehicle is operating.

The ODD for an automated driving system will need to be defined at type-approval. It is expected that the automated vehicle Safety Case submitted during type-approval will define the in-use monitoring required to verify that it remains inside it's approved ODD at all times during operation.

The Safety Case should also define the actions that will be taken upon detection of an ODD exit which may include; the execution of a Minimal Risk Manoeuvre (MRM) that brings the automated vehicle to rest in a Minimal Risk Condition (MRC).

For the in-use safety assurance regulator it's essential that ODD exit events, MRMs and MRCs are recorded.

In addition to the existing DSSAD requirements it is suggested that these specific occurrence flags (ODD exit, MRM, MRC) are recorded with date and timestamps to within a second and are supplemented with the location of the automated vehicle at the time of the occurrence.

In addition to recording the occurrence it's important to capture the *circumstances and situations* happening at the time which triggered the occurrence. This would include capturing the scenery, environment and dynamic elements to define the events and conditions present.

This information should be readily available from the automated driving software because it is required for detecting an ODD exit condition. It is also required for continually for execution of the OEDR during the DDT. This information is the software's digital representation of the physical world often referred to as a 3D world model, or more loosely, as digital twin.

8.2.1 The 3D World Model

The 3D world model is an abstract representation of the physical world's temporal and spatial dimensions. It may be considered as an output of the perception system that captures the scenery, environment and dynamic elements that define the ODD at the given time and location of the automated vehicle.

In Section 4.4.2 some examples of 3D World Model visualisations were presented. These computer game like visualisations enable humans to easily interpret the abstract digital world model representation being used by the automated vehicle for decision making.

3D world model visualisations are used as communication tools by Manufacturers to build public trust in the OEDR capabilities the AV is using to execute the DDT. A good example is

the 40 minute unedited video published by Intel Mobileye captured during a self-driving demonstration in the complex urban environment of Jerusalem[9].

The representation of the automated driving system's 3D world model is shown alongside an internal cockpit camera and an aerial viewpoint in which ego vehicle is shown surrounded with a blue circle. The automated vehicles' lateral location in lane and longitudinal location relative to the detected crosswalk is displayed. Other scenery shown includes the central reservation, side road and bus stop location in which a car is parked.

On the crosswalk you can see a pedestrian which is shown as red on the 3D world model visualisation to indicate the automated vehicle must yield. Also depicted in the scene are four cars travelling in the same direction as well as one car and one motorbike travelling in the opposite direction.

Additional scenes show the automated vehicle's 3D world model includes a representation of the location and state of traffic lights and dynamic elements including pedestrians, a motorcycle rider, cars and a bus.

Access the 3D world model being used by the automated driving software, for decision-making, planning and control, would provide valuable evidence for proving the vehicle continuously meets it's lawful use requirement defined by AEVA 2018. According to the ODD definition of BSI PAS 1883, the 3D world model used by the AV should include a representation of the scenery (static objects), the environment (weather) and the dynamic elements (other road users).

8.2.1.1 3D World Model - circumstances and situations

The use of in-use monitoring to ensure that the automated vehicle only operates inside its lawful circumstances and situations will depend upon the specification of the ODD permitted during type-approval.

Additionally, there is also a need for the in-use monitoring of circumstances and situations that occur within the approved ODD.

For example, an automated vehicle might be approved to operate in an ODD that includes cyclists, however, in-use monitoring may indicate a violation of required safe passing distances and speeds when a cyclist is encountered. Under the proposed regulatory framework this may be considered lawful use, but breaching the expectation of "safely" driving itself may incur sanctions imposed by the in-use safety assurance regulator.

8.2.1.2 3D World Model – safe driving

The in-use monitoring of safety critical events occurring within the approved ODD can also be conducted using data extracted from the automated vehicle's 3D world model.

For example, an automated vehicle might be approved to operate in an ODD which includes signal-controlled junctions, however, in-use monitoring would be required to detect if the AV drove through a red light. This safety critical event would occur during lawful use of the

[9] https://youtu.be/kJD5R_yQ9aw?t=252

automated vehicle and under the regulatory framework would incur a regulatory sanction. The in-use regulator would need to give consideration to the arguments made in the Manufacturers safety case and the evidence presented during the approval phase because withholding of falsifying information could lead to criminal sanctions.

The automated vehicles 3D World Model also provides a tool for assessing and assuring the safety performance of the Object and Event Detection and Response (OEDR) task.

In the Intel Mobileye video [10] there is a sequence of frames, separated by around 1/10th of a second, in which a number of other road users are missing from the world model. The missing road users can be clearly identified using the aerial drone footage as the world model ground truth.

Over the sequence of frames there is clear instability in the detection and persistence of objects (a bus and a car) within the world model representation.

Capturing the instantaneous 3D World Model representation upon key event occurrences is valuable, however, this can lead to a false sense of OEDR performance. As shown by the sequence above it's important to track the 3D World Model over time to provide evidence of OEDR performance stability.

It's obvious to human drivers that objects do not instantly appear in, or disappear from, the world, and that, if they become occluded, they are still present in the world. It's also true that human drivers understand that objects obey the laws of physics in the way that they move and that it is not possible for them to transform from one object class to another e.g. from a car to a bus or a bendy bus to separate lorries.

The same cannot be assumed for the perception systems of automated vehicles (as evidenced in the Intel Mobileye video). This means that it is critical for the in-use safety assurance scheme to track the 3D World Model and evaluate OEDR task performance.

The concept of using the 3D world model to evaluate OEDR performance is not new. In fact, the process of human commentary driving is based on the same principle, in which the examiner independently constructs a 3D world model to act as a ground-truth for the verbal descriptions and actions of the driver being assessed.

Similarly, Driver Monitoring Systems (DMS) are used in real-time to track human driver attention required for the OEDR task e.g if the driver is not looking out the windscreen it indicates that they are no longer completing the OEDR task successfully. A real-time DMS can also be used to alert the driver when inattention is detected and the OEDR is not being completed safety.

In-use monitoring could also be used to provide feedback to the automated driving software when unsafe outcomes are detected. The feedback mechanism may be real-time but it will always occur after the detected unsafe event has occurred – meaning that it cannot be used to prevent an unfolding collision. Instead, the primary value of this real-time feedback would be to mitigate the risk of future collisions.

For example, if in-use monitoring detects that a cyclist was detected late which resulted in the AV passing too close and too fast. Informing the AV of this safety critical event could be

[10] https://youtu.be/kJD5R_yQ9aw?t=252

valuable in improving the safety of cyclist interactions occurring later in the same journey. While the software isn't capable of immediate learning, it would have a number of mitigation options available. One of which would be to execute an MRM/MRC until a remote operator has reviewed the incident, identified the root cause, assessed the risk of reoccurrence and provide remote authorisation that automated driving can be safely resumed.

The judgement on how to respond in-use monitoring feedback could be left to the individual Manufacturer and described in their approved safety case. However, in the future the in-use regulator may define the expected response to certain safety critical event notifications e.g. a mandatory MRM/MRC stop might be required if in-use monitoring detected the passing of a traffic controlled junction while the traffic lights were on red.

8.2.1.3 3D World Model – collision reconstruction

The 3D World Model is also a valuable tool for collision reconstruction.

As was shown earlier, 4.4.4 Dash Cam footage from human driven vehicles is already being used to reconstruct collision events using AI to extract a 3D World Model representation and combine this with ego vehicle GPS location and acceleration data to create an event timeline.

Dash cams could be used in an identical manner for automated vehicles, however, just as with human driven vehicles, they only act as a proxy for the 3D World Model actually used by software to execute the OEDR/DDT.

In the case of human driven vehicles, dash cams can act as an independent digital witness to corroborate the human driver's recollection of events. This is useful given the unreliability of driver event recall and testimony.

However, for an automated vehicle it's reasonable to expect 100% recall and 100% reliability given that all events can be stored as digital data. Therefore, for AVs, rather than just having a dash cam proxy it's reasonable to expect recall of the actual 3D World Model used by the AV in it's decision making.

The NTSB investigation into the automated vehicle collision which caused Elaine Herzberg's fatality provides a good indication into the value of dash cam footage and 3D World Model reconstruction (NTSB, 2019).

The dash cam image below was captured around one second before impact. You can see Elaine Herzberg crossing the road as a pedestrian pushing a bicycle. The dash cam footage was easily available to the police and was released early to the public. It provided a false sense that it was impossible for the automated vehicle software to have detected Elaine Herzberg any sooner and that therefore the collision was inevitable and the result of pedestrian jaywalking.

Figure 9: NTSB: HWY18MH010, Tempe, Arizona - Uber ATG – Dash Cam Footage (-1s)

However, during the more detailed NTSB investigation a full timeline of event was reconstructed (see figures below) which provide evidence of the 3D World Model being used by the automated vehicle and the associated decisions and actions that were being made.

From the reconstruction it's clear that the automated vehicle detected Elaine Herzberg 5.6 seconds before the collision event but miss classified her as a vehicle and failed to detect her motion or predict her future path would interest the ego vehicle trajectory.

Figure 10: NTSB: HWY18MH010, Tempe, Arizona - Uber ATG – Collision Reconstruction

Time (s) relative to impact	Speed (mph)	Classification and Path Prediction[a]	Other Events / Details[b]
-9.9	35		Vehicle begins to accelerate from 35 mph due to an increased speed limit
-5.8	44		Vehicle reaches the speed of 44 mph
-5.6	**44**	**Classification: Vehicle - by radar** **Path prediction: None; not on the path of the SUV**	**Radar makes the first detection of the pedestrian and estimates its speed.**
-5.2	45	Classification: **Other** - by lidar Path prediction: Static; not on the path of the SUV	Lidar detects an unknown object; this is the first detection of that object by lidar, the tracking history is unavailable, and its velocity cannot be determined. ADS predicts the object's path as static.
-4.2	**45**	**Classification: Vehicle - by lidar** **Path prediction: Static; not on the path of the SUV**	**Lidar classifies a detected object as a vehicle; this is a changed classification of an object and without a tracking history. ADS predicts the object's path as static.**
-3.9		Classification: Vehicle - by lidar Path prediction: The left through lane (adjacent to the SUV); not on the path of the SUV	Lidar retains the classification "vehicle", and based on the tracking history and the assigned goal, ADS predicts the object's path as traveling in the left through lane.
3.8 > - 2.7	45	Classification: alternated several times between **vehicle and other** - by lidar Path prediction: alternated between static and left lane; neither were considered on the path of the SUV	The object's classification alternates several times between vehicle and an unknown. At each change, the object's tracking history is unavailable, and ADS predicts the object's path as static. When the detected object's classification remained the same, ADS predicts the path as traveling in the left through lane.
-2.6	**45**	**Classification: Bicycle - by lidar** **Path prediction: Static; not on the path of the SUV**	**Lidar classifies a detected object as a bicycle; this is a changed classification of the object, and without a tracking history. ADS predicts the bicycle's path as static.**
-2.5	45	Classification: **Bicycle** - by lidar Path prediction: The left through lane (adjacent to the SUV); not on the path of the SUV	Lidar retains the classification "bicycle" and based on the tracking history and the assigned goal, ADS predicts the bicycle's path as traveling in the left through lane.

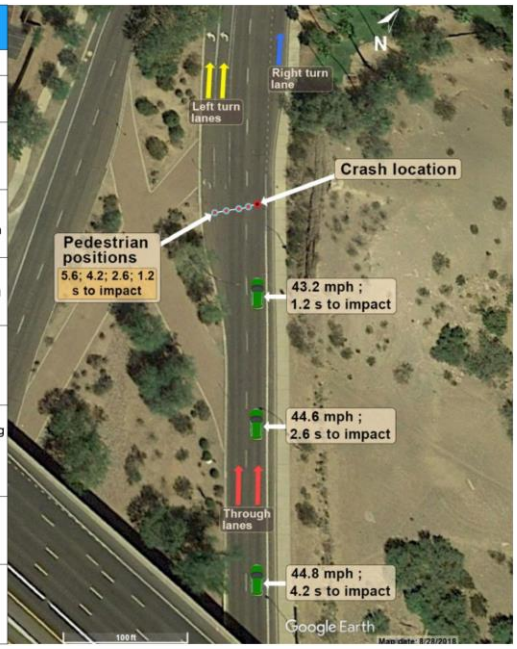


Figure 11: NTSB: HWY18MH010, OEDR Performance (from -5.6s to -2.5s)

Time (s) relative to impact	Speed (mph)	Classification and Path Prediction[a]	Other Events / Details[b]
-1.5	*44*[c]	Classification: ***Unknown*** - by lidar Path prediction: *Static*; partially on the path of the SUV	Lidar detects an unknown object; since this is a changed classification, and an unknown object, it lacks tracking history and is not assigned a goal. ADS predicts the object's path as static. Although the detected object is partially in the SUV's lane of travel, the ADS generates a motion plan around the object (maneuver to the right of the object); this motion plan remains valid—avoiding the object—for the next two data points.
-1.2	**43**	**Classification: *Bicycle* - by lidar** **Path prediction: The travel lane of the SUV; fully on the path of the SUV**	**Lidar detects a bicycle; although this is a changed classification and without a tracking history, it was assigned a goal. ADS predicts the bicycle to be on the path of the SUV.** **The ADS motion plan—generated 300 msec earlier—for steering around the bicycle was no longer possible; as such, this situation becomes hazardous.** **- Action suppression begins**
-0.2	40	Classification: ***Bicycle*** - by lidar Path prediction: The travel lane of the SUV; fully on the path of the SUV	Action suppression ends 1 second after it begins. The situation remains hazardous; as such, ADS initiates a plan for vehicle slowdown. *An auditory alert was presented to indicate that the controlled slowdown was initiating* [d]
-0.02	39		Vehicle operator takes control of the steering wheel, disengaging the ADS.
IMPACT			
0.7	37		Vehicle operator brakes

[a] Only changes in object classification and path prediction are reported in the table.The last reported values persist until a new one is reported.

[b] The process of predicting a path of a detected object is complex and relies on the examination of numerous factors, beyond the details described in this column

c The vehicle started decelerating due to the approaching intersection, where the pre-planned route includes a right turn at Curry Road. The deceleration plan was generated 3.6 seconds before impact.

[d] *While the system generated a plan for the vehicle slowdown, due to a slight communication delay, the data is unclear on whether the implementation of the slowdown plan started before the operator took control prior to impact.*

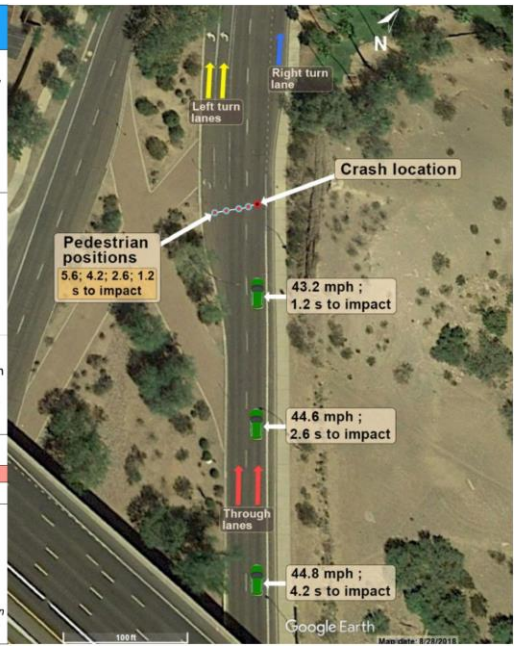


Figure 12: NTSB: HWY18MH010, OEDR Performance (from -1.5s to 0.7s)

The final NTSB investigation report was published 20 months after the collision event occurred.

It's important to recognised that, as the number of deployed automated vehicles and number of investigations increases it will be essential to scale the investigation processes to cope with a much larger volume of collisions events, near-miss events and traffic infractions.

A 3D World Model based approach to in-use monitoring will become a key tool in scaling the investigation process.

One can envisage that the 3D World Model reconstruction should be made as easily available and accessible to police officers as dash cam footage is now. Additionally, it would be possible for the automated vehicle to utilise mobile communication to share 3D World Model incident data instantly with a cloud hosted service to which all relevant agencies can subscribe, including the in-use safety assurance regulator.

8.2.1.4 3D World Model – perception, decision, reaction and outcome explainability

In it's feedback to Law Commission Consultation Paper 3 the ITU FG-AI4D described a model for explainability using the 3D world model as the source data. The model described the benefits of the 3D world model for enabling explainability of four key DDT topics;

1. Perception explainability
2. Decision explainability
3. Reaction explainability
4. Outcome explainability

More details of this proposed model and it's relevance to in-use modelling based around the AVs 3D world model can be found in Appendix D.

9 Event Reporting

Reporting schemes are common in regulatory frameworks. In aviation The CAA operates a Mandatory Occurrence Reporting scheme in accordance with Regulation (EU) No 376/2014. The objective of the MOR Scheme is to contribute to the improvement of flight safety by ensuring that relevant information on safety is reported, collected, stored, protected and disseminated.

A reportable occurrence in relation to an aircraft means any incident which endangers or which, if not corrected, would endanger an aircraft, its occupants or any other person (CAA, 2021).

The sole objective of occurrence reporting is the prevention of accidents and incidents and not to attribute blame or liability. The CAA gives an assurance that its primary concern in relation to the MOR scheme is to secure free and uninhibited reporting and that it will not be its policy to institute proceedings in respect of unpremeditated or inadvertent breaches of the law which come to its attention only because they have been reported under the Scheme, except in cases of gross negligence (CAA, 2021).

Similarly, for proposals in the VMAD – NATM group, the aim for In-Service Monitoring and reporting (ISMR) is to contribute to the improvement of road safety by ensuring that relevant information on safety is reported, collected, stored, protected and disseminated. Again, the sole purpose is the prevention of accidents and incidents and not to attribute blame or liability (VMAD SG3-14-14, 2021).

For an approval scheme, the prevention of incidents is also the priority. However, the GB scheme also aims to utilise in-use monitoring for to ensure continued compliance with the type approval requirements. Furthermore in-use monitoring data and reports are expected to support the application of civil sanctions where non-compliance is found (Law Commission & Scottish Law Commission, 2022). There is a potential conflict between encouraging reporting for the prevention of incidents free of blame while using the same information to monitor compliance and apply civil sanctions. Options to ensure civil sanctions are just and proportionate are discussed in Section 0. This section discusses what occurrences need to be reported, what information needs to be provided to regulators, what the processes could be for reporting, and how reporting can be encouraged.

9.1.1 Reportable occurrences

It is recommended that best practice from aviation be considered. Thus, reportable occurrences for this scheme should include:

Any occurrence which endangers or which, if not corrected, would endanger the AV, its occupants or any other person.

Sub definitions for such events have been defined in the In-use monitoring taxonomy (Reed, 2021) developed under this project, they are:

- Collisions
- Near-collisions
- Safety critical events

- Proximity conflicts
- Non-conflict critical incidents
- Safety-relevant infractions
- Traffic infractions

These events will enable the regulator to monitor the compliance of the AV as well as generate data for comparative assessment of safety performance. Collection of these events may provide evidence that indicates a violation or potentially violation of the safety case or safety performance declared at type approval. However, as detailed in Section 7.1.1, the Manufacturer would be expected to have monitoring in place to assess safety arguments made in their safety case, such as monitoring object detection failure rates or classification failure rate. As such, reportable occurrences should include any instances where arguments made in the safety case or safety performance has been invalidated

Within the stakeholder engagement activity of this project, it was identified that collecting, investigating and analysing AV safety data is a specialist, in-demand skill, requiring in-depth knowledge of the specific system in question. Those who have this skill would likely be working OEM/developers and not public bodies. As such, it is envisaged that the responsibility for interpreting AV safety data to determine whether there is potential non-compliance with type approval sits with the Manufacturer.

This would need oversight by the regulator. This could be achieved through the audit of the Manufacturer's Safety Management System to ensure there are proper monitoring, investigation and reporting processes in place to provide confidence.

Law commission envisages the in-use regulator will use a range for civil sanctions as part of the regulatory framework. However, if the Manufacturer deliberately misleads or fails to disclose safety critical evidence during type-approval, or during on-road operations, then criminal prosecution may be levied.

9.1.2 Information to be reported

Reporting requirements need to be specified robustly, otherwise there is a risk that developers and Manufacturers do not give adequate, consistent, and factual information to the regulator.

There is expected to be two modes of reporting which will have different information contained (VMAD SG3-14-14, 2021):

- Individual event reports and case study analysis
- Aggregated data and statistical analysis

Both modes of reporting could potentially identify a violation of type approval.

9.1.3 Individual event reports

Individual reports may be required for:

- Lagging measures – their low frequency would inhibit statistical evaluation, but their high data availability allows in depth analysis.

- Police reported events such as traffic infractions – where the regulator would require the Manufacturer to investigate the event.
- Events not recorded by the event data capture system but detected through other processes – indicating there is a potential failure
- Any other event requested by the regulator – For instance, where the regulator receives significant public complaints about a near collision event.

Individual event reports should be underpinned by case study analysis. The approach for investigating and analysing events should be detailed within a Manufacturer's Safety Management System. The event reports should display sufficient information required for the regulator to determine whether the event violates the safety performance on which the type approval was based. NHTSA have developed a reporting form for collisions involving ADS and ADAS features as a means of ensuring consistent and comparable information has been provided (NHTSA, 2021). Information contained in the NHTSA report should include:

- Reporting entity information
- Subject vehicle information, including:
  - VIN and registration
  - ADS/ADAS version
  - Operating entity
  - Make and Model
- Incident information, including:
  - How the incident was identified (claim, complaint, telematics, law enforcement of testing)
  - Date and time of incident
- Details of the incident scene, including weather, road and surface information and location
- Description of the crash, including:
  - Crash partner
  - Highest injury severity
  - Property damage
  - Vehicle damage of both subject vehicle and partner
  - Crash and post-crash details such as occupant restraint status and non-reversible restraints (e.g air bags) deployment status
  - Pre-crash speed
- Post-crash information, including:
  - Data availability
  - ODD status

- Investigation status (and details of investigator)

- A written narrative of pre-crash, crash and post-crash detail, including any ADS features engaged and any issues leading up to the crash.

This could be used as a basis for the specification of an individual event report for an in-use monitoring scheme with the addition of further data to aid in the regulators assessment based on best practice AV monitoring and continuous improvement process (Underwriter Laboratories, 2020;BSI, 2020). This includes:

- Whether the event realised is associated with a hazard identified within the safety case.
- Whether any risks associated with the event assessed in the safety case were accepted without mitigation.
- The credit, if any, that can be attributed to any safety arguments in mitigating the event.
- Whether the AV acted inconsistently with the behaviour and performance requirements.

9.1.4 Aggregated data

Aggregated datasets enable statistical evaluation of safety performance over a specified time periods. Aggregated data can be used during in-use monitoring as evidence that validates the AV safety level, and therefore the accepted level of residual risk, that were approved during authorisation are actually achieved during on-road operations.

Frequency of occurrence of the leading and lagging measures are defined in the dataset specification. They may be categorised in different ways to generate learnings of safety performance (ISO, 2018):

- Categorised by the number of false positives and confirmed events to determine the suitability of selected thresholds
- Categorised by the number of confirmed events investigated to root cause and their conformance with the safety arguments stated within the safety case
- The relevant crash characteristics that could inform blame determination. For example: The ego vehicle is hit from behind at a stop sign would by default be blamed on the trailing vehicle. An elevated rate of not-blamed loss events could still be indicative of safety issues, such as the AV behaving in a manner that provokes mistakes by human drivers.
- The demographic of involved parties. This is intended to identify patterns is safety performance for different user groups. For example, this may identify biases in machine learning training sets or defects in ODD construction so that they can be corrected.

Any other metrics defined in the safety case should also be reported on:

- Definition of the metric and the safety argument its associated with
- Value of the metrics, such as failure rates of a particular system or function

- Conformance with any metric targets or thresholds set during development
- Any positive contributions to public safety (VMAD SG3-14-14, 2021)

9.1.5 Reporting frequency

The VMAD NATM groups envisages in-service reporting taking two forms; periodic and reactive reporting. Periodic reporting would be at a fixed interval e.g. every six months. Reactive reporting would take place as soon as collected data (either aggregated data or case studies) provides evidence of an inconsistent ADS behaviour compared to the safety level declared prior to market introduction, or when collected data provide evidence of degradation of the safety margin) (VMAD SG3-14-14, 2021).

It is also recommended that collisions be reported to the regulator immediately in order to enable the regulator to manage crisis communications and media reporting to ensure public confidence in the regulator. It is recommended that the following collisions be reported immediately:

- Level 3: Police-reported collision with vehicle / property damage only.
- Level 4: Police-reported collision with possible or slight human injury.
- Level 5: Police-reported collision with serious human injury or fatality.

10 Sanctions

The Law Commissions propose that infractions (including non-safety related ones) would be dealt with through a graduated system of regulatory sanctions. The currently proposed sanctions are:

(1) informal and formal warnings;

(2) fines;

(3) redress orders;

(4) compliance orders;

(5) suspension of authorisation;

(6) withdrawal of authorisation; and

(7) recommendation of attendance at a restorative conference.

Outside of those sanctions recommended by the Law commission, there is growing support for the use of enforcement undertakings in civil sanctions regimes (DfT, 2015). An enforcement undertaking is, in essence, a binding agreement in relation to a breach that is created by a voluntary offer from the potential offender and subsequent acceptance of this offer by the relevant regulator. Enforcement undertakings are used extensively for environmental regulation. It has proved to be the case that, when faced with other civil sanctions, the overwhelming majority of offenders offer an enforcement undertaking (BCLP Law, 2019).

The action that the regulator can offer a person to undertake must be:

- action to secure that the offence does not continue or recur;
- restorative or remedial action;
- compensation; or
- other action prescribes, such as promotional and educational activities.

Regulators find that they are beneficial rather than the regulator actively expending effort to investigate and pursue enforcement action against the passive regulated sector, the regulated sector actively comes forward itself with a proposed suite of appropriate actions for the regulator's acceptance. (BCLP Law, 2019)

While safety regulation differs from environmental regulations in a number of important respects, it is possible that the benefits of enforcement undertakings would transfer well to in-use safety assurance of AVs.

10.1 In-use monitoring data to support sanctions

Once in-use monitoring detects a safety issue, it will be necessary for the regulator to determine the most appropriate corrective action. The data required to determine this is expected to come from in-use monitoring and reporting which can be used to understand the event. By understanding the extent to which ADS failures/non-compliances contributed to an event, it is possible to apply proportionate and just sanctions.

One option for assigning sanctions could be to establish a relationship between the type (severity and frequency) of the violation and the level of sanction applied. For example, a preventable fatal collision would incur suspension of authorisation. However, this approach fails to take advantage of in-use monitoring to understand events in greater context and take proportionate measures based on the risk.

Instead, a risk-based approach may be used to identify the proportionate level of sanction. The HSE recommends that organisations prioritising investigation of incidents and near misses based on the potential for risk if the event is not resolved. This takes into account the likelihood of recurrence of an event and the potential worst consequences of the event should it happen again. The HSE prioritisation matrix is given in Figure 13. Risk based decision making is also commonplace in market surveillance regulation. If non-compliance does not pose a risk to consumers, market surveillance authorities may limit their action to ensuring that the next batch of products which is placed on the market is compliant. In other cases, they may impose sanctions on non-compliant businesses (UNECE, 2012).

Likelihood of recurrence	Potential worst consequence of adverse event			
	Minor	Serious	Major	Fatal
Certain				
Likely				
Possible				
Unlikely				
Rare				

Risk		Minimal		Low		Medium		High
Investigation level		Minimal level		Low level		Medium level		High level

Figure 13: HSE prioritisation matrix for investigating near misses

By focusing on the outcome of investigation rather than severity of initial event, it allows for sanctions to more effectively manage risk as well as be more just for the Manufacturer. It also provides incentive for Manufacturers and operator to provide information on the context of the event. Without sufficient understanding of the event, it would not be possible to accurately assess potential risk. As such a more cautious approach would have to be taken which may involve more severe sanctions that are more disruptive to the Manufacturer/Operator.

Other factors would likely affect the application of the most appropriate sanction.

For example, it would be possible to build upon the Law Commissions' recommendations using the CAA approach which considers various aggravating or mitigating factors to help determine the severity of the event, and thus the sanction which should apply (CAA, 2020).

Examples of aggravating factors include:

- The offender's state of mind and level of culpability: deliberate, reckless, negligent or accidental;
- Awareness of the offence and the risk of harm likely to arise from the offence;
- Disregard of warnings from the regulator, or from within the workforce;
- Poor co-operation with the regulator; and
- The prevalence of the offence such that deterrence is a priority.

Examples of mitigating factors include:

- Prompt and full remedial action taken by the offender;
- Immediate and voluntary reporting of the offence;
- Admission of responsibility;
- Previous good compliance record;
- Preparedness to co-operate with the regulator; and
- Personal circumstances or case-specific factors.

10.2 Where sanctions apply?

10.2.1 Operators

The Law Commissions final report (Law Commission & Scottish Law Commission, 2022) provides a recommendation for the conditions under which a Operator Licence will be granted;

> ***Recommendation 54.***
>
> *To obtain a NUIC operator licence, the applicant should submit a safety case, showing how safety will be assured. Among other things, the applicant's safety case should set out:*
>
> 1. *(1) how oversight will be provided to vehicles, including suitable connectivity, equipment, staff training and rest breaks;*
> 2. *(2) incident management, including communication with passengers, road users and the emergency services, together with measures to remove vehicles causing an obstruction;*
> 3. *(3) systems, expertise and equipment to maintain vehicles, install updates and ensure cybersecurity;*
> 4. *(4) data management;*
> 5. *(5) whether safety relies on any element of remote driving, and (if so) how this will be done safely; and*
> 6. *(6) ways to learn from mistakes, including links with local authorities, highway authorities and the police.*

Where an ASDE and the NUIC operator are the same entity, the entity may submit a joint safety case covering both roles, to be assessed by the authorisation authority.

In other cases, the safety case should address the Manufacturer's written specifications for what must be done to ensure safe operation.

Additionally, Recommendation 56 states that the "*new Act should give the regulator powers to impose the following regulatory sanctions on Operators*".

The range of sanctions that can be imposed are the same as those listed for an ASDE at the start of this section; *informal and formal warnings; civil penalties; redress orders; compliance orders; suspension of licence; withdrawal of licence; and recommendation of attendance at a restorative conference*.

Where the Operator and the Manufacturer the same organisation it is recommended that the "*regulator can impose sanctions on the combined Manufacturer/Operator organisation, without having to establish in which role it was responsible for the fault*".

For Compliance Orders related to technical issues the recommendation is that they should be "*outcome-oriented: they should specify the result to be achieved, rather than the means for doing so*". It is expected that the Manufacturer/Operator may be able achieve the outcome either through changes to the vehicle, or changes to its operations. The in-use regulator "*would have power to issue a compliance order specifying the outcome to be achieved to the combined Manufacturer/Operator, without distinguishing between the two roles*".

Where the Operator role is separate the in-use regulator would still be able to apply a the range sanctions flexibly and depending upon the circumstances.

10.2.2 Manufacturer

The Law Commissions report outlines sanction applicable to manufacturers that fulfil the role of the ASDE. The ASDE is the entity that puts the AV forward for legal categorisation as self-driving and is legally responsible for how the AV performs dynamic control (Law Commission & Scottish Law Commission, 2022).

The ASDE "*must have been involved in the safety assessment. It must submit a safety case and other required documentation. It must also be willing to vouch for the information it has given the authorisation authority. If the information in it is inaccurate the ASDE might be guilty of a serious criminal offence*".

The Law Commission noted that a wide variety of organisations may work together to develop self-driving vehicles and there maybe a variety of structures to manufacture vehicles, to bring them to market or deploy them on the roads.

Notwithstanding the need for these structures, the recommendation is that a single entity, the Manufacturer, is registered with the authorisation and in-use authorities as the first point of reference in the event of problems. For example, the Manufacturer may be a vehicle manufacture, a ADS developer, a joint venture between the two or even a Operator.

The Manufacturer authorisation would be conditional on good repute, appropriate financial standing (funds or insurance to cover liability and recalls) and the submission of a safety case and equality impact assessment.

Additional ongoing conditions were also recommended;

Recommendation 13.

Authorisation should be conditional on the ASDE undertaking ongoing duties. These should include:

(1) assuring their AV will continue to drive safely and in accordance with road rules throughout the lifetime of the vehicle;

(2) disclosing information where required by law or if required to do so as a condition of the authorisation process;

(3) co-operating with the authorisation authority, the in-use regulator and the road collision investigation branch.

The law commissions intend for the sanctions to apply to Manufacturers that have committed an infraction against a GB approval scheme. However, it's important that the in-use regulator be able to identify when safety critical issues may be common across different Manufacturers that may be reliant upon elements for a single software supplier.

For example, the GB scheme is expected for focus on whole vehicle type-approval combining vehicle platform and ADS hardware and software. It is conceivable that the same ADS hardware and software might be fitted to a number of different vehicles from a number of different Manufacturers which have all been individually type-approved. If a safety critical flaw in the ADS has been identified for one Manufacturer but evidence suggested this may impact other Manufacturer's using the same ADS solution then it's is suggested that the in-use regulatory should have the authority to impose sanctions to all affects Manufacturer's.

This process would be similar to the existing vehicle recall system when a fault in a common hardware component, such as brakes or airbags, is detected which affect multiple vehicles individually type-approved by different Manufacturers.

Furthermore, the definition of the Manufacturer allows another scenario whereby an OEM can set up multiple Manufacturers specific to each deployment as a method of reducing their liability and exposure to sanctions, as shown in Figure 14 below. It should be noted that if this is considered a safety issue it could be addressed in secondary legislation.

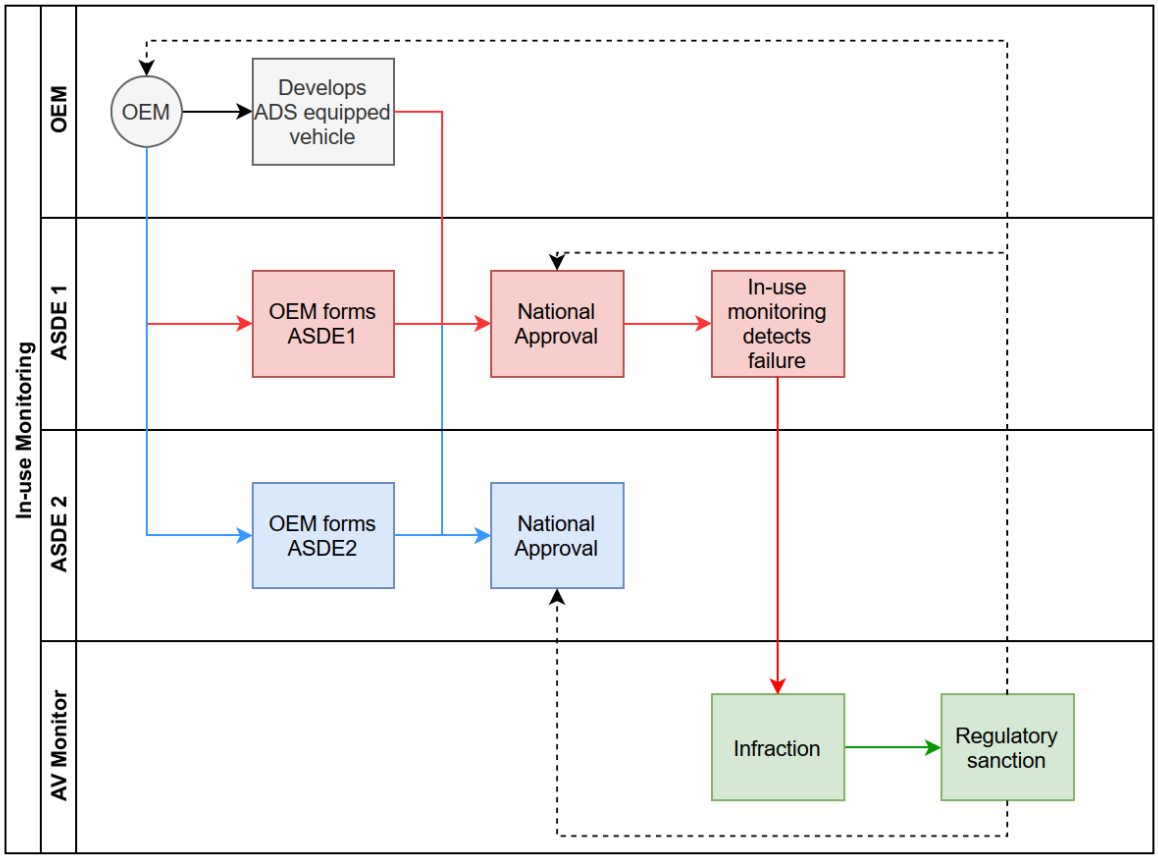


Figure 14: Options for the scope of where sanctions apply

Again, it is necessary in this case to ensure that sanctions and restrictions can be applied across Manufacturers where necessary in order to adequately control risk.

11 In-use monitoring framework

This document has aimed to collate all potential requirements of an in-use monitoring framework and then assess and propose possible approaches to implement it. Sections 0-0 of this report outlines the various elements of a monitoring framework and provides recommendations for how each element may be addressed in practice. Bringing all elements together, Figure 15 below proposes a framework for in-use monitoring for GB AV approval.

A high-level summary of the framework is given below to support readability of the diagram and refer the reader to relevant supporting text:

Pre-deployment:

12. The In-Use Regulator sets out the minimum requirements for monitoring that the Manufacturers and Operators must ensure compliance against prior to deployment (as part of approval requirements). A set of minimum in-use monitoring requirements has been proposed in the Minimum Dataset Specification report (Chapman and Perren, 2021).
13. The Manufacturer develops their Vehicle Safety Case (VSC) for approval. In the VSC, the Manufacturer should demonstrate compliance with the minimum monitoring requirements and also specify additional monitoring required to ensure compliance with the safety arguments stated within the VSC (See Section 7.1.1)
14. The Manufacturer's SMS determines their ability to conduct in-use monitoring and develop their VSC robustly. In order to ensure compliance, the Approval Authority (possibly supported by the In-Use Regulator) should conduct an audit of the Manufacturers' SMS prior to approval and periodically during deployment. Key recommendations regarding SMS requirements for in-use monitoring and Audits are given throughout this report where relevant and collated in the following section (Section 12).
15. The vehicle safety and security monitoring capability is assessed at type approval. Following/ as part of approval, the Manufacturer develops an operational manual that outlines safe use of the AV. It should also set out operational monitoring requirements specific to the AV. This document will be used by the AV operator to establish their monitoring responsibilities and define them within the deployment approval (and authorisation) step, in a Deployment Safety Case (DSC). Key monitoring responsibilities proposed to be overseen by the Operator are outlined in Section 7.1.3.

In-Service

16. When in-use, the Manufacturer and operator enact their responsibilities for in-use safety monitoring. The primary mechanism for monitoring is through the identification and subsequent data collection of unsafe events. The Road Incident Taxonomy Report for this project (Reed, N., 2022) defines and classifies events within scope of the scheme. The Minimum Dataset Specification report (Chapman and Perren, 2021) outlines the minimum set of leading and lagging measures used to identify events of interest.

17. In addition to the minimum data set, there are a number of additional mechanisms for event detection to account for the limitations in the ability of vehicle's event-based data capture. These are outlined in section 0.
18. Upon identification of an event of interest, it is necessary for the Manufacturer to investigate the event. Data recall requirments to enable investigation are outlined in Section 0. This involves identifying an actual risk event or an infraction occurred, classifying it (in line with the Road Incident Taxonomy Definitions (Reed, N., 2022)) and investigating it sufficiently to identify causal factors.
19. If the event was a severe event requiring immediate action (i.e. a collision involving injury), this should be immediately reported to the In-Use Regulator. It will also trigger the post-incident response to ensure police, regulator and independent investigating authority response is coordinated effectively as required. This is defined in the Post-Incident Response Framework report (C. Arnold, 2022) for this project. In addition to this, if the event show indicates that VSC/DSC has been invalidated, this should also be reported to the regulator immediately. Requirements for immediate reporting are outlined in Section 9.1.1.
20. For all other monitoring data, reports are expected to be made to the regulator periodically. This is to establish trends over time which could further indicate issues with the AV deployment and identify any actions required by the regulator to investigate further. The proposed requirements for periodic reporting and aggregated data analysis by the Manufacturer are provided in Section 9.1.4 and 9.1.5. The process for In-Use Regulator Analysis and monitoring of AV safety performance is provided in the Outcome Reporting report for this project (Reed *et al.*, 2022)
21. If the In-Use Regulator establishes non-compliance, they can take action against the Manufacturer or the operator (as required). For non-ADS issues, we propose this is handled by existing the Market Surveillance capacity of the DVSA. For ADS issues, the In-Use Regulator should establish the most appropriate course of action. Where the regulator was misled by the Manufacturer or operator (e.g. through misrepresentation or withholding of data), then criminal prosecution may be sought. For all other cases, the range of regulatory sanctions proposed by the Law Commissions (Law Commission & Scottish Law Commission, 2022) can be used. Considerations for the proportionate and fair application of sanctions are discussed in Section 0. The aim of the sanction is to initiate remedial or restorative action to improve AV safety and maintain compliance with AV approval and authorisation requirements.
22. Where there is non-compliance, but as a result of a potential deficiency in approval and authorisation requirements, this should instigate a review of potential changes to the requirements to close the gap and improve public safety. The Change Control Process report for this project outlines a recommended method for enacting considered, positive changes to the AV safety assurance scheme (Perren, 2022).

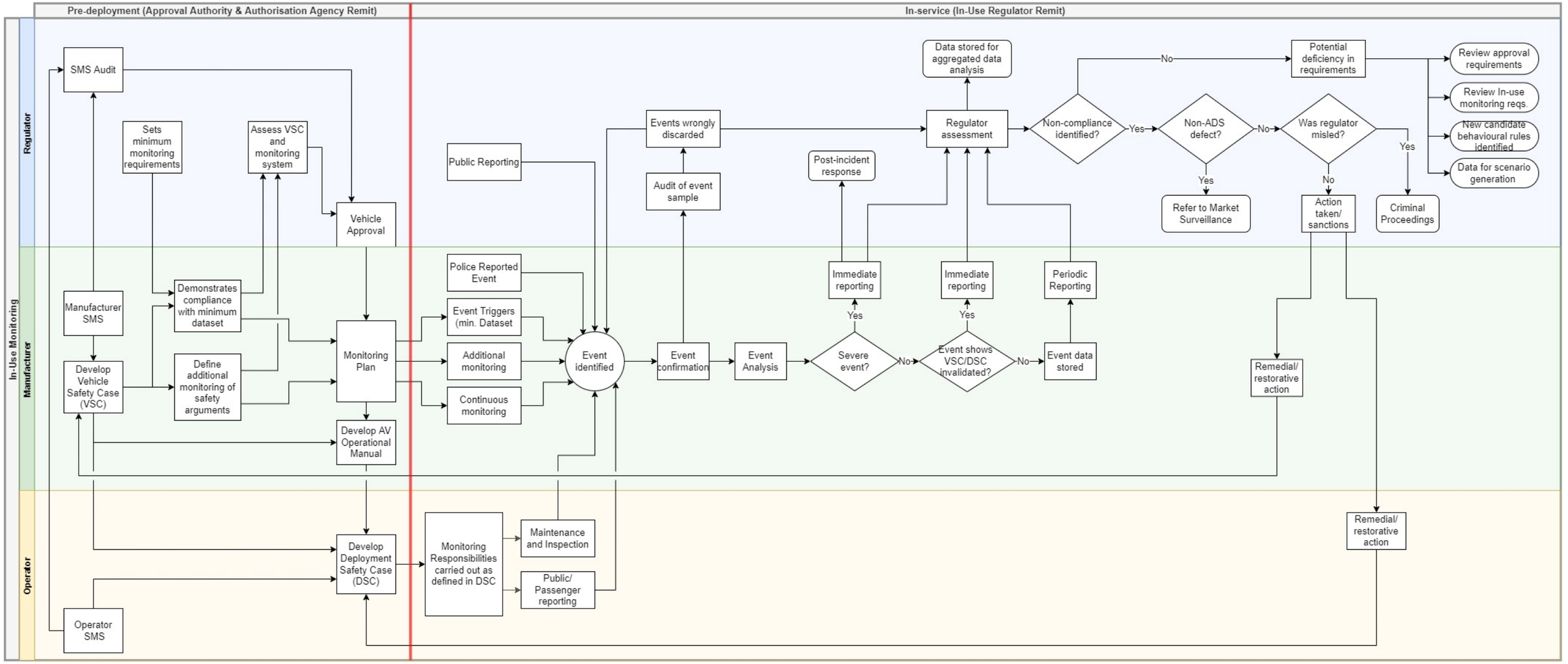


Figure 15: Process outline for the in-use monitoring framework

12 Conclusions and Key Findings

This report has reviewed the current approaches for in-use monitoring for automated vehicles, as well as similar approaches for other transport domains and in other industries. This report has also collated a complete list of requirements and objectives for in-use monitoring to achieve as set out by various stakeholders such as the Law Commissions, the DfT and the public. Following this, potential options were assessed for how key elements of an in-use monitoring scheme can operate in practice to meet these requirements considering:

- Event detection;
- Data recall and investigation;
- Event reporting; and
- Regulatory sanctions.

In summary, the key findings and recommendations are:

- The Law Commission recommend that the Secretary of State for Transport publishes a safety standard by which the safety of automated and conventional driving should be compared. A new AV in-use regulator would then be responsible for collecting the required data.
    - In-use monitoring data will be required to provide evidence that AVs fleets achieve the same safety performance as human driven fleets. These comparisons will be more accurate when there is access to common safety data between fleets. Existing voluntary fleet operator schemes, such as FORS, provide a useful reference for the type of data available for human driven fleets and which therefore might also be expected of AV fleets.
- Many ADS developers with deployed systems are already conducting some form of in-use monitoring and using the data collected to communicate with stakeholders and the public to assure confidence. Much of the data already being collected is the same as what is required to fulfil the objectives of this scheme. The is a perception that developers, Manufacturers or future Manufacturer's may not be able, or may be unwilling, to share data with the in-use regulator. This document highlights the importance of this data for maintaining public safety, growing trust in AVs and accelerating their deployment. The benefits proposed are aimed at making the proposed scheme attractive with limited burden on the Manufacturer to implement.
- It is inevitable that there will be instances where that collision avoidance fails and other situations where a collision is truly unavoidable. In these instances, the residual risk for not detecting a collision is high. If the AV fails to detect a collision it will not initiate the appropriate response (MRM, or E-stop, etc.) which could result in increased consequence severity and potential for secondary collisions before intervention. Collision detection should itself become a safety goal which much be argued in order to meet the definition of safe driving.
- Much of the data required by the proposed scheme is already being collected and utilised for existing processes such as telematics-based insurance models and fleet

monitoring. This suggest that the scheme, although novel in approach, does not rely on data that cannot feasibly be collected, highlighting the practicability of the scheme. Additionally, for AVs, data which provides evidence to the circumstances and situations under which the OEDR/DDT was performed at the time of an incident should also be readily available from the software. The most valuable source of data in this regard has been identified as the 3D world model representation used by the ADS for decision making.

- The hazard analysis has shown a high degree of coverage of the currently proposed leading and lagging measures to detect hazards that are credible for LSAV uses cases in scope of this project. It did however identify some gaps, primarily concerning events where there was some failure in perception meaning the AV is unaware that an event has occurred. A number of other potential measures have been suggested which could help to further identify these types of errors. There is however expected to be residual risk from not detecting hazards which have the potential to predict harm arising in the future. It is suggested that monitoring that is independent of the real-time system under test (either through independent technical solutions or operational processes) would be the only way to manage this risk. Several external monitoring options have been discussed that Manufacturers and Fleet Operators should consider.
- Outside of the minimum dataset specification, it is recommended that in-use monitoring plans developed and evidenced during approval include measures and methods to monitor that validity of the safety arguments and assumptions in the safety case on which the approval was based. This is in line with current best practice safety management systems. Any evidence that violates the validity of the safety case must be reported to the regulator immediately. Conversely, in-use monitoring should also demonstrate continuing compliance with type approval through periodic reports sent to the regulator. Reporting methods have been proposed for both modes of reporting.
- The in-use regulator will have statutory duties and powers to maintain in-use safety once AVs are deployed on GB roads. This includes having the powers to collect relevant data from Manufacturers and Operators. Trust between the Manufacturer/Operators and the in-use regulator will be essential for creating a no-blame safety culture. Self-reporting and the collection of in-use monitoring data should be encouraged by a proportionate approach to sanctions. There is a clear incentive for Manufacturer/Operators to share data in order to enable the in-use regulator to apply sanctions proportionately to the breach and ongoing safety risk. Failure to provide data would likely result in insufficient evidence for the in-use regulator to make safety critical judgements and would likely result in higher levels of sanctions to protect public safety e.g. suspension of operations (the grounding of operating fleets) might be the only appropriate response to a collision event where the Manufacturer/Operators cannot provide evidence to the contrary.
- In-Use Monitoring must be underpinned by a robust Safety Management System. Recommendations for Manufacturer and operator SMS are collated below:

Table 6: Recommendations for Safety Management Systems to support in-use monitoring

Manufacturers should be encouraged to identify further measures and other monitoring approaches to ensure tolerable coverage and evidence this as part of their type approval safety case.
In line with current best practice, it is recommended that the Manufacturer develops an in-use monitoring plan that is evidenced within the safety case. The in-use monitoring plan should evidence that there are processes in place to collect and investigate data in line with the minimum dataset specification as well as any additional processes required to monitor continued compliance with the safety case.
Ultimately the regulator will require evidence to support in-use monitoring investigations. This data may be made available through continual monitoring, through reliable event -based capture (though this is not expected to be fully reliable, at least initially), or through other operational means such as a steward in the vehicle. The data collected may also vary. While continual monitoring is recommended, it is thought that the best approach would be for the Manufacturer to define the appropriate level of continual data collection in order to control residual risk. This decision would be based on the deployment context, the Manufacturers Safety Management System (SMS), and their own cost-benefit analysis; balancing data storage and their risk tolerance for the exposure to claims and sanctions. This would be expected to be evidenced as part of their in-use monitoring plan within the SMS and the Safety Case.
It is envisaged that the responsibility for interpreting AV safety data to determine whether there is potential non-compliance with type approval sits with the Manufacturer. This would need oversight by the regulator. This could be achieved through the audit of the Manufacturer's Safety Management System to ensure there are proper monitoring, investigation and reporting processes in place to provide confidence.
Individual event reports should be underpinned by case study analysis. The approach for investigating and analysing events should be detailed within a Manufacturer's Safety Management System. The event reports should display sufficient information required for the regulator to determine whether the event violates the safety performance on which the type approval was based.

Taking into account the full recommendations made in this report, an in-use monitoring framework has been proposed.

13 References

Amodo 2022, *PHYD and PAYD: UBI most popular product models.*, <https://www.amodo.eu/en/blog/phyd-and-payd-ubi-most-popular-product-models>.

Aptiv 2021, *What Is a Driver-Monitoring System?*, <https://www.aptiv.com/en/insights/article/what-is-a-driver-monitoring-system>.

Automated Vehicle Safety Consortium (2021). *Best practice for metrics and methods for assesing safety performance of automated driving systems (ADS).*. SAE Industry Technologies Consortium.

Automotive World 2022, *The evolution of insurance telematics.*, <https://www.automotiveworld.com/articles/evolution-insurance-telematics/>.

AVS 2022, *AVS: Autonomous Visualization System.*, <https://avs.auto/#/about>.

BCLP Law (2019). *ENFORCEMENT UNDERTAKINGS: THE SHIFTING LANDSCAPE OF ENVIRONMENTAL CRIMINAL LIABILITY*. BCLP Law.

BSI (2020). *PAS 1881:2020 Assuring the safety of automated vehicle trials and testing – Specification*. British Standards Institute.

BSI (2020). *PAS 1883:2020 Operational Design Domain (ODD) taxonomy for an automated driving system (ADS) – Specifi cation*. British standards Institute.

BSI (2021). *A review of CAV safety benchmarking and a proposal for a "Digital Commentary Driving" technique*. British Standards Institute: London.

C. Arnold WP (2022). *GB Safety and Secturity Approval Scheme: Post event investigation framework*. TRL: Crowthorne, Berkshire.

CAA (2020). *Civil Aviation Authority: Regulatory Enforcement Policy*. Civil Aviation Authority.

CAA (2021). *Airspace Infringements: review and actions process*. Civil Aviation Authority.

Caesar H, Kabzan J, Tan KS, Fong WK, Wolff E, Lang A, Fletcher L, Beijbom O and Omari S (2021). *nuPlan: A closed-loop ML-based planning benchmark for autonomous vehicles*. Motional.

Caesar, Holger 2021, *Technically Speaking: Auto-labeling With Offline Perception.*, <https://motional.com/news/technically-speaking-offline-perception>.

Chapman S and Perren W (2021). *In-use monitoring dataset specification (draft)*. TRL.

Chen X, Lisee J, Wojtaszek T and Gupta A 2019, *Introducing AVS, an Open Standard for Autonomous Vehicle Visualization from Uber.*, <https://eng.uber.com/avs-autonomous-vehicle-visualization/>.

DeGould 2022, *Powered by AI: Automated vehicle damage detection and image augmentation.*, <https://degould.com/powered-by-ai/>.

DfT (2011). *STATS19 road accident injury statistics – report form*. Department for Transport.

DfT (2015). *Consultation on civil sanctions for the Civil Aviation Authority*. Department for Transport.

DfT 2021, *Road Safety Data.*, <https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data>.

DfT 2022, *Vehicle Market Surveillance Unit: Enforcement Policy.*, <https://www.gov.uk/government/publications/how-dvsa-makes-sure-businesses-make-or-sell-safe-vehicles-or-parts/vehicle-market-surveillance-unit-enforcement-policy>.

DVSA (2021). *Vehicle Market Surveillance Unit: Enforcement Policy*. UK Government: London.

FORS (2016). *CLOCS Toolkit: Managing collision reporting and analysis*. FORS.

FORS (2021). *Fleet Operator Recognition Scheme Standard V6.0*. FORS.

FORS 2021, *FORS Collision Manager.*, <https://www.fors-collision-manager.org.uk/>.
FORS 2021, *What is FORS and why is it for you?*, <https://www.fors-online.org.uk/cms/>.
ISO (2018). *ISO 45001:2018 Occupational health and safety management systems — Requirements with guidance for use*. International Standards Organisation.
ITU-FGAI4AD (2020). *The Molly Problem: Public Survey Results (preliminary)*. International Telecommunications Union.
Law Commission & Scottish Law Commission (2021). *Automated Vehicles: Summary of responses to Consultation Paper 3 and next steps*. Law Commission and Scottish Law Commission.
Law Commission & Scottish Law Commission (2022). *Automated Vehicles: joint report*. Law Commission: London.
LexisNexis (2018). *Moving Telematics-enabled UBI from Niche to Mass Market*. LexisNexis.
Life Insurance International 2021, *One third of UK SMEs with commercial vehicle insurance have a usage-based insurance policy.*, <https://www.lifeinsuranceinternational.com/comment/one-third-of-uk-smes-with-commercial-vehicle-insurance-ubi/>.
Matt Vitelli YCYYMWBOMNHGQHAJPO (2022). *SafetyNet: Safe planning for real-world self-driving vehicles using machine-learned policies*. Cornell University.
Motional 2022, *NuPlan: world's first benchmark for autonomous vehicle planning.*, <https://www.nuscenes.org/nuplan>.
Natwest 2020, *Telematics and your fleet - Big data and modern communications can transform the management of vehicle fleets, if managers focus on results, not data.*, <https://natwestbusinesshub.com/articles/telematics-and-your-fleet>.
Nexar 2021, *Get the insight to make better decisions: AI-based first notice of loss and collision reconstruction.*, <https://data.getnexar.com/solution/car-insurance/>.
NHTSA 2021, *Standing General Order on Crash Reporting for Levels of Driving Automation 2-5.*, <https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting-levels-driving-automation-2-5>.
Nissan 2019, *Seamless Autonomous Mobility (SAM): The Ultimate Nissan Intelligent Integration.*, <https://global.nissannews.com/en/releases/release-38d144e67f3bedef1b961fff830f08e9-seamless-autonomous-mobility-sam-the-ultimate-nissan-intelligent-integration>.
NTSB (2019). *Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian*. National Transportation Safety Board.
Peng H (2019). *Mcity ABC test: A concept to assess the safety performance of highly automated vehicles.* Ann Arbor: University of Michigan..
Perren W (2022). *GB AV Approval: In-Use Safety And Security Monitoring - Change Control*. TRL: Crowthorne.
Police.UK 2022, *Reporting a crime or incident.*, https://www.police.uk/pu/contact-the-police/report-a-crime-incident/, <https://www.police.uk/pu/contact-the-police/report-a-crime-incident/>.
RAND (2018). *Measuring Automated Vehicle Safety - Forging a Framework*. RAND: Santa Monica, California.
Reed N (2021). *In-use monitoring taxonomy (draft)*.
Reed N, Perren W, Kourantidis K and Simpson B (2022). *GB AV Approval: In-Use Safety and Security Monitoring - Outcome Reporting*. TRL: Crowthorne.
Reed, N. (2022). *GB Safety and Security Approval Scheme: In-use monitoring taxonomy*. TRL: Crowthorne, Berkshire.

SAE (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Society of Automotive Engineers.
Smart eye 2020, *Driver Monitoring (DMS) on its way to becoming mandatory in vehicles around the world.*, <https://smarteye.se/blogs/driver-monitoring-dms-on-its-way-to-become-mandatory-in-vehicles-around-the-world/>.
Tesla 2022, *Safety Score Beta.*, <https://www.tesla.com/support/safety-score>.
Tesla 2022, *Tesla Vehicle Safety Report.*, <https://www.tesla.com/VehicleSafetyReport>.
Underwriter Laboratories (2020). *ANSI/UL 4600 Standard for Safety for the Evaluation of Autonomous Products*. ANSI.
UNECE (2012). *Risk Management in Regulatory Frameworks: Towards a Better Management of Risks*. United Nations: New York and Geneva.
VMAD SG3-14-14 (2021). *Draft requirements for In-Service Monitoring and Reporting for the Automated Driving System (ADS)*. UNECE.
WP.29 (2019). *Framework document on automated/autonomous vehicles*. UNECE.

Appendix A Law Commissions Consultation Paper 3 Responses

Consultation Question 18 (Paragraph 10.83)

18.18 We provisionally propose that the enhanced scheme should give regulators the following responsibilities and powers:

*(1) scheme regulators should be responsible for **comparing the safety of automated and conventional vehicles** using a range of measures;*

(2) to do this the regulator should have power to collect information on:

*(a) **leading measures** (instances of bad driving which could have led to harm) and*

*(b) **lagging measures** (outcomes which led to actual harm);*

(3) regulators should have power to require an ADSE:

(a) to update software where an update is needed to ensure safety and continued compliance with the law;

(b) to keep maps up-to-date, where an AV relies on maps to ensure safety and compliance with the law;

(c) to communicate information about an ADS to users in a clear and effective way, including where necessary through training.

Do you agree?

Consultation Question 22 (Paragraph 11.24)

18.22 We provisionally propose that a statutory scheme to assure AVs in-use should:

*(1) investigate **safety-related traffic infractions** (such as exceeding the speed limit; running red lights; or careless or dangerous driving);*

*(2) investigate other t**raffic infractions**, including those subject to penalty charge notices;*

(3) if fault lies with the ADSE, apply a flexible range of regulatory sanctions.

Do you agree?

Question 23 (Paragraph 11.53) which was;

*18.23 We provisionally propose that the regulator which assures the safety of AVs in-use should have powers to impose the following **sanctions** on ADSEs:*

(1) informal and formal warnings;
(2) fines;
(3) redress orders;
(4) compliance orders;
(5) suspension of authorisation;
(6) withdrawal of authorisation; and
(7) recommendation of attendance at a restorative conference.

Do you agree?

The Law Commissions response summary;

*•Q18 Respondents **widely supported the collection of data on AV performance for in-use monitoring** (Q18) This extended to both **leading measures (instances of bad driving which could have led to harm)** and **lagging measures (outcomes which led to actual harm)**. Many emphasised the importance of choosing appropriate measures, which would not necessarily mirror some of the classic leading measures used to indicate "bad driving" in humans. Instead new measures would be needed, like the **frequency of emergency manoeuvres** or **unstable lateral positioning within lane**. We understand that Government is considering these issues as part of the CAVPASS programme.*

*• Q22 & Q23 In Chapter 11 we considered two challenges. The first was how to deal with breaches of traffic rules. The second was how to learn from collisions so as to promote a safety culture. In both cases we proposed a move away from the current emphasis on the criminal prosecution of human drivers. Instead, we proposed that the in-use safety assurance scheme should investigate breaches of traffic rules by AVs driving themselves and apply a flexible range of regulatory sanctions on ADSEs. **There was broad agreement on all the policies we put forward in this chapter.***

Specific responses for more context;

Five AI Responses

Q18

*Care will be needed to select leading measures that are relevant to AVs. Examples might include state changes such as the **frequency of transition to a MRC**, the u**se of an Emergency Manoeuvre** (as specified in the ALKS regulation), or **unstable lateral positioning within lane**.*

*Accordingly, **work will be required to define leading measures** that are meaningful indicators that the ADS is not operating as intended. An **ADSE/ADS developer could be encouraged to specify in their safety case what leading/lagging measures they will monitor** as part of their safety monitoring, with the r**egulator then having the power to ensure the ADSE then monitors those measures**.*

Q22

We agree in principle with (1)and (2). However, the focus should be on ***investigating infractions that have a meaningful bearing on the safety of the AV*** *(see further our response to Consultation Question 18 above),* ***or are issues of wider public interest****, rather than automatically focusing on infractions that are meaningful for human drivers. This area would benefit from further research to determine both which infractions are meaningful and which have most impact, to ensure best use of finite resources and to enable resources to be allocated appropriately between investigating different types of infractions.*

In regard to (3), we agree with the qualification that ***regulatory sanctions should only be applied if the "fault" is an indicator that there is an overall issue with the safety of the AV****.*

Q23

Any sanctions should be proportionate and applied fairly and consistently. Like in the aviation sector, the focus should be on improving safety and access to the benefits of self-driving. For this purpose, the regulatory system should encourage collaboration with the regulator and ***self-reporting****, encouraging finite resources to be* ***focused on what can be learnt from the incident and implementing that****, rather than diverting resources to contesting liability and self-defence.* ***When assessing proportionality, the focus should be on the conduct of the ADSE, the nature of the failure and the level of risk posed by it, rather than the actual result of the failure.*** *This is to reflect the necessary move away from the human centric rules and regulations that apply to human drivers to a product safety led system. For example, a failure that leads to a material deterioration in an AV's ability to drive safely in the rain, a relatively common occurrence, that has by fortune not caused any fatalities at the time of discovery, should be viewed more seriously than an AV's failure to deal appropriately with a very rare event that unfortunately leads to multiple fatalities in a single event.*

Oxbotica Responses

Q18

(1) & (2) ***The type and source of this data needs to be explicitly defined.*** *What data is being requested, and where is it expected to come from? Any data-sharing requirements should not cause an increased computational burden for the ADSE. It is very expensive for an AV to continuously log data, and impractical to upload it, even over 5G. The data requested by regulators through this enhanced scheme should be consistent with what is requested by international regulation.*

Q22

Yes, we agree.

Q23

We have no strong opinion. Perhaps this should be coordinated by the Traffic Enforcement Centre.

Renault Responses

Q22

No We disagree. In case of a problem (infraction…) We propose that in case of any traffic related infraction, et… the authorities should inform the OEM and should request the defect to be resolved. Sanctions can only be imposed on the OEM when technically this has been proven.

Q23

Agree, but these 7 steps come after the OEM is initially informed

Wayve Responses

Q18

Yes We broadly agree with this suite of tools being available to regulators for autonomous driving. These need to be developed in conjunction with AV developers, which are the only entities likely to have sufficient data and analysis for these powers today. We struggle to see how 3rd parties will be able to assess AVs without the deep understanding of AVs and the scale of data to support decisions. This suggests to us that ADSEs will need to collaborate closely with regulators and to some extent each other. On 1. ***We see risk in comparing autonomous vehicles to conventional (human) driven vehicles****. Top level population statistics have some merit, but we caution against granular comparisons.* ***We do not see merit in deeper comparisons between human-driven and autonomous vehicles for regulatory purposes****, though these have merit in supporting public acceptance. On 2. As an AV Developer* ***we would like to work with regulators to define and monitor leading indicators of safety in particular.***

Q22

Yes We support this as a mechanism to create a transparent no-blame culture for safety in use.

Q23

We agree that each of these could have a place in a regulator's toolkit, where measures are gradual and proportional. We suggest the approach taken by this regulator should encourage mutual learning in the industry during early AV deployments.

SMMT

Q18

40. We agree in principle with the above proposals.
41. As regards proposal (1). Comparing the safety of automated and conventional vehicles using a range of measures is integral to the principle of a positive risk balance that we support, as set out in paragraph 15 above. However, the range of measures and their metrics must be clearly defined and agreed with industry and stakeholders.
42. As regards proposal (2), ***we urge alignment with the data elements currently being discussed at VMAD.*** *Data on lagging measures is better understood and can be more easily obtained as it depends on actual outcomes, e.g. accidents. However, data on leading measures, as well as "bad driving", must be properly defined, the process for recording and*

accessing it clearly specified, and the handling and processing of the data compliant with data protection laws. We observe three concerning examples:

- *Paragraph 10.70 in the consultation paper suggests disengagement as one potential leading measure. Disengagement is often recorded as part of trials to better understand the context within which either the automated driving system (ADS) hands back control or the safety driver unilaterally decides to retake control. If recorded in deployed automated vehicles, it could unfairly penalise a perfectly-operating ADS but where the user is extremely risk averse and often unnecessarily retakes control.*
- ***Near-misses**, as suggested in paragraph 10.67 in the consultation paper, **to the extent that it is reasonable and feasible to collect such data, will also need to be contextualised.** A wellperforming ADS in an urban setting may have recorded an usually high number of nearmisses but no accidents because it has successfully avoided swerving cyclists and pedestrians carelessly stepping out into the road. **Data on near-misses must also not be transferred to any entity outside the regulator to prevent misuse**; for example, as a basis for insurers to increase insurance premiums.*
- *Depending on how it is implemented, "placing unobtrusive sensors on conventional vehicles", as suggested in paragraph 10.71 in the consultation paper, may not be entirely compatible with privacy laws and may only exacerbate society's resentment against surveillance by authorities.*

Q22

54. We agree in principle with the above proposals.

*55. There are clearly benefits to proposals (1) and (2), mainly the **information on traffic infractions can be utilised for safety recommendations**. However, the proposals are imprecise as to who has the investigatory power (e.g. the agency performing in-use assurance itself, the police), what data they have access to, whether the automated driving system entity (ADSE) would have the opportunity or an obligation to support the investigation, whether there would be an opportunity for the ADSE to see the data, and whether the ADSE could make submissions before facing a regulatory sanction.*

*56. In relation to **leading measures**, on which we have raised some concerns in paragraph 42 above, **certain technical complexities with in-use monitoring data capture by the automated driving system (ADS) must first be resolved**. For example, unless it results in an accident (a lagging measure) or is reported by road users, **an ADS that ran a red light would not "know" it has committed a traffic offence if it could not recognise the light was red in the first instance.***

*57. As such, we suggest investigations on **safety-related traffic infractions** should adopt a collaborative approach involving national and/or local enforcement bodies. The **in-use safety assurance agency, or the police, should inform the ADSE of the traffic infractions, present the supporting evidence and request the identified problems to be resolved.** This does not mean the ADSE could avoid being issued a penalty charge notice if the traffic offence is proven to be the fault of the ADS. Failure, or continuous failure, to resolve the identified problems should result in appropriate, or escalating, sanctions.*

Q23

59. We agree with the above proposed sanctions so long as:
*The **automated driving system entity must first be informed of the offence**(s) and be shown the evidence;*
*• Provision has been made for **self-reporting, collaborative investigation, resolution of the identified issues** and, if necessary, product recall; and*
*• The **sanction is proportionate to the offence and its consequences**, and*
escalated appropriately, with the severest, i.e. (5) and (6), reserved for cases of gross negligence or for serial offenders.

ROSPA

Q23

*RoSPA agrees with these proposals. **These measures will be essential to allow automated vehicles to operate safely.** The collection of data, to allow the regulator and Manufacturer to **act quickly if things go wrong, will be vital.** This should include collecting data on lagging and leading measures. While **lagging measures (such as counting casualties) provide the most accurate reflections of safety, they are rare events and, by definition, have resulted in harm**. By contrast, **leading measures (such as failures to follow road rules or "near-miss" events) can act as warnings.***

Appendix B Fleet Operator In-use safety frameworks

The voluntary Fleet Operator Recognition Scheme (FORS) provides an accreditation framework for managing Work Related Road Risk (WRRR) within a culture of continual improvement. The scheme currently has 4,900+ accredited members operating 103,000+ accredited vehicles to either Bronze, Silver or Gold standard (FORS, 2021).

FORS covers all vehicle types operating on public highways (Heavy Good Vehicles, Vans, Passenger Carrying Vehicles, Cars, powered two-wheelers) and all types of drivers (defined as a person driving or riding any vehicle for an organisation that is in scope of FORS accreditation).

There are four key areas to the FORS standard; Management, Vehicles, Drivers and Operations. Each area has a set of Requirements (Legal compliance, Safety, Efficiency, Environment, Security and counter terrorism), a Purpose for a required outcome and Demonstration details to show how the Requirement is to be met

Operators Requirement O3 specifies the need to "document and investigate road traffic collisions, incidents and near-misses". The Purpose of O3 is to "determine the contributory and root causes of road traffic collisions, incidents and near-misses to prevent recurrence and minimise road risk." FORS Operators shall Demonstrate they have "a policy and supporting procedures in place to record and investigate road traffic collisions, incidents and near-misses."

Silver accreditation Requires the FORS operator to "actively monitor and manage operational performance" with the Purpose to "improve operational performance, reduce costs and minimise impact on the environment". The following operational performance indicators should be reported for all vehicles under FORS accreditation;

- Total distance travelled and fuel used by vehicle type
- Total road traffic collision and incident data by vehicle type
- Total Reporting of Injuries, Diseases and Dangerous Occurrences Regulations (RIDDOR) reportable incident data
- Total transport related fines and penalty charges

Data capture and monitoring tools **should** include:

- Fleet management system
- Vehicle telematics
- Insurance reports
- Fuel card reports
- Penalty Charge Notice (PCN) management

Additionally, road traffic collision and incident data should be entered into the FORS Collision Manager tool (FORS, 2021). The tool provides "industry transparency relating to work related road safety enabling you to capture, report, investigate and analyse incidents, collisions and near-misses" including;

- Anonymous industry benchmarking

- Positive Corporate Social Responsibility
- Access to shared learning
- Compliance with CLOCS, FORS, WRRR
- Central database to help inform policy
- Integration with insurance companies
- Platform for monitoring collision KPIs

The Construction Logistics and Community Safety (CLOCS) Standard (FORS, 2016) aims to "ensure the safest construction vehicle journeys" and achieve this by placing "responsibilities and duties on the regulator, the client, the principal contractor controlling the construction site and the supply chain including the operator of any road-going vehicles servicing that project".

In the event of a collision the primary responsibility of the driver is "to help co-ordinate activities to make the scene safe, assess the well-being of any party involved and summon assistance from the appropriate emergency services". Afterwards the driver is expected to "capture collision data" which is that start of the CLOCS collision management and reporting requirements (FORS, 2016).

CLOCS reporting requirements can be met by using the FORS Collision Manager where incident types are categorised as; Near Miss, Damage Only, Personal Injury Only or Damage & Personal Injury. Private data, such as Vehicle Registration Number remains hidden and encrypted to facilitate anonymous benchmarking against peers.

Incidents are given a unique ID and can be filtered to view the FORS operator itself, similar FORS Operators or all Operators. Results can be sorted based upon incident type, date and plotted on a map. Dashboards and reports over specific time periods can be used to analyse KPIs and trends.

Incidents can be grouped by Public and Privately defined contracts and clients. While it's also possible to link incidents to insurance claims to simplify claims management. The types of data CLOCS requires for reporting are specified in Table 7 below.

Table 7: Data required for CLOCS reporting

Data	Why it is important to collect this data
Incident type	Whether the incident resulted in a KSI, damage to vehicle only or near miss will inform to a significant degree the resultant actions required. Any collision could result in criminal proceedings if offences are disclosed or identified. All personal injury collisions regardless of severity of injury are likely to involve the police.
Incident date and time	This will enable an understanding of the times of year and day that incidents are occurring
Location	This information is vital in enabling identification of any collision hotspots

Was collision on prescribed route?	Did the driver deviate from the route set by the Transport Manager? If so, it may be that the route the driver selected was inappropriate for the vehicle they were driving. Was a prescribed route provided?
Road type	The type of road is an important consideration. If incidents are occurring on a particular type of road (eg motorway) is specific training required?
Road condition	This will help identify if the road surface was a factor in the collision. Ice, oil or an uneven road surface can all cause a driver to lose control of a vehicle. A cyclist could be adversely affected by road conditions including pot holes and slippery manhole covers.
Road features (eg bus lane, cycle lane etc.)	A lack of segregated facilities requires vulnerable road users to share the carriageway with large vehicles. This may have been a factor in the collision/incident
Road hazards	Hazards on the road can result in a collision. These can include temporary road works or parked vehicles
Road speed limit	Speeding is often a factor in collisions. The speed the vehicle was travelling at should be collected, either from the driver, from on-board systems or the police (who can determine speed by marks left by braking)
Type of junction (if applicable)	If collisions are occurring at a particular type of junction, (such as a roundabout) are measures required to address this?
Signage	If there is a particular hazard at the location of the incident was this appropriately signed? If not then drivers should be warned and it may be appropriate to contact the Local Highway Authority. This is why it is important to report missing or damaged signage, did this affect the collision?
Weather	Weather can significantly impair drivers' ability to operate their vehicle safely. It is vital, in order to develop an understanding of how a collision occurred, that data relating to the weather conditions at the time of the collision is noted
Vehicle details	Information related to the vehicle is vital when undertaking post-collision analysis. The age and body type of vehicles could affect the nature of a collision and its resultant impacts
Vehicle Damage	Information relating to where vehicles were damaged as well as the cost implication help to understand the impact a collision has on your business
Vehicle safety features	If there were safety features such as blindspot cameras or side-proximity sensors fitted to the vehicle this may have prevented a collision occurring. Were they working at the time of the collision?
Vehicle movement	A significant proportion of cyclist fatalities resulting from a collision with an HGV occur when the vehicle is turning left and the cyclist is in the vehicles' blindspot. In order to determine if the direction the vehicle was manoeuvring was a factor in the collision

Driver details	Details of the driver are a critical element of any data collected. This could determine whether the driver is licensed to drive the type of vehicle they are operating, whether they require glasses and are wearing them and whether they were wearing a seatbelt
Third party involvement	Details on any other road users are vital in ensuring an understanding of how a collision occurred and the severity of any impacts. VRUs are often less visible and have less protection afforded to them. This means that a collision with an HGV is more likely to result in a KSI.
Causality and outcomes	The individual tasked with investigating the collision should assess how they think the collision was caused (eg impairment or distraction). Should there be a prevalent cause of collisions at an organisation this could be addressed by re-routing vehicles away from a particular 'problem' location or fitting safety equipment

Reporting also covers types of driver actions being taken at the time of the incident;

Table 8: Description of driver actions required in CLOCS reporting

Driver actions at the time....	In relation to the junction....
Changing lane to left	Approaching junction or waiting at junction approach
Changing lane to right	Cleared junction or waiting/parked at junction exit
Going ahead left hand bend	Emerging from slip road
Going ahead other	Entering main road
Moving off	Leaving main road
Overtaking on nearside	Mid junction - on roundabout or on main road
Overtaking stationary vehicle on its offside	Not at or within 50m of a junction
Parked	Not applicable/available
Reversing	
Slowing or stopping	
Turning left	
Turning right	
U turn	
Waiting to go ahead but held up	
Waiting to turn left	
Waiting to turn right	
Waiting to reverse	
Not applicable/available	

A CLOCS policy for collision reporting should also include an expected timeline of data collection, reporting and recording. For example;

Table 9: Recommended timelines for CLOCS collision reporting

When	Maximum time by which action should be completed	Action
At scene immediately after a collision	As soon as possible once immediate actions have been completed	Driver reports to Transport Manager
At base with information collected from scene	Recommended within 24 hours of incident and ideally prior to the end of shift	Driver reports to Transport Manager
Manager reviews evidence and determines what happened	Recommended within 24 hours of the driver reporting the collision at base	Transport Manager/ Office support
Information logged in Collision Manager	Recommended within 72 hours of incident	Transport Manager/ Office support
Information investigated, analysed and reported	As appropriate	Transport Manager reports to Company Director

It would seem practical, fair and logical that Licensed Fleet Operators for automated vehicles should be required to meet the same standards of safety, sustainability and efficiency as those expected of human driven fleets.

Appendix C FG-AI4AD - Perception, decision, reaction and outcome explainability model

C.1.1.1 The perception, decision, reaction and outcome explainability model

There are a number of tricolons commonly used in describing the high-level tasks required for automated driving;

- Sense, plan, act (common in robotics control)
- See, think, act (as trademarked by ZF[11])
- Perception, decision, reaction (as defined in the UNECE ALKS regulation[12])

The UNECE use of “perception, decision and reaction” refers to the performance model of ALKS where;

Traffic critical scenarios of ALKS are divided into preventable and unpreventable scenarios. The threshold for preventable/unpreventable is based on the simulated performance of a skilled and attentive human driver. It is expected that some of the "unpreventable" scenarios by human standards may actually be preventable by the ALKS system.

In a low-speed ALKS scenario, the avoidance capability of the driver model is assumed to be only by braking. The driver model is separated into the following three segments: "Perception"; "Decision"; and, "Reaction". The following diagram is a visual representation of these segments:

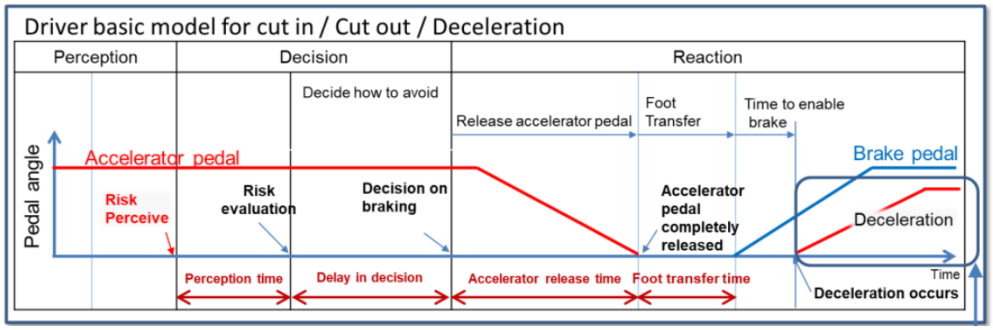


Figure 1 : UNECE skilled human performance model

[11] https://www.zf.com/mobile/en/technologies/see_think_act/see_think_act.html

[12] Appendix 3 page 41 Performance model of ALKS https://undocs.org/ECE/TRANS/WP.29/2020/81

The UNECE model of perception, decision, reaction maps well onto the public responses to The Molly Problem;

Perception *Decision* *Reaction*

Would you expect the software to recall the following?	**Yes**	**Unsure**	**No**
1 - Time of the collision	99.4%	0.4%	0.2%
2 - Location of the collision	98.9%	0.6%	0.4%
3 - Speed at the point of collision	98.5%	0.2%	1.3%
4 - When the collision risk was identified	93.2%	0.9%	6%
5 - If the object representing the child was detected	95.9%	1.5%	2.6%
6 - When the object representing the child was detected	95.9%	1.9%	2.1%
7 - If the object was detected as a human	90.2%	3.4%	6.4%
8 - When the object was detected as a human	88.9%	4.1%	7.0%
9 - Whether mitigation action was taken	97.4%	0.9%	1.7%
10 - When mitigating action was taken	95.9%	1.3%	2.8%
11 - What mitigating action was taken	95.9%	1.3%	2.8%

Table 2 – The Molly Problem – perception, decision, reaction

(data sample from 16/02/2021)

It's clear that the UNECE ALKS regulation embed a performance comparison to "*human driver"* described as *"skilled and attentive"* and elsewhere in the regulations as *"competent and careful"*.

It can be said that to define *"competent and careful"* driving the UNECE ALKS regulations make this comparison for the Object and Event Detection and Response (OEDR) task within the dynamic driving task (DTT).

Comparing the performance of self-driving vehicles with human drivers was raised in the consultation paper[13];

A political decision will have to be made as to the acceptable level of safety of AVs in comparison with human drivers, but in any event AVs should be made as safe as is reasonably practicable.

[13] Consultation Paper 3 page 83, para 5.103

It is also covered in a later section of this feedback[14]; where OEDR comparison is suggested as required for self-driving to human comparison within a safety culture of continual learning.

Explainability of the "perception, decision and reaction" of the ADS to the specific circumstances and situations is important, but so too is the "outcome". Each will be explored further below.

C.1.1.2 Perception explainability

Perception explainability can be split two different viewpoints 'the how' and 'the what?';

- 'the how' explains the software processes used to transform ADS sensor data into a digital representation of the 3D world.
- 'the what', on the other hand, explains what that ADS digital representation of the 3D world contains.

In the context of The Molly Problem 'the what' explains whether the ADS detected Molly, when she was detected, where she was located, where she was moving and the timing of those movements. Whereas 'the how' explains the process by which Molly was detected and tracked such as, the sensors and algorithms used.

'The how' is very specific to the design and architecture of the ADS and is also seen as highly sensitive proprietary intellectual property by the developer.

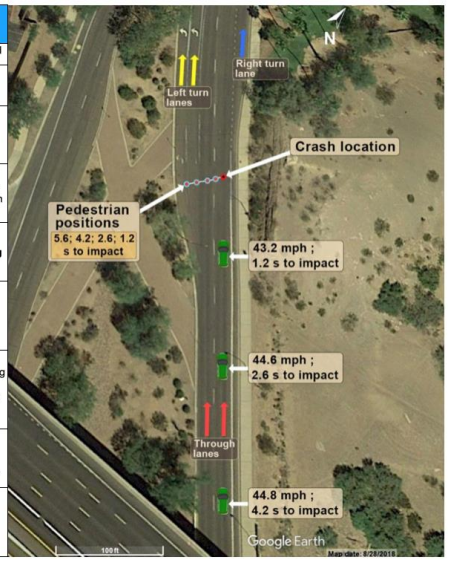
'The what' just the digital representation of the 3D world at a given time and location. It is independent from the ADS architecture and contains no proprietary information. It simply represents the output of the ADS perception system.

'The what' is critical for explainability of the collision or near miss events. It describes where the self-driving (ego) vehicle was located, the relative location other objects and how the ego vehicle and those objects moved over time. It provides the foundation for evaluating OEDR performance.

[14] Section 5 - Comparisons between human and automated driving safety

'The what' can also be used to identify ADS issues in the stability of the 3D world representation. For example, notice the multiple "classification" for the object representing Elaine Herzberg highlighted in NTSB report below[15];

Time (s) relative to impact	Speed (mph)	Classification and Path Prediction[a]	Other Events / Details[b]
-9.9	35		Vehicle begins to accelerate from 35 mph due to an increased speed limit
-5.8	44		Vehicle reaches the speed of 44 mph
-5.6	**44**	**Classification: Vehicle - by radar** **Path prediction: None; not on the path of the SUV**	**Radar makes the first detection of the pedestrian and estimates its speed.**
-5.2	45	Classification: **Other** - by lidar Path prediction: Static; not on the path of the SUV	Lidar detects an unknown object; this is the first detection of that object by lidar, the tracking history is unavailable, and its velocity cannot be determined. ADS predicts the object's path as static.
-4.2	**45**	**Classification: Vehicle - by lidar** **Path prediction: Static; not on the path of the SUV**	**Lidar classifies a detected object as a vehicle; this is a changed classification of an object and without a tracking history. ADS predicts the object's path as static.**
-3.9		Classification: Vehicle - by lidar Path prediction: The left through lane (adjacent to the SUV); not on the path of the SUV	Lidar retains the classification "vehicle", and based on the tracking history and the assigned goal, ADS predicts the object's path as traveling in the left through lane.
3.8 > - 2.7	45	Classification: alternated several times between **vehicle and other** - by lidar Path prediction: alternated between static and left lane; neither were considered on the path of the SUV	The object's classification alternates several times between vehicle and an unknown. At each change, the object's tracking history is unavailable, and ADS predicts the object's path as static. When the detected object's classification remained the same, ADS predicts the path as traveling in the left through lane.
-2.6	**45**	**Classification: Bicycle - by lidar** **Path prediction: Static; not on the path of the SUV**	**Lidar classifies a detected object as a bicycle; this is a changed classification of the object, and without a tracking history. ADS predicts the bicycle's path as static.**
-2.5	45	Classification: **Bicycle** - by lidar Path prediction: The left through lane (adjacent to the SUV); not on the path of the SUV	Lidar retains the classification "bicycle" and based on the tracking history and the assigned goal, ADS predicts the bicycle's path as traveling in the left through lane.



Time (s) relative to impact	Speed (mph)	Classification and Path Prediction[a]	Other Events / Details[b]
-1.5	*44*[c]	Classification: ***Unknown*** - by lidar Path prediction: *Static*; partially on the path of the SUV	Lidar detects an unknown object; since this is a changed classification, and an unknown object, it lacks tracking history and is not assigned a goal. ADS predicts the object's path as static. Although the detected object is partially in the SUV's lane of travel, the ADS generates a motion plan around the object (maneuver to the right of the object); this motion plan remains valid—avoiding the object—for the next two data points.
-1.2	**43**	**Classification: *Bicycle* - by lidar** **Path prediction: The travel lane of the SUV; fully on the path of the SUV**	**Lidar detects a bicycle; although this is a changed classification and without a tracking history, it was assigned a goal. ADS predicts the bicycle to be on the path of the SUV.** **The ADS motion plan—generated 300 msec earlier—for steering around the bicycle was no longer possible; as such, this situation becomes hazardous.** **- Action suppression begins**
-0.2	40	Classification: ***Bicycle*** - by lidar Path prediction: The travel lane of the SUV; fully on the path of the SUV	Action suppression ends 1 second after it begins. The situation remains hazardous; as such, ADS initiates a plan for vehicle slowdown. *An auditory alert was presented to indicate that the controlled slowdown was initiating* [d]
-0.02	39		Vehicle operator takes control of the steering wheel, disengaging the ADS.
IMPACT			
0.7	37		Vehicle operator brakes

[a] Only changes in object classification and path prediction are reported in the table.The last reported values persist until a new one is reported.

[b] The process of predicting a path of a detected object is complex and relies on the examination of numerous factors, beyond the details described in this column

c The vehicle started decelerating due to the approaching intersection, where the pre-planned route includes a right turn at Curry Road. The deceleration plan was generated 3.6 seconds before impact.

[d] *While the system generated a plan for the vehicle slowdown, due to a slight communication delay, the data is unclear on whether the implementation of the slowdown plan started before the operator took control prior to impact.*

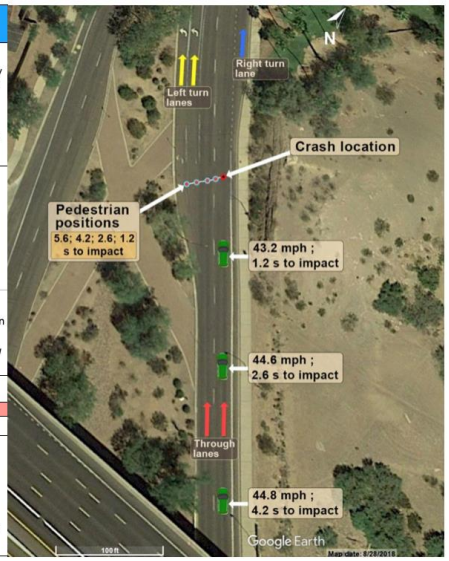


Figure 1 – NTSB HWY18MH010, Tempe, Arizona – Uber ATG

Elaine Herzberg was vulnerable road user crossing the road as a pedestrian while pushing a bicycle. However, before the collision the ADS classified her as a "vehicle, other, vehicle" repeatedly between -5.6 and -2.7 seconds, then bicycle between -2.6 and -1.5 seconds, "unknown" from -1.5 to -1.2 seconds and bicycle up until the impact.

'The what' explainability is capable of highlighting a significant issue with the ADS perception software for which the proposed safety assurance regulatory could issues sanctions including;

[15] NTSB HWY18MH010, Tempe, Arizona - Uber ATG

informal and formal warnings; fines; redress orders; compliance orders; suspension of authorisation; withdrawal of authorisation.

'The what' is valuable to the regulator for in-use safety assurance. Whilst, within the culture of continual learning, the ADSE is responsible continually improving 'the how' to prevent similar future incidents occurring.

It should be noted that 'the what' data is a digital representation of the 3D world which in many ways could be viewed as a compression of the raw sensor data into an encoded data model.

Streaming, monitoring and recording 'the what' encoded data model significantly reduces the network data rates, processing power and data storage required compared to raw sensor data required to evaluate 'the how'.

It is suggested that the safety assurance regulator be granted the powers to mandate the ADSE provide access to ADS digital representation of the 3D world ('the what the ADS saw') and that access includes a provision for real-time data access both onboard or remotely from the vehicle.

C.1.1.3 Decision explainability

Decision explainability provides insights in two safety critical tasks (both highlighted in the UNECE driver skill model); *the risk evaluation* and the selection of *appropriate action to mitigate the risk*.

In terms of the OEDR, the decision-making step plans the "R" (response). It therefore takes the "O" (objects) and the "E" (events) as input which are shared within the digital representation of 3D world model from the perception system.

Response planning requires reasoning and decision-making under uncertainty about both the current and future state of the world.

Uncertainties about the current state of the world are associated with confidence levels in the digital 3D world model e.g. how certain is the ADS perception system about; where the self-driving vehicle is located; where the other objects are located, what those other objects are (their classification) and; the speed and heading of the ego vehicle and the other objects relative to each other and the environment.

In relation to The Molly Problem current state of the world would include; how confident is the ADS in the presence of an object, the size of the object, the location of the object, the speed of the object, the heading of the object, as well as how confident the ADS is that this object is a child pedestrian.

Risk evaluation provides a measure of these uncertainties in the current state of the world. Humans suffer from these uncertainties too caused by variations visual perception such as diminished visual acuity, dazzling by sunlight, fog, rain and road spray. Risk mitigating action, such as slowing down, should planned as the *response* when uncertainties in the current state of the world occur.

Risk evaluation is critical for safety and it is clear from the NTSB analysis of Uber ATG's ADS that if uncertainties in the classification of Elaine Herzberg had been properly identified as a risk then there would have been 2.7-5.6 seconds to execute appropriate mitigating action.

It is suggested that the safety assurance regulator be granted the powers to mandate the ADSE provide access to the levels of uncertainty associated with the ADS digital representation of the 3D world ('the what the ADS saw') and that access includes a provision for real-time data access both onboard or remotely from the vehicle.

Risk evaluation also includes uncertainties about the future state of the world. These uncertainties are associated with confidence level in the ADS predictions. These prediction algorithms are responsible for determining: the level of road friction, the future motion of the other objects, the number of available planned paths and the risk of collision for these available planned paths.

The planned paths are the "response" and in the case of ALKS they may be constrained by design, such as only being able to brake while staying in lane.

Conceivably it would be possible to request the ADS published it's predictions for the future state of the world. These predictions could then be monitored in real-time against the actual state of the world at the end of the planning horizon when the action occur e.g. how close did ADS predict it would come to a pedestrian and how closed did the vehicle actually come.

However, these prediction algorithms are also seen as highly sensitive proprietary intellectual property by the developer as they require significant levels of capital investment to develop. These algorithms, and the future world state prediction that results, represent 'the how' in prediction of *risk evaluation*.

The safety assurance regulatory should again be concerned with 'the what' of *risk evaluation* e.g. what levels of risk were perceived by the ADS and were used to select what the ADS considered the most *appropriate action to mitigate the risk*.

It is suggested that the safety assurance regulator be granted the powers to mandate the ADSE provide access to the levels risk evaluated by the ADS and used to select the mitigating action; and where that access includes a provision for real-time data access both onboard or remotely from the vehicle.

C.1.1.4 Reaction explainability

Reaction explainability provides insight into the ADS *response* and it's *execution*.

Having detected the *objects* and *events*, conducted a *risk evaluation* and *planned mitigating action* all that is left is for the ADS to execute the *response* by controlling the vehicle actuators (brake, throttle, steering, signals etc).

Reaction can be simple. In the UNECE driver skill model, for low-speed ALKS comparison, avoidance capability assumed to be longitudinal deceleration only e.g. take the foot of the accelerator, actuate the brake with a target force and slow down the vehicle at a target rate.

Reaction to a single hazard can also be complex requiring a sequence of longitudinal and lateral inputs that require vehicle control at it's limits of handling. Anti-lock Braking Systems (ABS) and Electronic Stability Control (ESC) assist human drivers execute these complex emergency manoeuvres.

It should also be noted that the dynamic driving task (DDT) itself is a continuous sequence of *reactions* to the changing environment. These types of tasks were described by United States Air Force Colonel John Boyd as a continuous cycle of *observe-orient-decide-act* known as the OODA loop. In the context of the UNECE driver skill model "observe" is the perception phase, "orient-decide" is the decision phase and "act" is the reaction phase.

Reactions are the continuous control inputs and resultant vehicle dynamics in *response* to *circumstances* and *situations*. They are measurable outputs of the perception and decision-making steps which provide valuable insights into driver behaviour and risk. As such, the continual monitoring of reactions is already in use for telematics-based insurance and driver fleet management H&S compliance solutions.

For self-driving vehicles the big advantage is that continual monitoring of ADS *reactions* can be placed in the context of the *circumstances* and *situations* captured from the ADS *perception* as well as the predicted levels of *risk* captured from the ADS *decision-making*.

This combination of *perception*, *decision* and *reaction* in-use monitoring ensures a coherent and complete set of evidence that can be used to determine whether the ADS driving behaviour meets the standards expected of a *competent and careful* human driver.

It is suggested that the safety assurance regulator be granted the powers to mandate the ADSE provide access to the reactions of the ADS related to the OEDR response; and for that access to include provision for real-time data access both onboard or remotely from the vehicle.

C.1.1.5 Outcome explainability

Outcome explainability captures evidence of driving behaviour that results in 'real-life' exposure to harm which is captured by the 1968 Convention of Road Traffic as behaviour likely to ***endanger***.

> *Article 7.1 General Rules*
>
> *Road-users shall avoid any behaviour likely to **endanger** or obstruct traffic, to endanger persons, or to cause damage to public or private property.*

> *Article 21.1 Behaviour of Drivers*
>
> *Every driver shall avoid behaviour likely to **endanger** pedestrians.*

While in the revised version of the Road Traffic Act 1988 captures this under the term ***dangerous***;

> ***Dangerous** driving*
>
> *A person who drives a mechanically propelled vehicle **dangerously** on a road or other public place is guilty of an offence.*

> *…a person is to be regarded as driving **dangerously** if (a) the way he drives falls far below what would be expected of a competent and careful driver, and (b) it would be obvious to a competent and careful driver that driving in that way would be **dangerous**.*

> *…if it would be obvious to a competent and careful driver that driving the vehicle in its current state would be **dangerous**.*
>
> *…"**dangerous**" refers to **danger** either of injury to any person or of serious damage to property; and in determining for the purposes of those subsections what would be expected of, or obvious to, a competent and careful driver in a particular case, regard shall be had not only to the circumstances of which he could be expected to be aware but also to any circumstances shown to have been within the knowledge of the accused.*

Perception, *decision* and *reaction explainability* provides the evidence of "*any circumstances shown to have been within the knowledge of the accused {ADSE/ADS}"*.

Outcome explainability, is the independent assessment the provides evidence of "*the circumstances of which he {ADSE/ADS} could be expected to be aware".*

Outcome explainability takes *perception*, *decision* and *reaction* data as continuous input.

- Using the *ADS perception data stream*, it can determine how close the vehicle actually came to a pedestrian and make an independent assessment on how dangerous the driving was and how endangered was the pedestrian.
- It can compare endangerment to the *risk evaluation* within *ADS decision data stream* to establish if the risks were clearly understood.
- It can assess the timings of *perception*, *decision* and *reaction* to *objects* and *events* to determine if they are *reasonable* and *competent*.
- It can make counterfactual assessments of *perception*, *decisions* and *reactions* using cached recordings of the data streams and running analysis backwards in time.
    - In doing so it can be used to assess actual performance and help verify the assumptions used by "do not cause a fault accident" solutions such as Intel/Mobileye RSS, NVIDIA Safety Force Field etc.
- It can be used to independently assess safety envelope violations, not only for the self-driving vehicle, but the envelopes of other road users which are critical for determining near-miss events.

Outcome explainability, in the context of the Uber ATG NTSB investigation would have enabled:

- The object representing Elaine Herzberg to be tracked backwards in time e.g. from the point of collision to the point of detection

- The retrospective object tracking to then be compared with the *ADS perception data stream* to establish the stability of the object identification, instance identification and object classification.
- The actual trajectory of the Elaine Herzberg to be established, independent assessment of the point when the ADS would be expected to identify a collision course, and comparison of this reference data to the *ADS decision data stream* to establish when the ADS actually identified the risk of collision.
- The comparison of event timings of *perception* and *decisions* with the *reaction data stream.* In this specific case this would have identified the ADS perceived the object at -5.6 seconds, only established a collision risk at -1.2 seconds, and did not act until -0.2s due to the ADS using an *action suppression algorithm*. This deliberate 1 second delay is 3x worse than the response time of a drunk human driver.

Outcome explainability, in this context helps reconstruct the accident and could have been available to investigators immediately after the collision significantly reducing the duration required to reach safety critical recommendations for continual learning.

Outcome explainability is ex-post facto, it enables corrective actions to be taken after the fact. It cannot change the fact the collision occurred and is not intended to alter the decision-making of the ADS which remains responsible, at all times, for the dynamic driving task.

It should also be noted that Uber ATG's use of "*action suppression algorithm*" seems to have been originally motivated as a tool to reduce the number of hard braking incidents caused by false positive detections e.g. the detection and reaction to phantom objects that don't really exist. The *perception, decision, reaction and outcome explainability model* proposed above is equally well suited to capturing these "false positive events".

It is suggested that the safety assurance regulator be granted the powers to access perception, decision and reaction data streams published by the ADS for use in independent real-time processing of outcome explainability and for that access to include provision for real-time data access both onboard or remotely from the vehicle.

C.1.1.6 Outcome explainability and Elaine Herzberg's fatality prevention

Outcome explainability is a leading measure with the benefit that they can be acted upon before harm occurs.

Outcome explainability is intended to run continually every time the ADS operates. So, to answer the question it's important to know whether Uber ATG's ADS;

- Had driven that route before? Yes.
- Had driven in night-time conditions before? Yes.
- Had detected objects, events and planned responses before? Yes.
- Had been active before the collision? Yes for approx. 45 minutes.

Leading measures generated by *outcome explainability*, during any of these periods of operation, could have provided evidence of poor stability in perception, poor risk evaluation in decisions and the use of action suppression in safety critical reactions.

In the proposed UK regulatory structure, the safety assurance regulator would have been aware of that evidence and have the power to impose sanctions to redress, comply or suspend or withdraw authorisation. In which case the Uber ATG ADS might not even have been on the road that evening. This is the power in-use monitoring and leading measures have in assuring safety.

If the above suggestions for explainability are implemented, then in-use monitoring, and the generation of leading metrics, runs in real-time. There are significant additional safety benefits of this approach.

Consider a scenario where Uber ATG had released a new ADS software version that evening and the journey with Elaine Herzberg was the first on public roads. If any significant safety issues had been identified during the journey prior to the fatal collision, they could have been acted upon *immediately*.

For example, if there had been a similar near-miss event earlier in the journey what would be the public's expectation of safe behaviour? Stopping the journey until it's clear it's safe to proceed?

Referring back to the Road Traffic Act 1988 definition of dangerous driving would this situations seems to fall under *"any circumstances shown to have been within the knowledge of the accused"*? As such, it could be evidence of dangerous driving for the near-miss event itself, but also if that knowledge was ignored and resulted in a fatal collision, under similar circumstances, that seems to be significant evidence of dangerous driving?

So what role could the safety assurance regulator have played to prevent the fatal collision after the earlier near-miss? Perhaps the assumption in the currently proposed regulatory framework is none, they would only be able to act after the specialist collision investigation unit completes a report?

The significant safety advantage of real-time in-use safety assurance is the regulatory can also use their powers in real-time e.g. *immediately* on detection a safety critical event such as a near-miss.

In practice that would mean that the safety assurance regulator having the powers to immediately suspend authorisation to operate upon a safety critical near-miss event. The suspension might be limited to this one specific journey for this one specific vehicle until further evidence is gathered that may indicate a wider fleet issue.

This immediate suspension of authorisation could be implemented digitally onboard the vehicle, whereby *outcome explainability* process generates not only a trigger to record data but also a notification that can be consumed by the ADS and used to trigger execution of a minimal risk manoeuvre (MRM).

Perhaps the biggest lesson that can be learned from the Uber ATG fatal collision with Elaine Herzberg in Tempe, Arizona is that permitting ADS operations on the road without independent safety assurance can unnecessarily expose the public to risk of harm.

If Arizona had kept entry barriers low, by permitting Uber ATG to operate without testing, then would a real-time in-use safety assurance scheme have been sufficient to assure public safety within a no-blame culture of continual learning? Arguably, yes.

If Arizona had raised entry barriers, by only permitting Uber ATG to operate after detailed safety case review, then would this have created a culture of innovation and that protects intellectual property? Arguably, no. Uber ATG chose to deploy their ADS on the roads in Arizona to avoid these restrictions imposed by other US states.

The question for the proposed GB regulatory framework, which is ultimately a political decision, is which type of independent safety assurance can enhance public safety, promotes innovation and protects intellectual property?

It is suggested that the safety assurance regulator be granted the powers impose sanctions in real-time based upon outcome explainability leading metrics. These sanctions should include the ability to suspend authorisation to operate during a journey where there is

evidence to indicate careless or dangerous driving (such as near-miss events) which should be investigated. Under these powers the ADSE must ensure that the ADS has will respond immediately to automated suspension request by executing an minimal risk manoeuvre. The regulator should have additional power to impose further sanctions for non-compliance of the request to stop once authorisation is suspended.

It is also recommended that the final regulatory framework proposal defines the most appropriate safety assurance scheme that maintains public safety, promotes innovation and protects intellectual property. It is suggested that real-time in-use safety assurance based upon outcome explainability and leading metrics fulfils this requirement and should therefore be considered as foundation upon which additional assurance measures can be introduced.

Automated Vehicle Safety Assurance - In-Use Safety and Security Monitoring

Abstract

This report examines the feasibility of establishing a regulatory scheme for In-Use Safety and Security monitoring. This report assesses the legal, societal and technological requirements of such a scheme to continually monitor the safety of Automated Vehicles (AVs) using both in-vehicle data streams and operational data. A framework has been proposed which outlines a process of event based data capture, data recall, analysis, investigation, and regulatory intervention to maintain the safety of AVs deployed on roads in Great Britain.

Relevant Report

- Automated Vehicle Safety Assurance - In-Use Safety and Security Monitoring Task 1 – Road Incident Taxonomy; https://doi.org/10.58446/mvuc1823
- Automated Vehicle Safety Assurance - In-Use Safety and Security Monitoring Task 2 - Minimum Dataset Specification; https://doi.org/10.58446/nksn4732
- Automated Vehicle Safety Assurance - In-Use Safety and Security Monitoring Task 4 - Post Event Investigation Process; https://doi.org/10.58446/egfa6491
- Automated Vehicle Safety Assurance - In-Use Safety and Security Monitoring Task 5 - Outcome Reporting; https://doi.org/10.58446/qlpq9096
- Automated Vehicle Safety Assurance - In-Use Safety and Security Monitoring Task 6 - Data Privacy; https://doi.org/10.58446/dwll8689
- Automated Vehicle Safety Assurance - In-Use Safety and Security Monitoring Task 7 - Change Control; https://doi.org/10.58446/bpdl3309

TRL
Crowthorne House, Nine Mile Ride,
Wokingham, Berkshire, RG40 3GA,
United Kingdom
T: +44 (0) 1344 773131
F: +44 (0) 1344 770356
E: enquiries@trl.co.uk
W: www.trl.co.uk

ISSN: 2514-9652

DOI: https://doi.org/10.58446/sgxq7004

PPR2018