



# Law Commission & Scottish Law Commission

## Consultation Paper 3 A regulatory framework for automated vehicles

### FG-AI4AD Feedback & Comments

Chairman's Draft for Review

1.	Overview .....	5
2.	Executive Summary.....	6
3.	Insufficient in-use data to create a no-blame safety culture .....	8
3.1.	Discrepancy in near-miss leading metrics between UIC and NUIC vehicles .....	8
3.2.	Detection of near-miss and collision events are missing from the framework.....	8
3.2.1.	How can UIC or Licensed Fleet Operator report if they cannot detect collisions? .....	9
3.2.2.	Should the ADSE report near-miss and collision events part or the OEDR? ....	10
3.2.3.	Near-miss detection, low impact and VRU collision detection – missing regulation, standards, testing procedures or best practice. ....	10
3.2.4.	Extending eCall for VRU collision reporting for self-driving vehicles .....	12
3.3.	ADS activation status is insufficient for creating a no-blame safety culture -even if both time and location are recorded .....	13
3.3.1.	When is a ‘black-box’ not a “black-box”?.....	14
3.3.1.	Was the self-driving vehicle at the scene of a collision? .....	15
3.3.2.	Police incident witness appeals. Self-driving vehicles, can you help? .....	16
4.	Data required for Transparency and Explainability in a no-blame safety culture .....	19
4.1.	Enhanced powers to assure safety post-deployment.....	19
4.2.	Data required for continual learning .....	19
4.2.1.	Determining if a collision is “reasonably foreseeable and preventable” .....	20
4.2.1.	ALKS “reasonably foreseeable and preventable” .....	21
4.3.	The <i>perception, decision, reaction</i> and <i>outcome explainability</i> model.....	23
4.3.1.	Perception explainability .....	25
4.3.2.	Decision explainability .....	27
4.3.3.	Reaction explainability .....	28
4.3.4.	Outcome explainability .....	29
4.3.4.1.	Outcome explainability and Elaine Herzberg’s fatality prevention .....	32
4.4.	Powers to define and collect leading metrics .....	34
4.4.1.	Leading Metric Example – Safe Passing of Vulnerable Road Users .....	34
4.4.2.	Are aggregate statistics on leading metrics sufficient?.....	36
4.4.3.	Leading metrics should be defined in an open, fair and transparent manner	37
4.4.4.	Leading Metric example – safety relevant traffic infractions.....	38
4.5.	Which entity should be responsible for processing of leading metrics? .....	40
5.	Comparisons between human and automated driving safety .....	41

5.1.	Comparison at the level of Object and Event Detection and Response (OEDR) .....	41
5.1.1.	OEDR limitations of ALKS compared to a human driver and implications for self-driving categorisation .....	42
5.1.1.	OEDR limitations of ALKS tailgating scenario .....	44
5.2.	Digital Commentary Driving – the human equivalent of in-use monitoring of OEDR performance.....	50
5.3.	Digital Commentary Driving for human OEDR performance comparison .....	51
5.3.1.	Digital Commentary Driving - assessment of the ADS safety in offline mode .....	51
5.3.2.	Digital Commentary Driving - assessment of the human driver safety using offline ADS.....	52
5.3.3.	Opt-in for assessment of the human driver safety using ADAS.....	52
5.4.	The value of an all-vehicle safety assurance dataset .....	53
6.	Limitations, loopholes and liability gaps .....	56
6.1.	The “Full Self-Driving” Loophole for development and deployment .....	56
6.1.1.	When is “full self-driving” actually “not self-driving”? .....	57
6.1.2.	Drunk, drugged, drowsy & distracted – are self-driving behaviours too .....	58
6.1.3.	AVs for use only with Human Driver Supervision (HDS) .....	59
6.2.	Limitations imposed by the UIC and NUIC definitions and categorisation .....	60
6.2.1.	The OEDR continuum .....	60
6.2.2.	Mapping OEDR capability to the circumstances and situations .....	61
6.2.2.1.	OEDR, MRMs, MRCs, ODD and stopping in lane .....	62
6.2.3.	Path 1 UIC prevents passenger only or unoccupied modes of operation .....	63
6.2.3.1.	CAV benefits must be accessible for all.....	63
6.2.3.2.	The Designated Passenger .....	64
6.2.3.3.	People must remain in control of their transport choices with CAVs .....	65
6.2.3.1.	Why must private Self-Driving Vehicles owners give up their privacy for to access passenger only modes .....	66
6.2.4.	UIC does not permits unoccupied modes such as Valet Parking .....	67
6.2.4.1.	Valet Parking – can a single vehicle be both UIC and NUIC? .....	67
6.2.5.	UIC & NUIC definitions may restrict or distort competitive markets.....	69
6.2.6.	Policing limitation of not-self-driving, UIC and NUIC vehicle categorisations .....	70
6.2.6.1.	Failure to stop for the police – who is the “driver” for UIC or NUIC?.....	71
6.2.6.1.	Sleeping & self-driving, how can police identify permitted behaviours? .....	72

6.2.6.1.	Causing death by careless driving when under the influence of drink – is the ADSE responsible when the UIC is drunk in charge? .....	74
6.2.6.2.	Drink driving, drunk in charge – how can police identify offenses? .....	75
6.3.	Limitations of the Safety Case for of the ADSE Criminal Liability .....	76
6.3.1.	New offences based upon the Safety Case .....	77
6.3.2.	ADSE - aggravated offence for death or serious injury .....	79
6.3.3.	The significance of <i>misrepresentations or non-disclosures</i> related to 'real-life' safety performance .....	80
6.4.	Limitations in scaling investigations for both collisions and near-misses .....	83
6.4.1.	Who investigates collisions and safety relevant infractions? .....	84
6.4.2.	Moving beyond criminal liability to a culture of safety .....	84
6.4.3.	The need to address the scale, rate and duration of investigations .....	85
6.4.4.	The value of 'real-life' evidence to incident investigation efficiency and proportional use of suspension or withdrawal of authorisation .....	86
6.4.5.	"Data-dumps" are not just as safety case issue .....	88
6.4.6.	How fast is safe enough? .....	88
6.4.1.	Police - immediate on the spot access to ADS in-use performance data .....	89
6.5.	International safety assurance .....	91
6.5.1.	Leading metrics and collisions investigations at an international scale .....	91
6.5.2.	Great Britain, Northern Ireland, Ireland & Europe Cooperation .....	93
6.5.3.	1968 Convention on Road Traffic - Article 34 bis & global fragmentation in self-driving regulation .....	93
6.5.3.1.	1968 Convention on Road Traffic – importance for Article 34 bis, ALKS & safety assurance .....	95
6.5.4.	Safety Assurance Regulator – extended responsibility for CAV Ethics .....	101
6.5.5.	Safety Assurance Regulator – extended responsibility for UN Improving Global Road Safety resolution .....	103
6.5.6.	The monitoring test and control test – revisited and revised .....	104
6.5.6.1.	The AD control test to determine use only with Human Driver Supervision (HDS) .....	105
6.5.1.	ADSE regulatory sanctions– Penalty Points (Endorsements) .....	106

## 1. Overview

Commissioned by the UK Government Centre for Connected and Automated Vehicles (CCAV) the Law Commission and Scottish Law Commission have conducted a three-year project conducting three rounds of high quality and in-depth consultation to review the legal framework for automated vehicles, and their use as part of public transport networks and on-demand passenger services. This paper is a response to Consultation Paper 3.

On the 2<sup>nd</sup> March 2021 the International Telecommunications Union (ITU) Focus Group on AI for Autonomous and Assisted Driving (FG-AI4AD) welcomed the opportunity to partner with the Law Commission and Scottish Law Commission on a 3hr international workshop entitled; [\*A Regulatory Framework for Automated Driving: the Value of in-use Data for Creating a no-blame Culture of Safety\*](#). We would like to recognise that, although the focus of the consultation is upon domestic regulation, it adds significant value to the international discussion and debate on automated driving regulation and safety assurance.

On the 17<sup>th</sup> March 2021 the FG-AI4AD Chair attended the second meeting of the All Party Parliamentary Group (AAPG) on Connected and Automated Mobility (CAM) at which the Law Commission and Scottish Law Commission presented the outline proposals for the regulatory framework detailed in Consultation Paper 3. The comments and discussion with Members of Parliament provided a clear insight into the challenges of building public trust in automated driving technologies as well as the need for robust regulatory frameworks.

FG-AI4AD's activities are closely aligned with the CP3 proposals for in-use safety monitoring which we believe is foundation to the creation of a no-blame safety culture based around continual learning. The balance between the pre-deployment safety case and post-deployment in-use monitoring is critical to creating a vibrant, innovative and accessible market for automated mobility whilst also providing safety critical assurance of the automated driving software and mobility services during live operations.

Public trust in CAM is essential to achieving widespread deployment and adoption. The comments and suggestions below aim to further this objective. They represent the view of the FG-AI4AD Chair and as a reflection of previous discussions and debates held over the last 14 months within FG-AI4AD meetings. The presented poll results represent the views of participants attending the 2<sup>nd</sup> March joint workshop.

We hope that the following feedback enhances the output of this detailed consultation and helps to establish a solid and robust regulatory framework which positions the UK as the leader in the future of safe connected and automated mobility.

The suggestions in this paper do not represent the views of ITU or the ITU Member States. This Chairman's Draft for Review will be submitted to the ITU FG-AI4AD for further contribution and comment.

## 2. Executive Summary

The introduction self-driving within no-blame safety culture, based upon continual learning, is a positive forward step for road safety.

The definition of User-In-Charge (UIC) compromises the framework, makes liability overly complex, will create future public confusion, risks restricting and distorting competitive markets and does not meet the conditions of CAV public acceptability in ensuring benefits must be accessible for all and people must remain in control of their transport choices.

This appears to be an unfortunate consequence of attempting to categorise ALKS, and similar SAE Level 3 Conditional Driving Automation systems as “self-driving” (which as this paper suggests, both are clearly in contravention of articles related to driver behaviour defined in the 1968 Convention on Road Traffic).

The public expectation is that a “self-driving” vehicle capable of “safely driving itself” no longer requires a human driver i.e. it introduces new passenger only and unoccupied modes of use.

Passenger only and unoccupied modes should be available to privately owned and fleet operated vehicles.

Creating this parity simplifies the legal framework significantly, makes it easier for the public and legal entities to understand roles and responsibilities, creates more equitable access to mobility choices for “non-drivers” and establishes an open and fair market between private self-driving vehicles and fleet operated mobility services.

The framework correctly introduces the principal of safety assurance during both pre-deployment and post-deployment. The balance of the two will be critical to the success of a no-blame safety culture based upon continual learning. At present the framework has an overreliance on the pre-deployment safety case assessment while post-deployment ‘real-life’ safety assurance is limited in scope.

For example, an ADSE can be held criminally liable for misrepresentations or non-disclosures within the safety case but, unless the framework is tightened, would not be criminally liable for failing to report a fatal collision with a child pedestrian and, in fact, the ADSE is not explicitly defined as the legal entity responsible for detecting the collision and has no obligation to record any data as evidence in an investigation.

There are a number of suggestions made throughout this paper to help redress the balance, by improving post-deployment safety assurance measures including; the suggestion of making the ADSE responsible for detecting and reporting both collisions and near-miss events; the need for eCall regulations to be extended to include self-driving; and the need for the safety assurance regulator to have necessary powers to require the ADSE to publish

real-time data from the ADS under a *perception, decision, reaction and outcome explainability model*.

It should be noted that, while the quality of the safety case may be essential for initial self-driving authorisation, it is the quality of 'real-life' data that becomes essential in the consideration of suspension or withdrawal of authorisation.

A number of limitations, loopholes and liability gaps within the framework, both domestic and international, are discussed at the end of this paper.

The continued need for alignment between the type-approval authority and the UNECE World Forum for Harmonization of Vehicle Regulations (WP.29) is recommended.

It is suggested that the safety assurance regulator seek to align with the UNECE Global Forum for Road Traffic Safety (WP.1) and, in doing so, is explicit about the interpretation of Article 34bis and whether the driver behavioural expectations defined in the articles of the 1968 Convention on Road Traffic remain relevant for self-driving.

Furthermore, the safety assurance regulator appears best placed to address issues of CAV ethics to ensure, not only road safety, but privacy, fairness, explainability and responsibility in automated mobility. Whilst also ensuring that self-driving delivers the targets set by the United Nations resolution on Improving Global Road Safety and related Sustainable Development Goals.

This paper challenges the interpretation of the Automated and Electric Vehicles Act (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"* which has resulted in complex definition of a *"monitoring test"* and associated *"control test"*.

Instead, this paper proposed a much simpler test; if, to *"safely drive itself"*, the ADS must continually monitor for the presence of a licensed driver then it is not *"self-driving"*. However, if the same ADS passes the longitudinal and lateral control test it is clearly not *"driver assistance"* and may be categorised as an AV for use only with Human Driver Supervision (HDS). For public safety, confidence and trust it is suggested that a HDS ADS (SAE Level 2 and 3) should fall within the scope of the safety assurance regulator.

It's clear that the GB regulatory framework can provide global leadership on self-driving safety and the UK uniquely placed to seek international harmonisation on these requirements to ensure frictionless safe self-driving travel between all four UK countries as well as between Northern Ireland, Ireland and by extension Europe.

It is hoped that the comments, feedback and suggestion of this paper help achieve those aims.

### 3. Insufficient in-use data to create a no-blame safety culture

In response to Consultation Question 18 relating to the proposal to create an enhanced AV safety assurance scheme requiring additional statutory responsibilities and powers we received the following poll responses;

*Should the regulator have power to collect information on:*

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

- 83.3% agreed

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

- 92.6% agreed

*It is suggested that the proposed safety assurance scheme should indeed cover both leading measures on “near-miss” events as well as lagging measures on “collision” events.*

These results align closely with the 469 public responses FG-AI4AD received to The Molly Problem Survey in which 87% people agreed that data should be recorded for “near-miss” events in addition to “collision” events<sup>1</sup>.

#### 3.1. Discrepancy in near-miss leading metrics between UIC and NUIC vehicles

In the proposed framework there is a discrepancy between the reporting requirements between User in Charge (UIC) vehicles and No User-in-Charge (NUIC) vehicles<sup>2</sup>.

Whilst accident (collision) reporting is required for both; near-miss reporting is only required by the Licensed Fleet Operator responsible for a NUIC vehicle.

There is a risk is that if this discrepancy is not address it will reduce the amount of near-miss data available within the safety assurance scheme, reduce the opportunity for continual learning these leading metrics can provide and increase the number of collisions that may bias UIC vehicles.

*It is suggested that all self-driving vehicles report near-miss information to support the use of leading measures for safety assurance within the no-blame culture.*

#### 3.2. Detection of near-miss and collision events are missing from the framework

---

<sup>1</sup> [The Molly Problem Survey](https://www.itu.int/en/ITU-T/focusgroups/ai4ad/Pages/MollyProblem.aspx) : <https://www.itu.int/en/ITU-T/focusgroups/ai4ad/Pages/MollyProblem.aspx>

<sup>2</sup> See Summary of Consultation Paper 3: Figure 1.1 - Overview of the roles of the user-in-charge and licensed fleet operator



Without detection there can be no reporting; this applies to both near-miss and more importantly to collision events.

Without reporting a key foundation to the no-blame safety culture is missing.

Which legal entity;

- is responsible detecting collision and near-miss events?
- has access to the data required to detect these events?
- will need to implement the learnings from these events to ensure continual improvement?

#### 3.2.1. How can UIC or Licensed Fleet Operator report if they cannot detect collisions?

Currently the framework defines the UIC or Licensed Fleet Operator as responsibility for reporting events.

How would these entities know that a near-miss or collision has occurred?

In the case of a UIC vehicle, when the ADS is engaged, the driver becomes a “user-in-charge” and so by definition they are no-longer responsible for conducting the Object Event Detection and Response (OEDR).

However, in the current framework if the UIC fails to report a collision they can still face criminal charges, conviction and end up in jail. I would expect significant public push back on this remaining criminal liability.

For the Fleet Operator, who is remote from the UIC vehicle, it's even less clear how they would be made aware of a collision. There is a requirement to “operate remote supervision” but this does not extend to completing the OEDR which is a dynamic driving task executed by the ADS.

It is conceivable that the ADSE could design the ADS in such a way that it reports collision and near miss events to the UIC or Licensed Fleet Operator for them to report. However, within the current legal framework there is no legal responsibility for the ADSE to share this data which creates a liability gap.

What purpose does it serve to have the UIC or Licensed Fleet Operator be made responsible for reporting? Is the assumption that human verbal communication is required for the task? Is the assumption that the ADSE does not have data communications from the vehicle? Is the assumption that systems such as automated eCall could not be extended for this function within a self-driving vehicles?

*It is suggested that further clarity be provided as to why the proposed legal framework considers the UIC and Licensed Fleet Operator as the most appropriate legal entity for reporting of near-miss and collision or that these report responsibilities be move on to the ADSE.*

### 3.2.2. Should the ADSE report near-miss and collision events part or the OEDR?

Is it the Law Commissions position that the OEDR includes the near-miss and collision detection? If so, then detection is a function of the ADS and should be the legal responsibility of the ADSE.

The issue of collision detection was also explored in The Molly Problem Survey in which 97% of respondents believed that the self-driving software should be responsible for collision detection.

Additionally, over 94% of respondents believe that self-driving software should be responsible for post-collision response which includes stopping the vehicle, indicating a hazard to other road users and alerting the emergency services.

These results suggest is that the ADSE that should be responsible for the collision detection and that the ADSE should be responsible for collision reporting. As indicated earlier the ADSE should also be expected to detect and report near-miss events.

It's essential that the proposed framework clearly identifies the responsible entity for the detection of near-miss and collision events.

*It is suggested that framework explicitly defines the legal entity responsible for near-miss and collision detection and it is suggested that the ADSE, who is responsible for the ADS execution of the OEDR within the dynamic driving task, is the most appropriate entity.*

Under the currently proposed framework this would remove liability gap whereby an ADSE legally withhold near-miss and collision data from an UIC or Licenced Fleet Operator while those entities that remain legally liable for reporting.

*Should, under a revised framework, the ADSE is held legally responsible for collision and near-miss detection, then it is suggested that the ADSE also be held responsible for reporting of these events.*

### 3.2.3. Near-miss detection, low impact and VRU collision detection – missing regulation, standards, testing procedures or best practice.

As described above near-miss and collision detection are key foundations to the no-blame safety culture.

Within that safety culture, The Molly Problem results, indicate a zero tolerance for hit and run collisions with child pedestrians<sup>3</sup>.

There are currently no regulations, no standards, no testing procedures or even industry best practice for assisted or automated driving system *detection* of low-impact collisions, vulnerable road user collision, runover events or near-miss events<sup>4</sup>.

Detecting pedestrian collisions, low impact collisions and near-miss events remains the responsibility of the driver.

Driver assistance technologies such as Autonomous Emergency Braking (AEB) focus upon collision avoidance or harm mitigation, but they still rely upon the human driver to detect and report all collision events.

Pedestrian airbag systems<sup>5</sup> are typically designed to detect frontal only collisions with an object that has similar properties to those of the human leg. The trigger is based upon the force of the impact measured using sensors in the front bumper. However, these systems may have a restricted ODD, such as Volvo's current limitation to speeds of approximately 20-50 km/h (12-30 mph)<sup>6</sup>. These fitting of the systems is not mandatory.

European Union requirements for Event Data Recorder (EDR) and eCall are both triggered on high-impact collision events such as those sufficient to activate non-reversible restraints.

It seems clear VRU collision detection remains a critical safety gap for self-driving vehicles.

***Given the zero tolerance for "hit and run" collisions expressed by respondents to The Molly Problem public survey, it is suggested that the Law Commission address this safety critical gap of detection of low-impact collisions, vulnerable road user collisions, runover events and near-miss events within the final regulatory framework proposal<sup>7</sup>.***

The consultation paper asks whether "*for a vehicle to be classified as self-driving, it needs to record the location as well as the time at which the ADS is activated and deactivated<sup>8</sup>*". Which suggests that the criteria used to determine whether a vehicle is classified as self-driving is yet to be finalised.

---

<sup>3</sup> Follow up research will confirm whether this applies to other VRU age categories.

<sup>4</sup> EuroNCAP introduced VRU protection standards in 1997 but does not require detection of collision only their mitigation which may be achieved through passive or active means; <https://www.euroncap.com/en/vehicle-safety/the-ratings-explained/vulnerable-road-user-vru-protection/>

<sup>5</sup> First introduced by Volvo in 2012 <https://www-esv.nhtsa.dot.gov/Proceedings/23/files/23ESV-000447.PDF>

<sup>6</sup> <https://www.volvocars.com/en-th/support/manuals/v40-cross-country/2018w17/safety/airbags/pedestrian-airbag>

<sup>7</sup> In accordance with "the call for evidence: problematic scenarios" para 3.28 note (2) Stopping after an incident (Page 24 of the Full Consultation Paper 3)

<sup>8</sup> Consultation Paper 3, page 298, para 17.65

Pedestrian fatalities represented 27% of all fatalities on British roads in 2019<sup>9</sup>. They are among the vulnerable road user group that are already disproportionately at risk of collision, serious injury and death.

*It is suggested that the final regulatory framework specifically addresses the most important safety assurance question currently missing from the public consultation, that is whether; “for a vehicle to be classified as self-driving, it needs be able to detect and report vulnerable road user collisions?”.*

Ultimately, as pointed out in the consultation<sup>10</sup>;

*Section 1 of the Automated and Electric Vehicles Act 2018 places the decision on whether a vehicle is capable of safely driving itself on the Secretary of State. It also affords the Secretary of State some discretion (“in the Secretary of State’s opinion”).*

*The Secretary of State’s decision can (and should) be informed by technical experts.*

*...tests can inform political decisions, such as whether, for example, benefits to many road users outweigh the dangers to a few road users in some rare scenarios.*

*However, there will always be some risk. Whether this remaining risk is acceptable is a political decision.*

#### 3.2.4. Extending eCall for VRU collision reporting for self-driving vehicles

The Electronic Communications and Wireless Telegraphy (Amendment etc.) (EU Exit) Regulations 2019<sup>11</sup> grants *power to alter standards relating to eCall service*;

*The Secretary of State may by regulations made by statutory instrument make provision to replace the standards that are for the time being referred to in Articles 2(j) and (k), 3(1) and (4) and 4 of the eCall Regulation.*

The eCall Regulations 2(j) specifies the definition of;<sup>12</sup>

*‘minimum set of data’ (MSD) means the information defined by the standard ‘Road transport and traffic telematics — eSafety — eCall minimum set of data (MSD)’ (EN 15722) which is sent to the eCall PSAP;*

---

<sup>9</sup>[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/922717/reported-road-casualties-annual-report-2019.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/922717/reported-road-casualties-annual-report-2019.pdf)

<sup>10</sup> Consultation Paper 3, page 84, para 5.114 & 5.115

<sup>11</sup> <https://www.legislation.gov.uk/uksi/2019/246/regulation/5/made>

<sup>12</sup> [Commission Delegated Regulation \(EU\) No 305/2013 of 26 November 2012](#)

These powers to alter standards related to eCall service could be extremely beneficial, within the regulatory framework for automated driving, for creating a solid foundation for the no-blame safety culture.

As discussed above the detection *and* reporting of pedestrian collisions were both seen by the public as critical safety capabilities for self-driving software; 97% and 94%.

*It is suggested that there are a number of ways the existing eCall system, could be extended for self-driving vehicles;*

- 1. Automated eCall activation for vulnerable road user collisions*
- 2. Manual internal activation for passengers*
- 3. Manual external activation for pedestrians (similar protected design as existing bus emergency stop buttons)*
- 4. Manual eCall extended to include activation by remote operators*
- 5. Extension of eCall voice service to include external road users*
- 6. Extension of eCall voice service to include remote operators*
- 7. Additions to the minimum set of data (MSD) which specifically address new requirement introduced by self-driving vehicles such as; activation state of the ADS and data regarding the road traffic situation of the collision*

Research indicates that eCall systems, can save lives or decrease the seriousness of injuries by relaying crucial information to emergency responders and thus reducing their response time significantly by 50% in rural and 40% in urban areas<sup>13</sup>.

*It is suggested that, in preparing the final regulatory proposal, the Law Commissions review and comment on the potential to alter existing eCall regulation to enhance the levels of safety assurance of self-driving vehicles.*

### 3.3. ADS activation status is insufficient for creating a no-blame safety culture - even if both time and location are recorded

The framework proposal to extend the function of the existing DSSAD to record both time *and* location data for the ADS activation and deactivation are well placed and reasonable.

For Consultation Question 55.1 the workshop participants poll indicated 81.8% believed that;

*“for a vehicle to be classified as self-driving, it needs to record the location as well as the time at which the ADS is activated and deactivated<sup>14</sup>”.*

The Molly Problem Survey asked a related question;

---

<sup>13</sup> <https://heero-pilot.eu/view/en/ecall.html>

<sup>14</sup> Consultation Paper 3, page 298, para 17.65

Should self-driving software, which is unable to recall data about a collision events, be prohibited from driving? In the survey responses 75% yes, 14% were unsure and 11% said no.

For The Molly Problem the public agreed that eleven different sources of data about the collision event were relevant. This included the time and location of the collision event which both received 99% approval.

Prohibition of driving, in the case of The Molly Problem, implies both that;

- a vehicle should not be licenced to drive without this recall capability;
- an approved self-driving vehicle which suffers a data ingest or recording failure should immediately stop and not return to driving until the issue is resolved.

Responses to The Molly Problem indicate that 88% of the public believe the same data recall capabilities are should also be required for near-miss events not just collisions.

*It is suggested that for a vehicle to be classified as self-driving it should meet the recording and recall capabilities necessary for in-use safety assurance and not just activation status.*

#### 3.3.1. When is a 'black-box' not a "black-box"?

*It is suggested that, given the current specification, the UNECE DSSAD, would be insufficient to meet the public expectation of "black box" as it has been described<sup>15</sup>.*

The term "black-box" is commonly associated with the functions of an aircraft's Flight Data Recorder (FDR) and Cockpit Voice Recorder (CVR). Collectively these two recorders provide aviation investigators the necessary data required to reconstruct the specific circumstances and situations leading up to the collision event; the FDR provides data on the environment and aircraft status, whilst the CVR serves as a proxy for the awareness, decision-making and actions taken by the pilots in response to the safety critical situation.

The Molly Problem responses highlight the public's expectation that a self-driving black-box should also; record environment and vehicle status as well as the "awareness, decision-making and actions taken by the software in response to the safety critical situation".

The consultation paper reports the views expressed by the Air Accidents Investigation Branch (AAIB) in relation to safety culture<sup>16</sup>;

---

<sup>15</sup> <https://unece.org/transport/press/un-regulation-automated-lane-keeping-systems-milestone-safe-introduction-automated>

<sup>16</sup> Consultation Paper 3, page 186, para 11.56

*People [...] have learned that, if they actually report these things, when they come to our attention, they are dealt with in a very much no-blame environment. We go to great lengths to ensure that our reports and our investigations do not carry any blame or liability.*

The “black-box” provides the foundation for this no-blame safety culture within the aviation industry. It enables continual learning that both regulators and manufacturers to improve safety.

As a crude comparison, the current specification of the DSSAD, would be the aviation equivalent of “black box” that only records the “on/off” status the aircraft autopilot system. Would this be sufficient for an AAIB investigation? It seems obvious that it would not. So, can the UNECE “black-box” really be described as a “black-box”?

The proposed regulatory framework includes provision to enhance the specification of the DSSAD, to records both time *and* location information ADS activation state change. While this additional location data may be valuable it does not change the primary function of the DSSAD which could be described as a “black box” for “*apportioning blame*” and “*determining where liability should lie*” rather than “*learn lessons from the incident so that they can make recommendations and increase the overall safety*”.

*It is suggested that, in order for the regulatory framework to firmly establish a no-blame safety culture it will need define the specification for a self-driving “black-box” that meets the public’s expectation, the needs of specialist collision investigation unit and requirements of the safety assurance regulator for facilitating continual learning from safety relevant incidents expected of the manufacturers, developers and operators of automated vehicle.*

#### 3.3.1. Was the self-driving vehicle at the scene of a collision?

As will be explored further below even the proposed enhancements to the DSSAD will not always provide proof that the self-driving vehicle was at the location a collision at a given moment of time.

The consultation raises a particular example of a “police investigation of a motorcycle hit and run”<sup>17</sup>. Evidence would be required to place the vehicle at the scene of the collision.

The scenario, as described, makes the assumption that external evidence places the vehicle at the scene and that all that remains to be resolved is whether the ADS was engaged at the time, and location, of the collision in order to apportion blame.

---

<sup>17</sup> Summary of Consultation Paper 3 - Page 53 para 10.10



Location is considered as critical for a situation in which the driver activates the ADS immediately after the collision in order to avoid liability. This location activation data would be used to indicate if the ADS was activated before or after the collision site.

However, apportioning blame is only of value if evidence can prove that self-driving vehicle was present at the scene of the collision. The current UNECE DSSAD specification will not provide this evidence. The addition of location data to the DSSAD activation status does not provide this information.

In fact, if the driver was in control of the vehicle at the time of the collision their best option to avoid liability for the hit-and-run would be to avoid activating the ADS because this will provide incriminating time and location data placing them at the scene.

If the ADS was in control situation is somewhat worse because there certainly would not be any definitive time and location data to place the vehicle at the scene. For a self-driving highway pilot system, limited to the motorway ODD in which the motorcycle collision occurred, then the time and location data may only indicate the junctions at which the vehicle entered and exited the motorway. Whilst for an ADS with an extended ODD the time and location data may not provide any evidence the vehicle used to motorway at all.

*In line with The Molly Problem responses, it is suggested that the regulatory framework ensures that ADS time and location data be recorded in a manner that it can be used as evidence to reliably place the vehicle at the scene of collisions and near-miss events.*

### 3.3.2. Police incident witness appeals. Self-driving vehicles, can you help?

Police witness appeals for serious and fatal road traffic collisions can be a valuable source of evidence. These appeals are often accompanied by a "Can you help?" yellow road sign at the site of the collision specifying the date and time of the incident.

Will self-driving vehicles be able to help with these police witness appeals? If ADS recorded evidence that may be useful for the investigation should the ADSE make this data accessible? If the self-driving vehicle was present at the location during the specified time window should the ADSE be compelled to provide evidence to prove they were not involved in the collision?

Consider again the consultation's example of the vehicle brush collision with a motorcycle on a motorway, resulting in a hit-and-run<sup>18</sup>;

- Should the police have the powers to issue a witness appeal to all self-driving vehicles operating in the area during a given time period?
  - Or should these powers be granted to the safety assurance regulator?
- Which entity should be responsible for responding to the appeal?

---

<sup>18</sup> Summary of Consultation Paper 3 - Page 53 para 10.10



- Is the ADSE best placed to respond to the appeal request; especially if they are also made responsible for reporting events ([see above](#))?
- Should witness appeal response be a legal requirement for the ADSE?
  - Should failure to respond to an appeal have similar consequences to non-reporting of collision and near miss events?
- Should participation in witness appeals form a key foundation to the no-blame safety culture?
- Would establishing a self-driving witness appeal process provide an additional public safety-net for ADS non-detection and non-reporting of collisions?
- Could the self-driving witness appeal process also provide a mechanism for evidence gathering in the case of near-miss events reported by other road users?
- If the self-driving vehicle is shown to be present at the location and time window of the collision would this be sufficient grounds for the police to request additional detailed data, recorded by the ADS, as evidence for use in the collision investigation.
  - This evidence should provide insight into the road traffic situations, the actions taken by the ADS in the environment and risk-based justifications of decision-making by the ADS.

The consultation also cites an additional example of an insurance claim against a self-driving vehicle that is alleged to stop for no reason resulting in the following vehicle also braking and then being hit from behind<sup>19</sup>.

The concerns raised for this specific scenario was;

*...that without location data to show whether the vehicle was at the scene it might be difficult to distinguish between genuine and fraudulent claims.*

And the resulting conclusion was;

*In our view, before a vehicle can be classified as self-driving it will need a data storage system that records whether the ADS was engaged with both a time stamp and location.*

As previously mentioned, the recording the time and location data for changes in ADS activation status might be valuable but it not sufficient to provide evidence that *"the vehicle was at the scene"*.

However, this is another good example of the value of including a self-driving witness appeal process as a key part of the no-blame safety culture.

This also raises an open question as to whether human drivers would be able to opt-in to the no-blame safety culture and witness appeal process. In this case it could be envisaged that even when the ASD is not engaged, sufficient data could still be gathered by the ADS

---

<sup>19</sup> Summary of Consultation Paper 3 - Example 2 : Insurance Claim page 52 para 10.13 to 10.16

regarding the surrounding road traffic situation which could act as evidence for appeals and increase the amount of data available for continual learning. A related topic is explored in a [later section](#).

*It is suggested that the final regulatory framework addresses the legal implications of creating a self-driving witness appeal system; describing the benefits to the no-blame safety culture and evidence gathering for safety critical incidents; the roles responsibilities of the legal actors, safety assurance agency and police; and any related data privacy issues.*

## 4. Data required for Transparency and Explainability in a no-blame safety culture

The framework's proposal for the creation of a no-blame safety culture based upon continual learning has many potential benefits.

Continual learning requires the capture of data that explains events including collisions, near-misses and safety-relevant infractions whilst being transparent about the actions and decisions taken during the event.

In the case of The Molly Problem, in which an unoccupied self-driving vehicle collides with a child pedestrian without any eyewitnesses present, it becomes clear how valuable data from the ADS to provide evidence and reasoning for the self-driving vehicle's behaviour at the scene – which may or may not have contributed to the collision

By making this data available the ADS and the ADSE can act as a witness to the collision. For example; in the case of the self-driving vehicle collision with the motorcycle above, this data would help answer;

- whether the collision “reasonably foreseeable and preventable”?
- whether the actions of the rider were a contributing or casual factor in the collision?
- whether physical contact (the glancing blow) with the motorcycle actually occurred?
- whether the ADS actions caused a near-miss event resulting in rider taking evasive action and loss of control causing them to be thrown from the motorcycle?

### 4.1. Enhanced powers to assure safety post-deployment

In the workshop poll related to Consultation Question 17<sup>20</sup> a total of 88% of respondents agreed with the statement that;

*“legislation should establish a scheme to assure the safety of automated driving systems following deployment, giving scheme regulators enhanced responsibilities and powers”.*

***It is suggested that these powers should include the authority to mandate an ADSE to make sufficient data available for the purposes of continual learning within the no-blame safety culture which should include the provision for accessing real-time data both onboard or remotely from the vehicle.***

### 4.2. Data required for continual learning

---

<sup>20</sup> Consultation Paper 3 page 169 para 10.82

The Molly Problem survey highlighted the transparency and explainability expected by the public should an unoccupied self-driving vehicle be involved in a collision with a child. The survey identified 11 different types of information the public expect self-driving software should be able to recall, whereby recall capability implies that the source data is available, it can be stored and it can be retrieved (see Table 1).

These public expectations provide evidence on;

- the road traffic situation,
- the actions taken by the ADS in the environment
- the risk-based justifications of decision-making by the ADS.

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 1 – The Molly Problem Survey (data sample from 16/02/2021)

#### 4.2.1. Determining if a collision is “reasonably foreseeable and preventable”

The Molly Problem data above, provides key evidence to establish if the collision was “reasonably foreseeable and preventable” in accordance with the UNECE WP.29 safety vision for autonomous vehicles<sup>21</sup>;

*The level of safety to be ensured by automated/autonomous vehicles implies that “an automated/autonomous vehicle **shall not cause any non-tolerable risk**”, meaning that automated/autonomous vehicle systems, under their automated mode ([ODD/OD]), **shall not cause any traffic accidents resulting in injury or death that are reasonably foreseeable and preventable**.*

In the case of The Molly Problem, what evidence is there that the collision was, or wasn't, “reasonably foreseeable and preventable”?

For example, if Molly was static in the middle of the road 5 seconds before impact was the collision reasonably foreseeable and preventable?

<sup>21</sup> <https://unece.org/DAM/trans/doc/2019/wp29/ECE-TRANS-WP29-2019-34-rev.1e.pdf>

Is evidence about the specific incident essential in making this judgement? If so, it is concerning, that under the proposed regulatory framework, no evidence would be available from the active ADS.

Without incident specific evidence how is it possible demonstrate that the ADS behaved in “*reasonably foreseeable and preventable*” and that it “*operated as represented*” by the safety case?

Are the claims, arguments and evidence presented in the safety case sufficient?

For example, if the ADS passed all required pedestrian detection and avoidance tests prior to deployment is that sufficient evidence that this specific collision not “*reasonably foreseeable and preventable*” based purely on the fact for the collision to have occurred it at all it must have been an unavoidable and not foreseeable circumstance or situation.

This type of argument is often associated with scenario-based testing used for verification and validation (V&V) as part of the safety case evidence required for deployment approval. The position being that if “enough” pedestrian scenario tests have been passed the ADS is safe to deploy.

Was the ‘real-life’ collision scenario the same, or similar, to those tested? It’s impossible to know without ‘real-life’ evidence from the specific collision event. If there were safety critical limitations in the V&V process, or ADS software, it’s also impossible to identify these without the same ‘real-life’ evidence and therefore impossible to make recommendation for improvements required for the continual learning culture.

*It is suggested that the regulatory framework balances the requirement to prove the ADS avoids all “reasonably foreseeable and preventable” collisions with a requirement for the ADSE to provide transparent and explainable evidence sufficient for the investigation to prove how the ADS performed in the specific incident.*

#### 4.2.1. ALKS “reasonably foreseeable and preventable”

This “reasonable foreseeable and preventable” language has been carried forward into the UNECE Automatic Lane Keeping System (ALKS) regulation<sup>22</sup>;

*The activated system shall not cause any collisions that are **reasonably foreseeable and preventable**. If a collision can be safely avoided without causing another one, it shall be avoided. When the vehicle is involved in a detectable collision, the vehicle shall be brought to a standstill.*

---

<sup>22</sup> <https://undocs.org/ECE/TRANS/WP.29/2020/81>

The ALKS regulations defines;

- tests with the purpose to verify the technical requirements on ALKS
- conditions under which Automated Lane Keeping Systems (ALKS) shall avoid a collision
- traffic disturbance critical scenarios {which} are those which have conditions under which ALKS may not be able to avoid a collision (Cut-in, cut-out and deceleration)
- traffic critical scenarios of ALKS {which} are divided into preventable and unpreventable scenarios

Do these tests and scenarios define the legal limits of “reasonable foreseeable and preventable” collisions? Does that mean that if these tests are passed at the approval stage are all subsequent ‘real-life’ collisions deemed to be beyond these approved limits – even without evidence of the specific collision conditions?

As the consultation points out<sup>23</sup>,

*Under section 1 of the AEV Act 2018, the Secretary of State must be of the opinion that a vehicle can safely drive itself, at least in some circumstances. If the vehicle is not safe, it should not be listed as self-driving.*

...and provisionally proposes;

*propose a national process for deciding whether a vehicle is safe enough to be allowed on the road without human monitoring. This means that (for example) some ALKS may meet the tests for self-driving while others may not. It also involves identifying an entity with a presence in Great Britain (the Automated Driving System Entity, or ADSE), to take responsibility for the way the vehicle drives on an ongoing basis.*

In relation to the use of ALKS on the road network in Great Britain, one specific self-driving safety issue relates back to the definition of the operational design domain (ODD) which the regulations states<sup>24</sup>;

*ALKS can be activated under certain conditions on roads where pedestrians and cyclists are prohibited and which, by design, are equipped with a physical separation that divides the traffic moving in opposite directions and prevent traffic from cutting across the path of the vehicle. In a first step, the original text of this Regulation limits the operational speed to 60 km/h maximum and passenger cars (M1 vehicles).*

---

<sup>23</sup> Consultation Paper 3 page 32 para 3.69 and page 33 para 3.73

<sup>24</sup> <https://undocs.org/ECE/TRANS/WP.29/2020/81>

This definition has been interpreted as motorways within Great Britain, however, the motorway network includes roundabouts<sup>25</sup> which do not appear in the ALKS regulation and were not mentioned in the ALKS call for evidence<sup>26</sup>.

The regulation defines the following condition: which must be valid for activation of the ADS, and, for which the ADS must deactivate if it becomes invalid;

*The vehicle is on roads where pedestrians and cyclists are prohibited and which, by design, are equipped with a physical separation that divides the traffic moving in opposite directions.*

Under this activation on a motorway exit slip road in traffic queuing for a roundabout would be valid and so to would navigation around the motorway roundabout.

*It is suggested that the regulatory framework clarifies whether motorway roundabouts are excluded from the operational design domain for which ALKS systems would be approved.*

#### 4.3. 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<sup>27</sup>)
- Perception, decision, reaction (as defined in the UNECE ALKS regulation<sup>28</sup>)

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*

---

<sup>25</sup> Such as M25 J4 & M23 J9

<sup>26</sup> [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/921409/Safe-Use-of-Automated-Lane-Keeping-System-ALKS-Call-for-Evidence-FINAL-accessible.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/921409/Safe-Use-of-Automated-Lane-Keeping-System-ALKS-Call-for-Evidence-FINAL-accessible.pdf)

<sup>27</sup> [https://www.zf.com/mobile/en/technologies/see\\_think\\_act/see\\_think\\_act.html](https://www.zf.com/mobile/en/technologies/see_think_act/see_think_act.html)

<sup>28</sup> Appendix 3 page 41 Performance model of ALKS <https://undocs.org/ECE/TRANS/WP.29/2020/81>

segments: "Perception"; "Decision"; and, "Reaction". The following diagram is a visual representation of these segments:

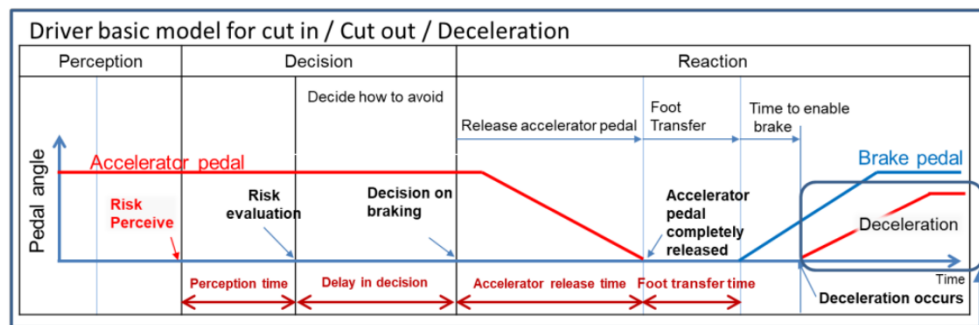


Figure 1 : UNECE skilled human performance model

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 (DDT).

Comparing the performance of self-driving vehicles with human drivers was raised in the consultation paper<sup>29</sup>;

<sup>29</sup> Consultation Paper 3 page 83, para 5.103



*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.*

It is also covered in a later section of this feedback<sup>30</sup>; 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.

#### 4.3.1. 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.

---

<sup>30</sup> [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<sup>31</sup>;

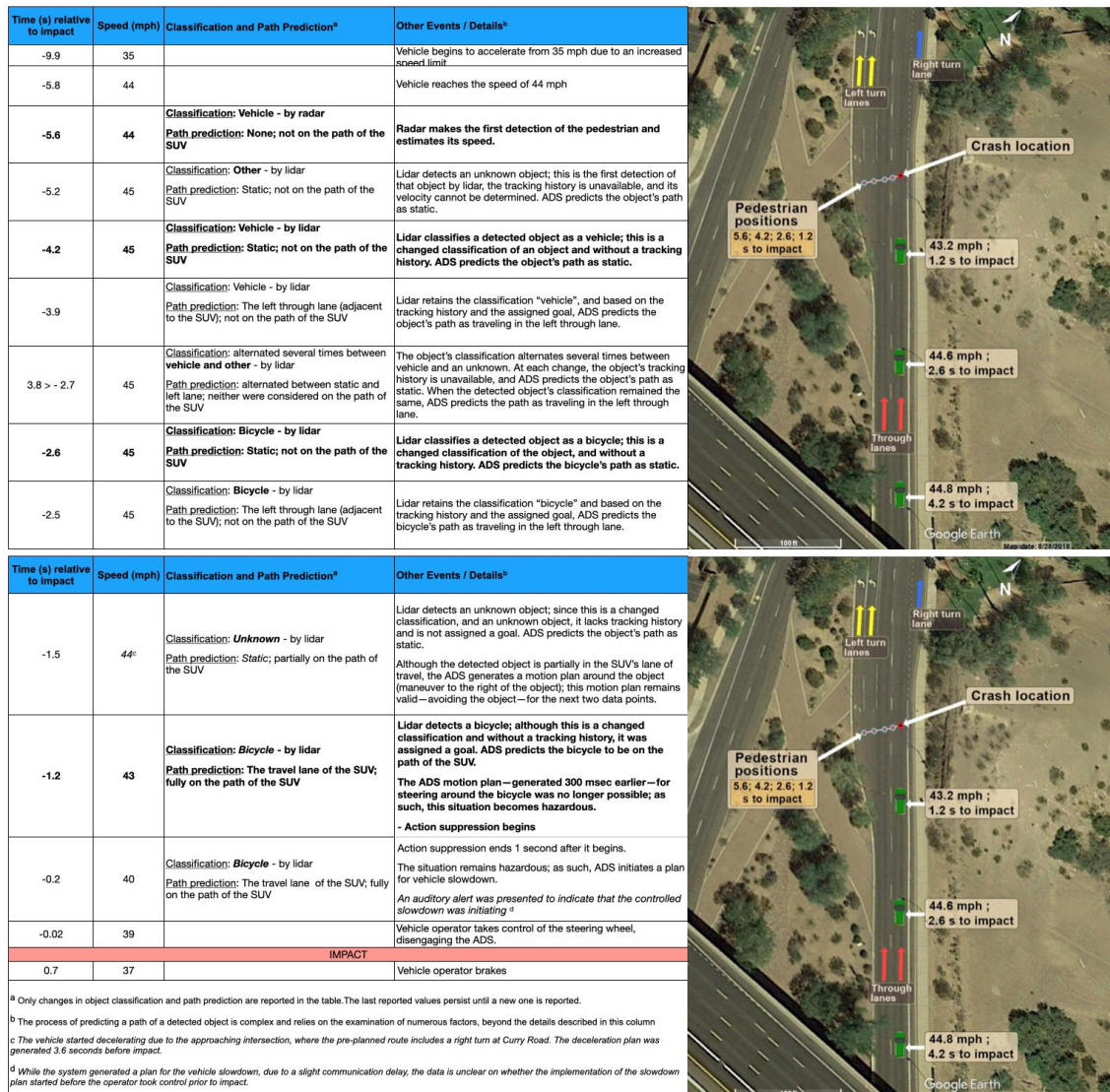


Figure 1 – NTSB HWY18MH010, Tempe, Arizona – Uber ATG

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

<sup>31</sup> [NTSB HWY18MH010, Tempe, Arizona - Uber ATG](#)

'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.*

#### 4.3.2. 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 be 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 its 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 occurs e.g. how close did ADS predict it would come to a pedestrian and how close 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.*

#### 4.3.3. Reaction explainability

Reaction explainability provides insight into the ADS *response* and its *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 its 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.*

#### 4.3.4. 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 SafetyForce 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.*

#### 4.3.4.1. 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.*

#### 4.4. Powers to define and collect leading metrics

Leading metrics measure outcomes. The safety assurance regulator must have the powers to collect these leading metrics and 83.3% of workshop poll respondent agreed.

As discussed in detail above, outcome explainability based upon the ADS perception, decision, reaction data streams are the foundation to in-use safety assurance monitoring and essential for generating the leading metrics and fundamental to the success of the safety assurance scheme.

*It is suggested that the safety assurance regulator is given the powers to define;*

- *the leading metrics which are most appropriate for safety assurance*
- *the thresholds at which these metrics must be recorded, reported and automated*
- *the source data an ADS must publish to be monitored by these metrics*

##### 4.4.1. Leading Metric Example – Safe Passing of Vulnerable Road Users

As an example, proposed government changes to The Highway Code include the following in relation to passing distances;

*give motorcyclists, cyclists, horse riders and horse drawn vehicles at least as much room as you would when overtaking a car (see Rules 211 to 215). As a guide:*

- *leave a minimum distance of 1.5 metres at speeds under 30 mph*
- *leave a minimum distance of 2.0 metres at speeds over 30 mph*
- *for a large vehicle, leave a minimum distance of 2.0 metres in all conditions*
- *pass horse riders and horse-drawn vehicles at speeds under 15 mph and allow at least 2.0 metres space*
- *allow at least 2.0 metres space where a pedestrian is walking in the road (e.g. where there is no pavement) and you should pass them at low speed*

- *you should wait behind the motorcyclist, cyclist, horse rider, horse drawn vehicle or pedestrian and not overtake if it is unsafe or not possible to meet these clearances*
- *take extra care and give more space when overtaking motorcyclists, cyclists, horse riders, horse drawn vehicles and pedestrians in bad weather (including high winds) and at night.*

How will the safety assurance regulator ensure that an ADS always meets these expectations? What leading metrics are required, what thresholds and what data;

- Leading metric = proximity.
  - How close does the vehicle come to a motorcyclists, cyclists, horse riders, horse drawn vehicle and pedestrians (VRUs)?
- Thresholds at which a near-miss event is triggered;
  - < 1.5m under 30mph;
  - < 2.0m over 30mph;
  - < 2.0m for large vehicles at all speeds;
  - < 2.0m and < 15mph for passing horse riders and horse-drawn
  - < 2.0 metres at “low speed” for pedestrian is walking in the road
- Source data required;
  - “Perception data stream” = 3D local world model captured by the ADS
    - must contain the proximity of objects to the ego vehicle
    - must be able to measure proximity accurately below 2m
    - must contain object identification and classification for motorcyclists, cyclists, horse riders, horse drawn vehicle and pedestrians
    - must include segmentation of pavement from road
    - must include location of pedestrian as on pavement or road
    - must be able to track the motion path of the classified objects to ensure the ADS’s *“waits behind the motorcyclist, cyclist, horse rider, horse drawn vehicle or pedestrian and not overtake if it is unsafe or not possible to meet these clearances”*
    - must ensure track the motion path of the classified objects to ensure the ADS does *“take extra care and give more space when overtaking motorcyclists, cyclists, horse riders, horse drawn vehicles and pedestrians in bad weather (including high winds) and at night”*
    - must contain environmental conditions including bad weather (including high winds) and day/night
    - must be able to measure proximity accurately beyond 2m to ensure more space is given under specified environmental conditions.
    - must contain object identification, classification, localisation and tracking for other vehicles (cars, vans, buses, trucks etc) in order to determine “safe/unsafe” passing e.g. maintaining 1.5-2m proximity to a VRU by crossing into the path of an oncoming vehicle which needs to take evasive action is “unsafe”

- “Reaction data stream” = vehicle dynamics
  - must include vehicle speed
- “Decision data stream” = risk evaluation by the ADS
  - ADS risk must be able represent awareness of “safe/unsafe” passing conditions e.g. maintain a “safe” level of risk if required to “wait behind” the VRU and accurately predict the level of risk expected for the full overtaking manoeuvre before it is initiated.
  - must be able to represent risk as a continual value that can be tracked over time
- *Outcome explainability* (independent assessment of exposure to harm/endangerment)
  - Outcome risk assessed independently
    - e.g. passed pedestrian on road with proximity of 0.5m at 40mph lead to a higher exposure to harm than a proximity of 0.5m at 5mph
  - must be able to compare against the ADS predicted risk
  - Outcome endangerment can be considered a measure of “severity”

So, in this example, for the safety assurance regulator to publish leading metrics on VRU near-miss events it must also have the powers;

1. to mandate the ADSE to publish the required data streams from the ADS
2. to define the metrics (algorithms) to process the data
3. to define triggering thresholds for those metrics

#### 4.4.2. Are aggregate statistics on leading metrics sufficient?

There is an argument to say that the safety regulator only needs to receive aggregate statistics on leading metrics and to do so only on a periodic basis.

The California DMV requirement to annually report ADS disengagements is a good example<sup>32</sup>. The report must include;

*(3B) The circumstances or testing conditions at the time of the disengagement including:*

- (i) The location: interstate, freeway, highway, rural road, street, or parking facility.*
- (ii) Whether the vehicle was operating with or without a driver at the time of the disengagement.*
- (iii) A description of the facts causing the disengagements, including: weather conditions, road surface or traffic conditions, construction, emergencies, accidents or collisions. The description should be written in plain language with enough detail that a non-technical person can understand the circumstances triggering the disengagement.*

<sup>32</sup> [CALIFORNIA AUTONOMOUS VEHICLE REGULATIONS](#)

- (iv) *The party that initiated the disengagement (autonomous technology, autonomous vehicle test driver, remote operator, or passenger).*
- (4) *...the total number of miles each autonomous vehicle tested in autonomous mode on public roads each month.*

The counter argument is that a culture of “continual learning” within data driven software development cannot be based around “annual reviews” of “written reports”.

The safety assurance regulator should have the authority operate in real-time on streaming data. In this context, the adaption of (iii) above could read;

**Data** description of the facts causing the **safety critical event**, including: weather conditions, road surface or traffic conditions, construction, emergencies, accidents or collisions. The **data** should be **published in real-time** in a **standardised format** with enough detail that a non-technical person can understand the circumstances of the **safety critical event (when using suitable data visualisation tools)** and an **expert can use the data within event reconstruction, simulation, testing and development tools.**

*It is suggested that the proposed regulatory framework moves beyond the idea that periodic reporting of aggregate leading metrics in a written form is sufficient to create a culture of continual learning for automated driving software development. It is suggested that a “digital first” model of safety assurance based around real-time streaming data is required to establish a no-blame safety culture.*

#### 4.4.3. Leading metrics should be defined in an open, fair and transparent manner

Within the framework's no-blame safety culture the safety assurance regulator should have the authority to define the process for determining how data, metrics and thresholds are specified.

The most open, fair and transparent process would be data driven and both consultative and collaborative with the ADSE.

To prevent undue burden on the ADSE, standards for ADS data publication should be defined first. Recorded data from individual ADSE's should be aggregated, and anonymised, into single dataset held under the authority of the safety assurance regulator.

Open collaborative research conducted on this dataset should be used to inform the selection of the most appropriate metrics and associated performance thresholds. These should remain under continual review, and a change management process, as new data is acquired and new algorithmic tools developed.

*It is suggested that the safety assurance regulator should have additional powers to request leading metrics data is collected during pre-deployment verification and validation testing in all environments; simulation, proving grounds and on-road trials.*

#### 4.4.4. Leading Metric example – safety relevant traffic infractions

Consultation Question 22 of the consultation<sup>33</sup> addresses the whether the in-use safety assurance scheme should;

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

The *perception, decision, reaction* and *outcome explainability model* described above can be used to identify safety-related infractions and set thresholds upon leading metrics which define when driving behaviour is considered to have fallen below level of a competent and careful human driver and therefore could be considered careless or dangerous.

With reference to the previous example of safe passing of Vulnerable Road Users (VRUs)<sup>34</sup> the safety assurance regulator should have the powers to determine the thresholds at which the exhibited driving behaviour crossed between “careless” and “dangerous”.

For example, when passing a cyclist on the road;

- The defined thresholds for the proximity leading metric taken from the highway code recommendations would be;
  - < 1.5m under 30mph
  - < 2.0m over 30mph
- Associated thresholds for classification of a measured leading metrics could be;
  - Proximity 1.75m @ 30mph – Competent & Careful Driving (not recorded)
  - Proximity 1.45m @ 29mph – Careless Driving
  - Proximity 0.5m @ 50mph – Dangerous Driving

The benefit of the *perception, decision, reaction* and *outcome explainability model*, as being developed for standardisation within the ITU FG-AI4AD, is that the process, data and metrics can be applied internationally and independently of domestic traffic regulation.

This is possible by defining a minimal set of safety assurance metric that are based upon physics of the environment e.g. they are universal metrics based upon the time & spatial relationship between static and dynamics objects is used to determine safety critical near-miss events.

---

<sup>33</sup> Consultation Paper 3 page 179 para 11.24

<sup>34</sup> [Section 4.4.1 - Leading Metric Example – Safe Passing of Vulnerable Road Users](#)

Within this framework, common data and metrics can then subject to domestic policy and regulation which defines the *safety thresholds* at which sanctions or legal penalties apply. This approach aligns with application of the 1968 Convention on Road Traffic internationally agreed requirements, such as, “*every driver shall avoid behaviour likely to endanger pedestrians*”<sup>35</sup> which still permits the domestic regulators to define “endanger” for their own citizens.

Incorporating safety relevant traffic infractions, based upon domestic traffic rules, can also be incorporated into the *perception, decision, reaction* and *outcome explainability model* simply by extending the data published by the ADS and adding leading metrics specific to the rule infractions.

For the example of a red-light traffic infraction;

1. The “*perception data stream*” 3D world model should already include location and classification data for the white “Stop” line at a junction and also for “advanced stop lines” which protects cyclists positioned ahead of the traffic.
2. The 3D world model should be extended to include traffic light objects, locations, status and time of status change events.
3. New leading metrics should be defined for detection of white stop line crossings which require cross reference with traffic light status.
4. New leading metric should be defined that compares ADS perceived risk (*decision data stream*), ADS vehicle control inputs and vehicle dynamic response (*reaction data stream*) to traffic light status changes e.g. Green to Amber to Red (*perception data stream*).
5. The *outcome data stream* should then be extended to include red-light traffic violations and cross reference these with near miss event identification to indicate both *traffic rule infraction* and *endangerment*.

Additionally, thresholds for endangerment in relation to the traffic infraction also need to be considered e.g.

6. Thresholds should be defined at which driving may be considered careless or dangerous;
  - Encroaching past an advanced stop line
    - Careless: slow speed, small encroachment distance, no cyclists present
    - Dangerous: emergency braking, large encroachment, cyclists present
  - If the time of passing the stop line after red light activation is consider important to safety;
    - Careless: passed stop light < 0.5 seconds after activation
    - Dangerous: Passed stop light > 0.5 seconds after activation
7. Appropriate level of sanction for traffic infraction and threshold exceed would also need to be defined.

---

<sup>35</sup> 1969 Convention on Road Traffic – Article 21.1 Behaviour of drivers towards pedestrians



*It is suggested that the safety assurance regulator is given the powers to define which traffic infractions are deemed safety relevant and specify the appropriate leading metrics, thresholds and powers to access to ADS data require to assess the infraction.*

#### 4.5. Which entity should be responsible for processing of leading metrics?

Within the work of ITU FG-AI4AD three entities have been identified as potential entities responsible for consuming, processing and storing this behavioural safety assessment data;

1. The ADSE as part of a self-assessment programme (controlled by the safety assurance regulator)
2. The insurer as part of an independent assurance programme (controlled by the safety assurance regulator)
3. The safety assurance regulator or 3<sup>rd</sup> party representative (controlled by the safety assurance regulator)

Additionally, within the context of the proposed GB regulatory framework, a fourth entity could be defined for NUIC vehicles only;

4. The Licensed Fleet Operator – as part of their requirement to report near-miss and collision events (controlled by the safety assurance regulator)

The ALKS regulations states only the manufacturer and type-approval authority are responsible<sup>36</sup>;

*The manufacturer shall have processes to monitor safety-relevant incidents/ crashes/collisions caused by the engaged automated lane keeping system and a process to manage potential safety-relevant gaps post-registration (closed loop of field monitoring) and to update the vehicles. They shall report critical incidents (e.g. collision with another road users and potential safety-relevant gaps) to the type-approval authorities when critical incidents.*

However, there appears to be more flexibility to define the role of each entity with proposed regulatory framework so;

*It is suggested that the scheme regulators should be given the power to define the entity responsible for;*

- *The publication of the source data*
- *The processing of the source data to create leading metrics*
- *The management and update of thresholds set against leading metrics*
- *The generation of event triggers based upon these thresholds*
- *The storage of the source and processed data based upon event triggers*
- *The reporting of the leading metrics and events*

---

<sup>36</sup> ALKS Regulations Page 35 Section 3.5.4 <https://undocs.org/ECE/TRANS/WP.29/2020/81>

## 5. Comparisons between human and automated driving safety

In the joint workshop poll 80.7% of respondents agreed the proposal of Question 18.1 that<sup>37</sup>;

*“scheme regulators should be responsible for comparing the safety of automated and conventional vehicles using a range of measures”.*

*It is suggested that, while the question positions the comparison at the level of “vehicles”, it would be more appropriate to make the comparison at the level of “drivers” be them human, automated or in future remote, and their proficiency at executing the dynamic driving task in a competent and careful manner.*

### 5.1. Comparison at the level of Object and Event Detection and Response (OEDR)

Safe execution of the dynamic driving task requires situational awareness, risk-based judgements and risk mitigating actions in response.

For automated driving SAE J3016 defines these skill as Object and Event Detection and Response (OEDR) which includes the following tasks;

- *Monitoring the driving environment via object and event detection, recognition, classification, and response preparation (operational and tactical)*
- *Object and event response execution (operational and tactical)*

For human drivers object and event detection relies upon vision. Driver licensing requirements define minimum “standards of vision for driving” as a measure of the spatial resolution of the human visual processing system. It ensures that all drivers on the road have the visual acuity to discern the shapes and details of objects and the environment that are relevant to the driving task. For example the law states<sup>38</sup>;

- *You must be able to read (with glasses or contact lenses, if necessary) a car number plate made after 1 September 2001 from 20 metres.*
- *You must also meet the minimum eyesight standard for driving by having a visual acuity of at least decimal 0.5 (6/12) measured on the Snellen scale (with glasses or contact lenses, if necessary) using both eyes together or, if you have sight in one eye only, in that eye.*
- *You must also have an adequate field of vision - your optician can tell you about this and do a test.*

---

<sup>37</sup> Consultation Paper 3 page 169 para 10.83

<sup>38</sup> <https://www.gov.uk/driving-eyesight-rules>

The “standards of vision for driving” requirements for Lorry and bus drivers are more prescriptive representing the risk these vehicles present to other road users and occupants;

- *You must have a visual acuity at least 0.8 (6/7.5) measured on the Snellen scale in your best eye and at least 0.1 (6/60) on the Snellen scale in the other eye.*
- *You can reach this standard using glasses with a corrective power not more than (+) 8 dioptries, or with contact lenses. There's no specific limit for the corrective power of contact lenses.*
- *You must have an uninterrupted horizontal visual field of at least 160 degrees with an extension of at least 70 degrees left and right and 30 degrees up and down. No defects should be present within a radius of the central 30 degrees.*

Currently there are no minimum standards for OEDR performance and no minimum “standards of vision for driving” that the Secretary of State can use to define a self-driving vehicle is capable of safely driving itself in a given circumstance or situation.

The safety assurance regulator should be given the powers to define minimum standards for “vision for driving” and OEDR performance. The regulator must also be given the powers to monitor the real-world in-use OEDR performance. These in-use monitoring standards should also be applied within the pre-deployment approval phase as part of the automated driver licencing process.

#### 5.1.1. OEDR limitations of ALKS compared to a human driver and implications for self-driving categorisation

Section 7 of the UNECE ALKS Regulations<sup>39</sup> does cover sensing requirements for the OEDR for this very limited automated driving use case;

*The ALKS vehicle shall be equipped with a sensing system such that, it can at least determine the driving environment (e.g. road geometry ahead, lane markings) and the traffic dynamics:*

*(a) Across the full width of its own traffic lane, the full width of the traffic lanes immediately to its left and to its right, up to the limit of the **forward detection range**; this declared value shall be **at least 46 metres**.*

*(b) Along the full length of the vehicle and up to the limit of the **lateral detection range**; the declared range shall be sufficient to cover the full width of the lane immediately to the left and of the lane immediately to the right of the vehicle.*

The first thing to note is that ALKS has zero requirement for rear facing vision. This violates the intent of the 1968 Convention on Road Traffic Annex 5 Technical Conditions concerning Motor Vehicles and Trailers states;

---

<sup>39</sup> <https://undocs.org/ECE/TRANS/WP.29/2020/81>

*Chapter III, 47. Driving (rear-view) mirror*

*Every motor vehicle shall be equipped with one or more driving (rear-view) mirrors; the number, dimensions and arrangement of these mirrors shall be such as to enable the driver to see the traffic to the rear of his vehicle.*

A driver should check their rear-view mirrors every five to eight seconds<sup>40</sup> and The Police Driver's Handbook<sup>41</sup> and The Highway Code<sup>42</sup> both state that you must use your mirrors before changing speed.

Why are the rules and expectations for ALKS as an automated "driver" different? Why is it no longer necessary to check the road behind before moving off or while driving up to 37 mph on a motorway?

If current driving safety standards are to be maintained when ALKS is in use is the human driver required to complete the rearward OEDR? If the driver does need to complete the dynamic driving task OEDR then by definition ALKS cannot be categorised as a self-driving vehicle.

Recent US research found that 73% of automated mode crashes involve being struck from behind and that automated vehicles in autonomous mode were found to be struck from behind at 4.8 times the rate of human-driven vehicles<sup>43</sup>.

In the UK rear end collisions are the most common collision type representing 30% of all crashes in 2019<sup>44</sup>. The Association of British Insurers (ABI) estimate that there are;

*More than 1,500 whiplash claims are made in the UK every day, costing the insurance industry more than £2 billion a year – and adding £90 to the average annual motor insurance premium.*

The consultation paper makes specific reference to the difficulty in dealing with insurance claims related to rear-end collision whiplash injuries<sup>45</sup>. It was noted that without data from the automated vehicle claims that a self-driving vehicle stopped for no reason would be difficult to defend.

---

<sup>40</sup> <https://blog.motoringassist.com/news/using-information-mirror-checks/>

<sup>41</sup> Roadcraft – The Police Drivers Handbook Chapter 2 The System of Car Control page 27

<sup>42</sup> The Highway Code General Rule 161 <https://www.highwaycodeuk.co.uk/using-the-road.html>

<sup>43</sup> Comparison of automated vehicle struck-from-behind crash rates with national rates using naturalistic data <https://www.sciencedirect.com/science/article/abs/pii/S0001457521000877?dgcid=author>

<sup>44</sup> <https://www.ax-uk.com/news/rear-end-collisions-down-27-in-2020>

<sup>45</sup> Consultation Paper 3 page 292 para 17.36

As already mentioned, the UNECE DSSAD specification, and even the enhanced proposal to include both time and collision of activation, would not provide data about the collision event itself.

The *perception, decision, reaction and outcome explainability model* described in this paper would provide evidence of whether the ADS stopped suddenly due to a false positive (an phantom object not present in the real-world but present in the ADS digital representation of the world).

However, in the case of ALKS, if there are no rear-facing sensors the perception data stream would not contain any objects behind the vehicle making it difficult to reconstruct the rear-end collision and create any continual learnings.

The issue of rearward perception capability was also raised in the DoT ALKS call for evidence<sup>46</sup>;

*Whilst an ALKS-capable vehicle will not be able to pull over, it may be able to issue a transition demand to the driver. If listed under AEVA (as an automated vehicle), the driver would not be responsible for responding to the signalling of the police vehicle, only to a transition demand. However, there is no explicit requirement in the ALKS Regulation for the vehicle to possess rear-facing sensors. The vehicle may therefore struggle to 'know' to make a transition demand if it is being requested to stop by the police.*

***It is suggested that for a vehicle to be categorised as self-driving it should at least meet the OEDR standards of a competent and careful human driver and at minimum, to meet this requirement, the ADS must have forward, lateral and rearward perception capability.***

#### 5.1.1. OEDR limitations of ALKS tailgating scenario

The lack of rearward perception for ALKS raises another concern: tailgating by a following car.

In the UK more than 130 people killed or seriously injured in incidents involving people driving too close in 2018<sup>47</sup> and in relation to tailgating Highways England report that<sup>48</sup>;

- *Nearly 9 out of 10 (87%) drivers say they've experienced or witnessed it.*
- *Tailgating is a factor in 1 in 8 casualties on the strategic road network.*

---

<sup>46</sup>[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/921409/Safe-Use-of-Automated-Lane-Keeping-System-ALKS-Call-for-Evidence-FINAL-accessible.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/921409/Safe-Use-of-Automated-Lane-Keeping-System-ALKS-Call-for-Evidence-FINAL-accessible.pdf)

<sup>47</sup> <https://www.gov.uk/government/news/caught-on-camera-10000-tailgaters-spotted-in-just-two-weeks>

<sup>48</sup> <https://highwaysengland.co.uk/road-safety/stay-safe-stay-back/>

The research above showed that self-driving vehicles are 4.5x more likely to be hit from behind, perhaps one might also expect higher numbers of tailgating incidents too.

Why would tailgating become relevant for ALKS in a motorway scenario?

The ALKS regulation limits the vehicle speed to 60 km/h (37mph) on a motorway with a speed limit of 70mph. The intended function of ALKS is as a traffic jam pilot where actually traffic flow speeds are low.

What should the ALKS do at the end of this traffic jam zone when the traffic speeds increase back to 70mph? The regulation states<sup>49</sup>;

*(5.1.4) A transition demand shall not endanger the safety of the vehicle occupants or other road users.*

*(5.4.1) The activated system shall recognise all situations in which it needs to transition the control back to the driver.*

Given these requirements one might reasonably expect that, as the traffic returns to normal speed, the ALKS should recognise the situation and execute a transition demand. However, this behaviour is not clearly defined in the regulations and is not included in the ALKS testing criteria.

This raises a number of important questions;

- Would it be considered safe for the ALKS to accelerate from a standstill with the other traffic only to stop acceleration at 37mph when the road ahead remains clear?
  - Would this be an expected behaviour for the road users behind?
  - Would those road-user be guilty of a tailgating offence if their surprise at the self-driving vehicle behaviour resulted in a following distance lower than the safe following distance?
- Would it be considered safe for the ALKS to continue driving at the maximum 37mph in a middle lane of a motorway when surrounding traffic has returned to 70mph?
  - For how long would this behaviour be considered acceptable?
  - Would the middle-lane hogging behaviour of the ALKS be considered careless driving as per government fixed penalty notice guidance?<sup>50</sup>
  - By definition, the careless driving offence would mean ALKS fell below the standards of a competent and careful driver.
- If the ALKS stayed below it's maximum operating speed would it still be considered to be operating within it's specified ODD and therefore operating

---

<sup>49</sup> <https://undocs.org/ECE/TRANS/WP.29/2020/81>

<sup>50</sup> <https://www.gov.uk/government/news/new-penalties-to-tackle-tailgating-and-middle-lane-hogging>

- according to the regulation and therefore considered “safe”, even when the speed of the surrounding traffic is moving 33mph faster?
- Would the user-in-charge be expected to detect the change in traffic speed and disengage the ALKS?
    - If so, the human is completing the OEDR and is therefore a driver which would mean ALKS is not self-driving.

If the ALKS exhibits this driving behaviour it is reasonably foreseeable that it would introduce unnecessary risk to other road users, increasing the likelihood of both tailgating and lane change manoeuvres from following vehicles.

Even if the ALKS does recognise the “unplanned event” of traffic speed change and issues a transition demand then the following regulatory requirement applies;

*(5.4.4.1) In case the driver is not responding to a transition demand by deactivating the system (either as described in paragraph 6.2.4. or 6.2.5.), a minimum risk manoeuvre shall be started, earliest 10 s after the start of the transition demand.*

If we assume the unplanned event is detected when the ALKS vehicle reaches its 37mph maximum, then within the 10 second period before execution of a minimal risk manoeuvre, at a comfortable acceleration of 1m/s<sup>2</sup> the surrounding traffic would have already reached 60mph and the gap to the car immediately ahead would have increased by 50 metres.

The safety risk in this scenario is even higher because not only would the likelihood of tailgating be increased but the end of the transition demand the ALKS would execute a minimum risk manoeuvre (MRM). The regulation states that;

*During the minimum risk manoeuvre the vehicle shall be slowed down inside the lane or, in case the lane markings are not visible, remain on an appropriate trajectory taking into account surrounding traffic and road infrastructure, with an aim of achieving a deceleration demand not greater than 4.0 m/s<sup>2</sup>*

As stated above, a driver, before executing a speed change, is expected to check their mirrors. However, without rearward perception capability it's impossible for the ALKS to check the traffic behind before braking.

According to recent research the ALKS maximum deceleration of 4m/s<sup>2</sup> and other values above 2m/s<sup>2</sup> can be considered “aggressive driving”<sup>51</sup>. The values are depicted in the diagram below;

---

<sup>51</sup> [Toward a Comfortable Driving Experience for a Self-Driving Shuttle Bus](#)



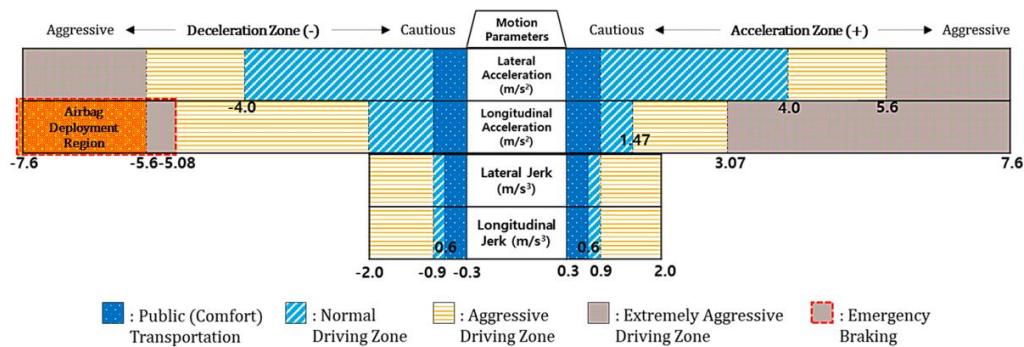


Figure 3 – Normal, aggressive, extreme and emergency acceleration levels

At the maximum 4m/s<sup>2</sup> deceleration the ALKS vehicle would go from 37mph to a complete stop in around 4 seconds, while the surrounding traffic lanes would have continued to accelerate to reach 69mph.

So, the additional questions raised include;

- Is this circumstance and situation is the ALKS executing the “*minimum*” risk manoeuvre?
- Would any of this behaviour be reasonably foreseeable for a following human driver? Especially if the following driver can see that the road ahead remains completely clear.
- Would this be considered that the ALKS was “safely driving itself” in this circumstance and situation according to the Secretary of State’s definition of self-driving?
- Would a rear-end collision that results from the actions of the ALKS have been reasonably foreseeable and preventable?
  - Arguably, yes. It would be great to see research to explore this issue further.
- Could the ALKS have taken other control options to mitigate the risk of a rear-end collision?
  - Yes. If the driver is not incapacitated, then continuing in lane at a constant speed would be another option.

Within the regulatory framework consultation it is proposed that<sup>52</sup>;

*...following the end of the transition demand period, the user-in-charge would re-acquire the legal obligations of a driver. Even if they have not taken control of the vehicle, they would be deemed to be a driver and their immunity for dynamic driving offences would cease.*

*In the case of ALKS, if a user-in-charge fails to take over, the vehicle would stop gradually in lane on a motorway with hazard lights flashing. It is a criminal offence for a person using a motorway to stop unnecessarily in an active lane. By failing to*

<sup>52</sup> Consultation Paper 3 page 198 para 12.33-12.35

*take over, the user-in-charge would become a driver and would be liable for that offence.*

*This does not mean that failing to respond to a transition demand would automatically lead to an offence. If the vehicle is able to park itself safely and legally at the side of the road, no offence will have been committed.*

Consultation Question 22 asks whether<sup>53</sup>;

*At the end of the transition demand period:*

- (1) the user-in-charge should re-acquire the legal obligations of a driver, whether or not they have taken control of the vehicle; and*
- (2) if, following a failure to respond to a transition demand, the vehicle stops in a manner which constitutes a criminal offence, the user-in-charge should be considered a driver and should therefore be liable for that offence.*

How would making the *driver* criminally liable for the ADSE's capability to *safely* execute a *minimum risk condition* appropriate to the *circumstances and situation* improve safety? How does it enhance a culture of continual learning? How does this approach address the "*radical asymmetry*" between the treatment of human drivers and ADSEs<sup>54</sup>?

Even the consultation states that<sup>55</sup>;

*Eventually, we would hope that transition demands would not be necessary: the vehicle would be able to achieve a minimal risk condition in all circumstances without endangering road safety or obstructing traffic flow.*

It seems to be recognised already that an ADSE's minimal risk condition may actually *endanger road safety and obstruct traffic flow*. If this is the case, would an ADS, approved to ALKS regulations, really meet the requirement of AEVA that a self-driving vehicle must be capable of "*safely driving itself*" for *circumstances and situations* that arise on a motorway<sup>56</sup>?

In order to differentiate between an ADS that is permitted to endanger to achieve a minimal risk condition (MRC) and an ADS that must achieve the same (MRC) safely, in what appears to be a legal separation between SAE Level 3 and SAE Level 4 driving automation, the following is proposed;

---

<sup>53</sup> Consultation Paper 3 page 199 para 12.37

<sup>54</sup> As raised by Gowling WLG in their response to Consultation Paper 1, see analysis, paras 7.154 to 7.164.

<sup>55</sup> Consultation Paper 3 page 199 para 12.36

<sup>56</sup> Automated and Electric Vehicles Act 2018 <https://www.legislation.gov.uk/ukpga/2018/18/enacted/data.pdf>

*As a matter of law, it is important to make a clear distinction between **transition demands**, to which users must respond, and “**transition information**” to resume driving, response to which is purely optional.*

*Under our provisional proposals, the distinction has legal consequences. Following a failure to respond to a **transition demand**, the **user-in-charge becomes criminally liable** for the consequences. By contrast, following **transition information**, the user-in-charge may resume dynamic driving but failure to do so **has no legal consequences**.*

Concerns over misleading consumer information and marketing of self-driving vehicles was raised in Chapter 10 of the consultation which referenced the German authorities taking action against Tesla for its use of the word “autopilot”. In 2020 the term “autonowashing” was coined to describe the gap in the presentation of automation and the actual system capabilities is put forth<sup>57</sup> which also referenced Tesla’s Autopilot and Full Self-Driving (FSD) products which are both SAE Level 2.

It would seem that the proposed regulatory framework would add further to this confusion by categorising vehicles as “self-driving” but then distinguishing between them based on safety critical capabilities and legal liabilities. The public/consumer has to understand four different categories of “self-driving”;

- *Not-self-driving vehicles;*
  - can be described as “full self-driving”
  - *legal responsibility* for the human to supervise and intervene in *all circumstances and situations* (Driver Assistance, SAE Level 2)
- *Self-driving vehicles (transition demand);*
  - can be described as “full self-driving”
  - *legal responsibility* for the human to supervise and intervene in *minimal risk manoeuvre circumstances and situations* (UIC, SAE Level 3 and ALKS)
- *Self-driving vehicles (transition information);*
  - can be described as “full self-driving”
  - *no legal responsibility* for the human to supervise and intervene under *any circumstances and situations* (UIC, ‘passenger’ mode, SAE Level 4)
- *Self-driving vehicles (no transition);*
  - can be described as “full self-driving”
  - *no legal responsibility* for the human to *supervise* and intervene under *any circumstances and situations* (NUIC, ‘passenger only mode, SAE Level 4)

From the above analysis it seems logical that if the human is legally responsible for safe execution of any part of the dynamic driving task, including the minimal risk manoeuvre, through a requirement to supervise or intervene in said dynamic driving task, then the

---

<sup>57</sup> Autonowashing: The Greenwashing of Vehicle Automation -  
<https://www.sciencedirect.com/science/article/pii/S2590198220300245>

vehicle does not meet the AEVA threshold of “safely driving itself” required to be categorised as “self-driving”.

*It is suggested that for a vehicle to be categorised as “self-driving” it must be capable of “safely driving itself” under all circumstances and situations of the entire dynamic driving task, including the execution of minimal risk manoeuvres, within the approved operational design domain. There should be no legal responsibility on the human to supervise or intervene during execution of the entire dynamic driving task and no liability for traffic infringements, including careless and dangerous driving.*

*It is suggested that, to ensure consumer confidence and clarity, there should be parity in the definition and meaning of “self-driving” and “safely driving itself” applied to the proposed UIC and NUIC categories.*

## 5.2. Digital Commentary Driving – the human equivalent of in-use monitoring of OEDR performance

Commentary Driving is a roadcraft technique in which the driver verbally describes what they can or can't see, the hazards present in the environment and the intended actions to mitigate the hazard risks<sup>58</sup>.

By listening to the commentary an independent observer, sitting in the vehicle, can assess the driver's level of situational awareness, the quality of their hazard and risk assessment and their decision-making in the selection of appropriate mitigating actions.

Commentary Driving is a great example of a continual in-use assessment of the drivers Object and Event Detection and Response (OEDR) performance. It is a technique that is widely adopted internationally for advanced and emergency response driver training, where research has shown it can improve hazard awareness and response.

Can commentary driving technique be applied for automated driving software?

Yes. A digital form of commentary driving is foundation for FG-AI4AD's proposed in-use behavioural assessment approach based upon the *perception, decision, reaction and outcome explainability model* described earlier<sup>59</sup>.

For automated driving the verbal commentary can be replaced by digital *perception, decision, reaction* data streams which are published by the automated driving software and consumed by an independent software observer who assesses the *outcome*.

---

<sup>58</sup> Roadcraft – The Police Drivers Handbook Chapter 3 Information, Observation and Anticipation page 51  
<http://www.roadcraft.co.uk/roadcraft/>

<sup>59</sup> [Perception, decision, reaction and outcome explainability model](#)

The independent software observer executes a number of leading metrics in order to assess the automated driving software's situational awareness, hazard and risk assessment and execution of mitigating action.

Digital Commentary Driving describes the replacement of the verbal commentary by a set of continuous data streams.

*It is suggested that the safety assurance regulator have the powers to define, and mandate the use of, a Digital Commentary Driving as part of the in-use safety monitoring scheme for automated driving.*

### 5.3. Digital Commentary Driving for human OEDR performance comparison

As described above Digital Commentary Driving can be used to monitor the performance of an automated driving system during execution of the dynamic driving task.

One additional advantage is that the technique can also be applied at times when the ADS is not engaged in the dynamic driving task. In the "offline" Digital Commentary Driving mode the ADS vehicle control outputs are not sent to the actuators. The offline mode can be used for two different purposes.

1. Assessment of the ADS performance – in which an expert human driver executes the dynamic driving task, while a digital observer monitors the ADS performance and an additional human observer compares performance to the real-world.
2. Assessment of human driver performance – in which the ADAS/ADS feeds data to digital observer to assessing driver behaviour.

#### 5.3.1. Digital Commentary Driving - assessment of the ADS safety in offline mode

Assessment of the ADS in offline mode has significant safety advantages.

It can be used during pre-deployment trials, licensing tests and period technical inspections.

It can also be used post-deployment at all times when the ADS is not engaged. This can be used to increase leading metric data and enhance safety case validation when the car is manually driven within the ADS ODD. Whilst, when used outside of the approved ADS ODD the additional data can provide real-world evidence to support ODD expansion.

During the offline assessment of the ADS an expert human driver always remains in active control of the vehicle and fully responsible for the dynamic driving task. However, in parallel that ADS *perception* (3D World Model) and *decision* (risk assessment) systems can still remain active. This means that a digital observer can still consume these data streams for safety assurance of the ADS situational awareness and hazard assessment.

In addition, a human observer, such as a driving test examiner, can compare their own model of situational awareness and hazard assessment with that generated by the ADS. In this scenario it would be beneficial if the data streams were converted into a human interpretable representation, such as a 3D environment visualisation or speech. This additional observer role is common practice within self-driving development testing.

*It is suggested that the proposed regulatory framework consider both the safety and continual learning advantages that offline ADS performance assessment can offer within a no-blame safety culture and that the safety assurance regulator have the powers to implement such processes within the scope in-use monitoring.*

#### 5.3.2. Digital Commentary Driving - assessment of the human driver safety using offline ADS

Assessment of the human driver using offline ADS data has significant advantages by enabling safety assurance for *all miles driven* not just automated miles driven.

This human performance mode can be easily understood as an extension to existing telematics-based insurance which already use vehicle dynamics to build behavioural models of drivers.

By extending this approach, using the ADS *perception data stream* (3D World Model), the driver's vehicle dynamic behaviours can be assessed in the context of the road traffic situation that was present at the time e.g. it would be able to identify if a hard braking incident the result of a safe emergency response to a child running into the road; or the result late response to a clear hazard due to distracted driving.

This human driver assessment using data generated by an offline ADS can also be used to monitor safety relevant traffic infractions, such as the earlier example of passing too close to a vulnerable road user. In a no-blame safety culture this data should focus on performance comparison and continual learning rather than prosecution and should require opt-in by the human driver as described below.

*It is suggested that the proposed regulatory framework consider both the safety and continual learning advantages of monitoring human driver performance using offline ADS can offer within a no-blame safety culture and that the safety assurance regulator have the powers to implement such processes within the scope in-use monitoring.*

#### 5.3.3. Opt-in for assessment of the human driver safety using ADAS

In addition to ADS equipped vehicles it would also be possible to extend the same human driver assessment approach can apply to lower levels of automation (SAE L0-2) which include environment perception systems.



This could even include leading metrics, from *outcome explainability*, on interventions such as AEB activations or the continual monitoring of range information within *perception data streams* to ensure safe following distances.

For the safety assurance regulator this additional human driving assessment using ADAS data would provide the most accurate, relevant and comparable data suitable for evaluating the in-use safety performance of all drivers, all vehicles and in all environments.

In order to ensure driver privacy, it is likely that an opt-in procedure would be required. Once the opt-in is complete this data could be used to significantly enhance the quality of comparison between human and automated driving.

It also opens up an interesting question of data ownerships and the rights the vehicle owner has over the data being collected by the onboard systems.

*It is suggested that the proposed regulatory framework consider both the safety and continual learning advantages of allowing human drivers to opt-in to the no-blame safety culture by providing safety assurance regulator access to in-use monitoring data of human driving performance. Further consideration should be given to any new regulatory powers required to ensure ADAS data is available after the opt-in permission is granted.*

#### 5.4. The value of an all-vehicle safety assurance dataset

An “*all-vehicle safety assurance dataset*” could be created by combining the suggestions above. While this may seem like an ambitious leap for the situation today it does appear there is unique opportunity to create a regulatory framework uses a no-blame safety culture to push further towards the 50by30 safe mobility target and beyond to vision zero by 2050.

This *all-vehicle safety assurance dataset* would have significant benefits in reducing the scarcity of safety critical event data and generation of associated leading metrics important for continual learning.

The *all-vehicle safety assurance dataset* provides a foundation for objective and meaningful comparison between automated and conventional vehicles across a range of measures. Meeting the expectations of 80.7% of workshop poll respondents who agreed with proposal of Consultation Question 18 that<sup>60</sup>;

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;

---

<sup>60</sup> Consultation paper 3 – page 169 para 10.83



The *all-vehicle safety assurance dataset* can also offer significant value to road network operators and City Mayors by providing leading metrics on the safety of roads, infrastructure and multimodal mobility. By visualising this leading metrics dataset as road network heatmaps the locations, times, conditions and road users classes at the highest risk can be determined and used to inform mitigation strategies through traffic controls or infrastructure planning.

This approach would complement the work of the International Road Assessment Programme (iRAP)<sup>61</sup> which recently launched the AiRAP initiative;

*Big data and artificial intelligence can play a lead role in building the social and economic business case for safer roads and create the scale of change needed to save millions of lives. The accelerated and intelligent collection and coding of road attribute data has the potential to reduce the time and effort required to undertake road safety assessments, reduce the costs and improve accuracy.*

*Together with its road safety Star Ratings and fatality and serious injury estimation models, iRAP's AiRAP initiative has the potential to put this road safety data at the fingertips of road authorities, policy-makers, investors and road users worldwide.*

By moving to leading metrics safety issues relating to the interaction between infrastructure and ADAS/ADS software can be identified before resulting in harmful outcomes.

For example, in one recent example in Somerset<sup>62</sup>, “a number of incidents involving cars leaving the road” were required before it was identified that some vehicles fitted Autonomous Emergency Braking (AEB) systems were unexpectedly executing emergency stops due to installation of a metal cattle grid in the road.

This is a great example where leading metrics on AEB activations would have identified a new safety hotspot immediately after the cattle grid installation. The higher frequency of these individual AEB activation events would have been detected much earlier than the lagging metrics requiring multiple car interactions resulting in cars leaving the road. While in this instance “no serious collisions” occurred, it clearly demonstrates the increasingly complex interactions between ADAS/ADS and road infrastructure which are essential to manage within the no-blame safety culture or continual learning. In this specific example Somerset Country Council were the entity that took mitigation action by reprofiling the road to reduce the likelihood of false positive AEB activations.

There is also value in the *all-vehicle safety assurance dataset* incorporating leading metrics from ABS, ESP, Traction Control activations but also specific algorithms that use low level

---

<sup>61</sup> <https://www.irap.org/project/ai-rap/>

<sup>62</sup> <https://www.bbc.co.uk/news/uk-england-somerset-55571080>

vehicle dynamics, such as wheelspeeds and accelerations, to identify changes in road surface friction, even before these negative outcome events occur. Recent collaborative work between Audi, NIRA Dynamics and HERE Technologies show how useful this data can be<sup>63</sup>.

*It is suggested that the safety assurance regulator be given sufficient powers to develop a no-blame safety culture that extends beyond automated vehicles and in doing so provide a framework for continual learning of safety benefit to all road users, ADAS development and infrastructure improvements. It is suggest that an all-vehicle safety assurance dataset would make a valuable contribution to this future.*

---

<sup>63</sup> <https://niradynamics.se/precise-data-for-greater-safety-nira-dynamics-launches-road-surface-alerts-with-audi-to-improve-slippery-roads-warning-system/>

## 6. Limitations, loopholes and liability gaps

The consultation paper is incredibly compressive and covers the full breadth of issues related to the creation of a regulatory framework for automated vehicle.

This section is intended to provide insights based upon more in-depth analysis on a narrow set of issues raised, which may nonetheless be seen as fundamental to the success of any future regulation.

The highlighted limitations, loopholes and liability gaps should in no way be seen as criticism of the outstanding work conducted by the Law Commissions. They should merely serve as additional considerations that can be used to further shape the proposed framework.

### 6.1. The “Full Self-Driving” Loophole for development and deployment

Recently released communications between the California DMV and Tesla Inc. highlight a significant loophole that exists in the regulation of “Full Self-Driving” systems classified as SAE Level 2 Partial Driving Automation<sup>64</sup>.

While the consultation raises important issues related to Consumer Information and Marketing, these are mainly concerned with the prevention the dissemination information that maybe “misleading or encourage irresponsible driving behaviour”. For example, should it be acceptable to market an SAE Level 2 system as “Full Self-Driving” if it leads to a public expectation that the system completes the “full OEDR” and therefore does not require human driver oversight?

However, there is a broader issue that the regulatory framework for automated driving should address;

*there is no technical difference between an ADS classified as SAE Level 2 and ADS classified as SAE Level 5. In fact, the only difference is liabilities.*

Or, within the context of the proposed GB regulatory framework;

*there is no technical difference between “not self-driving, driver assistance”, “self-driving with User in Charge UIC” and “self-driving with no-user in change NUIC”.*

Meaning that a vehicle with the technical capabilities to be categorised as “self-driving with no-user in charge” could instead be sold as a “not self-driving, driver assistance” vehicle.

---

<sup>64</sup> <https://www.plainsite.org/documents/242a2g/california-dmv-tesla-robotaxi--fsd-emails/>

This can be simply be achieved by forcing the driver to accept a liability waiver before the ADS can be activated.

In the case of Tesla's Full Self-Driving Beta the waiver includes acceptance that the system *"may do the wrong thing at the worst time"*.

This issues which may result are discussed in the sections below.

#### 6.1.1. When is "full self-driving" actually "not self-driving"?

The framework proposed a two-step approval process for a manufacturer or developer wishing to have an ADS approved for UIC or NUIC use.

This raises a few interesting questions;

- What happens if that manufacturer/developer chooses not to apply for approval because they believe a driver liability tick box deems their "full self-driving" system to be a "not self-driving/driver assistance" system?
  - How will these instances of non-application be captured?
    - Will the existing vehicle type approval process be able to identify these cases and have the authority to raise them to the ADS approval scheme?
- What happens when an ADS approval application results in a "not self-driving/driver assistance" categorisation?
  - Does that level of automated driving system sit completely outside of the no-blame safety culture and is exempt from sharing any safety performance data to aid continual learning?

*It is suggested that the regulatory framework provides clarity on the questions above relating to non-application for self-driving approval as well as whether some systems categorised as not self-driving should still fall within scope of the in-use safety assurance scheme.*

Consultation Question 9 proposes that<sup>65</sup>;

1. *unauthorised automated driving systems should be prohibited; and*
2. *this should be subject to an exemption procedure by which the Secretary of State may authorise unauthorised systems to be used in tests and trials.*

If it's as simple as including a driver liability tick box to have a "full self-driving system" categorised as "not self-driving" then neither the prohibition nor the exemption procedure for testing and trials would apply.

---

<sup>65</sup> Consultation Paper 3 – page 12 para 8.17

This is exactly the same loophole that exists within the California DMV regulation. By classifying the ADS as “not self-driving” the manufacturer/developer can bypass the requirement to apply for a “permit to test autonomous vehicles with a safety driver”<sup>66</sup>. This removes them from the following requirements;

- *pre-deployment testing under controlled conditions that determine if the ADS is safe to operate in the target ODD*
- *employing and training the safety driver*
- *reporting collisions*
- *reporting in-use leading metrics such as disengagement*

There are commercial considerations too, for example;

- *if permitted to test an autonomous vehicle with a safety driver the manufacturer may not receive a compensation for providing a ride to members of the public.*
- *whilst, a manufacturer that uses a driver liability tick box for their ADS is at liberty to ask a member of the public to pay \$10,000 to take part in full self-driving beta software testing.*

As more regulations, standards and best practices are defined for UIC and NUIC vehicles the incentive to use the “not self-driving/assisted driving” route becomes even more appealing.

*It is suggested that the GB regulatory framework for automated driving closes the loophole whereby a vehicle categorised “not self-driving” may be used for self-driving testing and be sold to a user who perceives the systems to be “self-driving” through experience, even if they’ve signed a liability waiver protecting the ADSE.*

#### 6.1.2. Drunk, drugged, drowsy & distracted – are self-driving behaviours too

A recent video of Tesla’s Full Self-Driving 8.2 Beta release being used in Oakland, USA<sup>67</sup> highlights in-use driving behaviours typically associated with drunk, drugged, drowsy & distracted human driving.

In just a single 30 minute drive;

- Does FSD accelerate confidently while actually trying to drive off the road and through a metal fence? Yes. See 12m46s.
- Does FSD hallucinate as if drugged? Yep. Sudden braking for objects that don’t exist at 2m35s.
- Does FSD lack judgement and risk awareness as if drunk? Sure. At 5m04s FSD decides to overtake a cyclist at the highest point of risk, crosses a double yellow line into oncoming traffic without slowing down.

---

<sup>66</sup> <https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/testing-autonomous-vehicles-with-a-driver/>

<sup>67</sup> <https://www.youtube.com/watch?v=antLneVlxc8>

- Does FSD ride up curbs as if distracted? Yes. At 9m41s FSD is only prevented from riding the sidewalk by rapid human steering.
- Does FSD fall asleep? Sure. FSD can disengage from driving at any point it chooses; 6m48s and 7m33s.
- Does FSD swerve as if drunk or drowsy? Yep. At 5m40s FSD gets a green light then staggers slowly into a four-way intersection unable to decide what to do until, cause other drivers to honk their horns before the human intervenes.
- Does FSD attempt to turn on “a no turn red light”. Sure. At 7m20s the driver has to intervene and take control to prevent it.

It seems clear that SAE Level 2 Partial Driving Automation could in fact be described as Impaired Driving Automation.

Human driving regulations provide deterrents against impairment expected to lead to careless and dangerous driving. It's clear from the video above that even with an alert driver supervising an SAE Level 2 Partial Automated Driving is engaged that instances of careless and dangerous driving may still occur before human intervention is possible.

*It is suggested that the proposed regulatory framework addresses the safety issues in which Partial Driving Automation systems can result in careless and dangerous driving even when a human driver is executing supervision and intervention as expected.*

#### 6.1.3. AVs for use only with Human Driver Supervision (HDS)

Building public trust in automated driving is recognised as being essential for adoption of automated driving. Both the technology and the legal framework in which it is regulated contribute to building public trust.

It is suggested that the “not self-driving/assisted driving” categorisation be altered to better reflect the function of SAE Level 2 Partial Driving Automation systems. This could be done by including a specific categorisation within the safety assurance scheme;

- AVs for use only with Human Driver Supervision (HDS)

Human Driving Supervision defines ADS that is not capable of “safely driving itself” without being “monitored by an individual”.

*It is suggested that the proposed regulatory framework recognises the risks that SAE Level 2 Partial Driving Automation systems create within the no-blame safety culture and need for these ADSEs to participate in the culture of continual learning.*

## 6.2. Limitations imposed by the UIC and NUIC definitions and categorisation

The consultation describes two paths in which automated driving might be developed<sup>68</sup>.

- *In Path 1, automated features are increasingly incorporated in vehicles sold across borders to a mass consumer market. Initially, these vehicles will continue to have a human in the driving seat though, as the technology improves, the human would be able to cede the driving task to the ADS in more circumstances.*
- *...in Path 2, vehicles are deployed without a human driver in limited local contexts, followed by a gradual expansion of their range of use.*

These definitions are not intended to be fixed or firmly defined and do not serve as legal definitions. However, each path has a safety critical decision point which needs to be determined;

- Path 1 the “*circumstance*” in which the human can “*cede the driving task to the ADS*”
- Path 2 the “*circumstance*” in which an ADS can be deployed without human fallback

*It is suggested that objective measures of OEDR capability is critical to determining both these circumstances.*

### 6.2.1. The OEDR continuum

Path 1 already has a long history. It started with the introduction of cruise control to maintain vehicle speed without driver throttle actuation as early as 1948 (or even 1908). However, these early systems lacked environmental perception and still required the driver to execute the full OEDR task.

In 1999 adaptive cruise control (ACC) started to *partially* complete of the OEDR task with the introduction of radar sensors to detect and respond the changes in speed of leading vehicles.

Path 1 represents a continuum increasing OEDR capability for automated driving systems. Whilst at the same time creates a diminishing requirement for human OEDR capability within the safe execution of the dynamic driving task.

The decision point at which the ADS OEDR capability is sufficiently to replace the human OEDR capability for safe execution of the dynamic driving task under a given circumstance or situation is central to the proposed regulatory framework.

*It is suggested that the regulatory framework for use of automated driving explicitly defines this dependency and ensures the safety assurance regulatory has the powers to*

---

<sup>68</sup> Consultation Paper 3 page 16 para 2.29



*define measurable OEDR capabilities which can be assessed during the approval process and verified through in-use monitoring.*

#### 6.2.2. Mapping OEDR capability to the circumstances and situations

The Automated and Electric Vehicles Act 2018 states that<sup>69</sup>;

*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.*

Defining the *circumstances* and *situations* in which a vehicle can be considered capable of *safely driving themselves* requires a mapping of OEDR capabilities to the definition of *safely* within these *circumstances* and *situations*.

The OEDR capabilities for *safely driving* within these *circumstances* and *situations* are independent whether the driver is a human or ADS.

An additional mapping is therefore required between these required OEDR capabilities and the specific OEDR capabilities of the ADS seeking approval or meeting in-use safety assurance requirements.

The consultation states<sup>70</sup>;

*...it would be crucial to distinguish between systems which can (and cannot) "safely drive themselves" without being "monitored by an individual".*

This is a binary decision (can or cannot) for each specific *circumstance or situation*.

The implication being that this binary decision is made at the point at which the human OEDR capabilities required executing the dynamic driving task under specific circumstances or situations are sufficiently met by the ADS to enable the vehicle to safely drive itself.

The circumstances and situations may be specific to the intended operational design domain (ODD) for which the ADS is seeking approval, or has been approved, for use.

*It is suggested that the safety assurance regulator should be responsible for, and have the necessary powers, to collect data to enable the following mappings;*

---

<sup>69</sup> Automated and Electric Vehicles Act 2018 <https://www.legislation.gov.uk/ukpga/2018/18/enacted/data.pdf>

<sup>70</sup> Consultation Paper 3 page 29 para 3.53

- *Minimum ODER capabilities that define safely driving for a specific set of circumstances and situations*
- *ADS OEDR capabilities to these minimum ODER capabilities*
- *Circumstances and situations to specific ODDs*

#### 6.2.2.1. OEDR, MRMs, MRCs, ODD and stopping in lane

These OEDR capabilities required under specific circumstances or situations should include the definition of minimal risk manoeuvres (MRM) and minimal risk conditions (MRC) which may be specific to the circumstances or situations (or ODD).

The consultation raises the issue of whether “coming to a gradual stop in lane”, should a User-in-Charge fail to respond to a transition demand, is a safe MRM resulting in a safe MRC.

The example cited for motorway driving, relevant to ALKS, is that<sup>71</sup>;

*“...ideally, drivers should exit the motorway or reach an emergency refuge area, the UK Government advises drivers who cannot do this to “move into the left-hand lane and put on your hazard lights. This suggests that even careful and competent drivers cannot always be expected to move out of an active lane of traffic.”.*

This does indeed suggest that if a driver is already present in the left-most lane then this would result in a gradual stop in lane. However, there would need to be good reason for not exiting the motorway, reaching an emergency refuge or using a hard shoulder.

What it does not suggest is stopping that in a lane to the right of the left most lane of a motorway is an acceptably safe behaviour for a competent and careful driver.

As described, above, if a DDT critical issue occurs while the driver in *any* to the right of the left most lane they are expect to execute a lane change manoeuvre to enter into the left-hand lane (this may be a hard shoulder or indeed active lane on a smart motorway).

*It is suggested that the safety assurance regulator must have the authority to define MRMs and MRCs which are consider as acceptably safe for a given circumstance and situation (such as an ODD). They should also be responsible for ensuring the defined MRM/MRCs are aligned to the expectations placed upon all other road users and remain predictable to all other road users.*

*It is suggested that the safety assurance regulator conducts further research and gathers sufficient evidence before determining whether “coming to a gradual stop in lane” can ever be considered a safe MRM/MRC for a motorway driving.*

---

<sup>71</sup> Consultation Paper 3 page 53 para 4.84 (3)

### 6.2.3. Path 1 UIC prevents passenger only or unoccupied modes of operation

The legal actors' diagram<sup>72</sup> defines Path 1 as; "AVs that can only be used with a User-In-Charge (UIC)" imposing a restriction that they must have "a human in the vehicle or insight of the vehicle, with access to the controls" who must;

- *Be qualified and fit to drive*
- *Be receptive to a transition demand*
- *Maintain and insure the vehicle*
- *Report accidents*

The definition of UIC would prevent Path 1 vehicles from being used in a "passenger only mode" or in an "unoccupied mode".

It would appear that the Path 1 UIC definition has been written to accommodate ALKS or SAE Level 3 Conditional Driving Automation systems. However, in doing so it removes many significant benefits of self-driving in which the driver becomes a "*passenger*".

The 2019 "CAV public acceptability dialogue: Engagement report" for the Department for Transport identified six key conditions which government and other stakeholders should consider when overseeing the introduction of CAV technology<sup>73</sup>;

1. *CAVs must be proven to be safe and secure*
2. ***Benefits must be accessible to all***
3. *CAVs must be good for society and good for jobs*
4. ***People must remain in control of their transport choices***
5. *There must be clear guidance on accountability*
6. *New bodies for oversight should be created*

The current definition of UIC appears to be in contradiction with key conditions (2) and (4) which will be explored further below.

#### 6.2.3.1. *CAV benefits must be accessible for all*

When describing "**benefits must be accessible for all**" the CAV public acceptability dialogue conclusion was that participants;

---

<sup>72</sup> Summary of Consultation Paper 3 - *Figure 1.1 - Overview of the roles of the user-in-charge and licensed fleet operator*

<sup>73</sup>[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/951094/cav-public-acceptability-dialogue-engagement.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/951094/cav-public-acceptability-dialogue-engagement.pdf)

*“insisted that automated vehicles **should not be allowed to reinforce current inequalities**. Consequently, there was a strong demand for government to ensure the introduction of CAVs would be **equitable** and would take into account the need of **minority groups**.”*

The UIC restrictions imposed for on the introduction of privately owned CAVs creates the exact the opposite situation;

- it only benefits existing drivers
- it reinforces current inequalities
- ignoring the needs of minority groups
- would not lead to equitable introduction of the technology

If a UIC has to be “qualified and fit to drive” this continues to exclude those individuals considered ineligible to take a driving test upon medical or disability grounds. It would also exclude around 70,000 drivers each year that have their licenses revoked on medical grounds<sup>74</sup>.

#### 6.2.3.2. *The Designated Passenger*

*To ensure the equitable mobility requirement that “CAV benefits must be accessible for all” it is suggested that the regulatory scheme for Path 1 vehicles be extended to include “passenger only” modes of operation. More consideration should be given to how this could be achieved but one suggestion could be the inclusion of a new legal definition for Designated Passenger (DP).*

The Designated Passenger would;

- not need to be qualified and fit to drive,
- not need to be receptive to a transition demand
- met be required to report accidents

The Designated Passenger could still have additional responsibilities that may include;

- in-use safety such as ensuring a child under 14 years old is wearing a seatbelt
  - it should be noted that this requirement be equally applicable for NUIC passenger only rides

Using the Designated Passenger would also enable;

- the DP minimum age limit to be set lower than for a licenced driver,
  - for example, at 14 years old a passenger is considered responsible for their own safety belt use.

---

<sup>74</sup> <https://motorway.co.uk/blog/2019/09/25/uk-licence-revocation>

As well as covering SAE Level 3 *transition demands* the User-In-Charge definition appears also to cover SAE Level 4 journeys with restricted ODDs that require a human driver to complete at least part of the journey e.g. a motorway only automated driving function without transition demand (described elsewhere as *transition information*).

However, the limitation of the UIC definition is that it explicitly prevents anyone other than a driver license holder from using a SAE Level 4 function capable of making a door-to-door journey without any human intervention.

Can we really call these vehicles “self-driving” if a user always requires a “driving licence”?

The UIC designation also significantly reduces many of the perceived benefits of self-driving vehicle. For example;

- Sleeping is excluded under the requirement to be receptive to a transition demand
- Using your own vehicle as a taxi service to get to a hospital or return from evening drinks is excluded under the requirement to fit to drive
- Sitting in a more comfortable or more social rear seat position is excluded under the requirement to be receptive to a transition demand as well as the need to have access to the controls

So, the ability to create a Designated Passenger status, or similar, not only improves equitable access for excluded user groups will increase the freedom of mobility choices available for privately owned self-driving vehicles.

#### 6.2.3.3. *People must remain in control of their transport choices with CAVs*

The CAV public acceptability dialogue report concluded that, with the introduction of CAVs, “people must remain in control of their transport choices”. Which was described as;

*It would be important for future travel to be **at least as convenient and flexible as current options** to convince people to change. Participants were reluctant to cede control over journey planning as they did not trust CAVs to choose the best, most reliable, or preferred routes. In addition, **many participants associated vehicle ownership with personal freedom**, and feared that **shared CAVs would be much less convenient**.*

An argument raised against altering the Path 1 UIC only designation is that excluded groups have the “option” to access CAV mobility through use a Path 2 NUIC approved vehicles operated by Licensed Fleet Operators.

However, by imposing this restriction, the proposed regulatory framework limits both choice and personal freedom in relation to the private ownership of self-driving vehicles. It also forces vehicle owners to enter into service contracts with Licensed Fleet Operators

before they have the option to use *their* self-driving vehicles in a passenger only or unoccupied mode.

It is also clear that, even if the regulation forces a self-driving vehicle owner into a contact with a licenced fleet operator, the framework itself offers no protection on the service contract terms and conditions.

Rather than having personal mobility freedom through ownership of a self-driving vehicle, the service contract means mobility options are at the discretion of the Licenced Fleet Operator e.g. it becomes their choice of how and when to offer the service which may result in self-driving unavailability for the owner at a time of need even though their own vehicle is available.

Left unregulated, these service contracts may further perpetuate inequitable access to mobility e.g. if Licensed Fleet Operator select journeys that offer the highest financial return rather than those with the highest social need e.g. those of fleet owned vehicle rather than personally owned; or choosing to monitoring short simple city journeys compared to long complex rural journeys etc.

*It is suggested that final regulatory framework considers options that enable privately owned self-driving vehicles to be used for passenger only and unoccupied journeys without forcing owners into unregulated contracts with Licensed Fleet Operators. It is suggested that this is fundamental to meet the requirement that with the introduction of CAVs "people must remain in control of their transport choices".*

*6.2.3.1. Why must private Self-Driving Vehicles owners give up their privacy for to access passenger only modes*

As participants noted in the report, private vehicle "ownership is associated with personal freedom".

Control over the when, where and how the vehicle is used for personal journeys is essential to maintain this sense of autonomy in mobility. Private ownership and private use are by their very nature private. You do not need to inform anyone when you decide to make a journey.

The current proposal removes this sense of freedom, control, autonomy and privacy for privately owned self-driving vehicles used in passenger only or unoccupied modes.

The contract required with a Licenced Fleet Operator to enable passenger only modes could result in the owner having to seek permission for every journey. In addition, it means they are forced to agree to handover personal journey information regarding start and end locations in addition to being forced to agree to continual remote surveillance throughout the trip.

Most concerning is that these intrusions on privacy would disproportionately affect the most vulnerable and marginalised groups who are already excluded from accessing private mobility options due to ill health or disability e.g. they are forced into a contract with a Licenced Fleet Operator whether using a UIC or NUIC.

Future of Mobility Minister Jesse Norman stated, in 2019, that<sup>75</sup>;

*"Self-driving technologies could greatly improve the mobility of vulnerable user groups, helping to address problems of isolation and loneliness across the country. The needs of older people, and those with visible or hidden disabilities, must be at the heart of all new modes of transport."*

*It is suggested that these issues of equitable mobility are reviewed and addressed in the final proposals for automated driving regulation. In particular that self-driving technologies extend mobility freedom and privacy to those unable to hold a driving license without forcing them into restricted licensed fleet operator services.*

#### 6.2.4. UIC does not permits unoccupied modes such as Valet Parking

The legal actors' diagram<sup>76</sup> incorrectly cites "valet parking" as a use case for which a User-in-Charge (UIC) may complete part of the journey.

The user expectation of "valet parking" is both leaving the vehicle and no longer being in control of the vehicle. For conventional vehicles it's the valet parking attendant fulfils the full DDT upon drop off and return. The user expectation of "valet parking" would therefore be that the ADS completes this full DTT function.

However, if a UIC must be a "human in the vehicle or insight of the vehicle, with access to the controls" then automated valet parking would be prohibited function.

To be used "out of site", in an unoccupied mode, the proposed framework would require the vehicle to be categorised as NUIC and would require the self-driving vehicle owner to have a contract with a Licensed Fleet Operator.

##### 6.2.4.1. Valet Parking – can a single vehicle be both UIC and NUIC?

Valet parking is a personally owned vehicle function which enables the most convenient drop off location, whilst enabling more efficient use of nearby parking facilities yet still maintaining fast on-demand vehicle availability for the next journey.

---

<sup>75</sup> <https://www.gov.uk/government/news/accessibility-must-be-at-the-heart-of-new-transport-tech>

<sup>76</sup> Summary of Consultation Paper 3 - Figure 1.1 - Overview of the roles of the user-in-charge and licensed fleet operator



Under the current framework “valet parking” would require NUIC categorisation and Licensed Fleet Operator monitoring.

This raises important issue in the way the responsibilities of the legal entities are defined.

The definition on Path 1 and Path 2 vehicles is an open flexible description. However, the resulting legal entities of UIC and Licensed Fleet Operator appear prescriptive in their obligations and responsibilities. Which opens some interesting questions;

- If a self-driving vehicle is privately owned and has functions categorised as UIC can it also be designated as a NUIC for other functions, such as valet parking?
- If so how are responsibilities intended to change between the User-In-Charge and the Licensed Fleet Operator?
  - For example, the responsibility to “maintain and insure the vehicle” and well as “report accidents and near-misses” must currently be assigned to either the UIC or the Licensed Fleet Operator.
  - For a personally owned vehicle, used for valet parking mode, can the Licensed Fleet Operator be expected to “maintain the vehicle”? Should this remain the responsibility of the private vehicle owner?
  - Would separate insurance policies be required for the single vehicle by two different legal entities? Or should only the private vehicle owner insure?

A note on reporting;

The responsibility for reporting accidents and near miss events was mentioned early in this feedback. It critical to remember that no entity can report without an event first being detected. The earlier suggestion was that only the ADSE has access to the required data from the ADS to detect events, so it logically follows that the ADSE should also have the legal responsibility to report.

If this earlier recommended is approved then reporting of accidents and near misses in the “valet parking” use case becomes simpler, it would be the ADSE who is responsible for reporting during both the occupied and unoccupied phases of the journey.

As noted earlier, if the current framework proposal is maintained, then in order for the UIC or Licensed Fleet Operator to fulfil their legal requirement to report, the ADSE must also be under a legal obligation to share accident and near-miss event data with the other legal entities.

*It is suggested that the regulatory framework clarifies the valet parking use case and at the same time clarifies whether a vehicle can indeed be categorised at both UIC and NUIC to enable a privately owned vehicle to be used in such an unoccupied mode.*

#### 6.2.5. UIC & NUIC definitions may restrict or distort competitive markets

The prohibition of passenger only and unoccupied modes for personally owned Path 1 self-driving vehicles, using the UIC designation, may result in the proposed regulatory framework restricting or distorting competitive markets.

Under the current proposal “passenger only” and “unoccupied” modes would only be available to a NUIC vehicle under the supervision of a Licenced Fleet Operator.

The “passenger only” and “unoccupied” modes are key selling features for a privately owned self-driving vehicle. If a manufacturer and ADSE can deliver these features safely and directly to the consumer then this would create open and fair competition with Licensed Fleet Operators services.

Consider a family celebration at a restaurant in which alcohol consumption is involved. There are currently two passenger vehicle options available;

- a. Nominate a designated driver who will not drink.
  - They must drive to the venue, park and drive home.
- b. Take a taxi to and from the event operated by a licenced taxi service.

How does the introduction of CAV technologies using the proposed regulatory framework change the situation? It doesn't;

- a. Under the UIC designation a “designated user-in-charge” is still required to be fit to drive
- b. A robotaxi NUIC vehicle operated by a Licenced Fleet Operators still operates exactly like a taxi.

However, the introduction of privately owned self-driving vehicles with “passenger only” and “unoccupied” modes opens up new options for CAV mobility for the situation;

- a. Nominate a designated driver who drives to and from the venue but uses an unoccupied “automated valet parking” mode to drop off and pick up outside the restaurant
- b. Drive to the restaurant, use unoccupied “automated valet parking” and then for the return journey engage a “passenger only” mode if unfit to drive.
- c. Use “passenger only” mode to and from the restaurant and unoccupied “automated valet parking” which keeps the vehicle local but incurs parking fees.
- d. Use “passenger only” mode to and from the restaurant and “unoccupied” mode which enables the vehicle to return home to park where no additional fees are incurred before returning on-demand to the restaurant.

There are potentially wider benefits for enabling “unoccupied” mode for personally owned self-driving vehicles. One benefit is that could be used to create more single car households.

For example, where a privately owned vehicle is used for the commute of one member of the household, an unoccupied mode would enable the vehicle to return home afterwards for use by another for the remainder of the day. It seems clear that, in this scenario, the initial household cost premium for this “unoccupied” feature could be offset against the household cost saving from operating only a single vehicle. The reduction in household car numbers would also reduce the requirement for private and public parking space.

If a manufacturer and ADSE can offer a privately owned self-driving vehicle with “passenger only” or “unoccupied” modes that can operate safely without requiring “remote supervision” this would further increase market competition.

*It is suggested that the proposed regulatory framework ensure that definitions of legal entities, services that can be offered by UIC and NUIC vehicles, do not restrict or distort the competitive market around “passenger only” or “unoccupied” mobility modes. It is suggested that, in fact, the framework should create new more equitable mobility options.*

#### 6.2.6. Policing limitation of not-self-driving, UIC and NUIC vehicle categorisations

The SAE Levels of Driving Automation refer to modes of operation for specific functions, whereby a single vehicle can include functions across all defined levels;

- Level 0 – Anti-lock brakes systems (ABS), Electronic Stability Control (ESC), Autonomous Emergency Braking (AEB)
- Level 1 – Adaptive Cruise Control (ACC) or Lane Keeping Assist (LCA)
- Level 2 – Supervised Highway Pilot and Autopark
- Level 3 – ALKS traffic jam pilot, AD with human fallback
- Level 4 – Highway Pilot, AD for specified circumstances or situations
- Level 5 – Chauffeur, AD for all circumstances or situations

As such, there is no such thing as an SAE Level 4 vehicle, only a vehicle which may have a function that operates at SAE Level 4.

It should be noted that some functions move upwards as the levels advance e.g. while AEB is a Level 0 function its purpose is collision avoidance, which means under J3016 it would become the responsibility of the ADS at Levels 3-5.

It should also be noted that from Level 2 upwards the automated driving system has both longitudinal and lateral control on a sustained basis i.e. it can execute the DDT. Although at the different levels the ADS may require supervision, fallback readiness or ODD restrictions to ensure safety.

If current trends continue, it's likely that personally owned vehicles will be sold with fixed hardware capabilities and upgradeable software functionality e.g. the owner/user will pay extra to unlock more capable automated driving functionalities.

This software upsell business model is likely to mean that a single vehicle on the road could be categorised as not-self-driving, UIC and NUIC depending upon which software is installed and which modes it enables.

There have already been reports in the US where a Tesla as configured included enhanced Autopilot and Full Self-Driving but, upon resale, these software features were remotely disabled reducing the functionality to limited driver assistance<sup>77</sup>.

Within the proposed regulatory framework, how would it be possible to identify which software functionality is available within the vehicle and which functions are enabled at any given time?

*It is suggested that within the proposed regulatory framework it should be possible to identify which software functionality is available within the vehicle and which functions are enabled at any given time and location. It is suggested that, at minimum, this may require the DSSAD to record an additional data field which indicates which automated driving mode is enabled.*

#### *6.2.6.1. Failure to stop for the police – who is the “driver” for UIC or NUIC?*

If a police constable or traffic officer requests a driver to stop, they have a legal obligation to do so and face criminal charges if they do not.

Who is the “driver” for a UIC or NUIC?

The proposed regulatory framework is unclear as to whether stopping is the responsibility of the ADSE whose ADS is executing the dynamic driving task? Or whether the stopping responsibility rests with the User-In-Charge or the Licensed Fleet Operator?

*It is suggested the legal responsibility to stop the vehicle upon police constable or traffic officer request should be clarified in the final regulatory framework and included on the legal actors' diagram.*

*It is suggested that the detection of the police request to stop and the execution of the required manoeuvre to “pull over and stop as soon as it is safe to do so” are elements of the dynamic driving task carried out by the ADS, so the ADSE is the logical legal entity responsible.*

Consider a situation where an unoccupied NUIC self-driving vehicle is on a motorway and is being requested to stop but fails to do so. What should happen next?

---

<sup>77</sup> <https://jalopnik.com/tesla-remotely-removes-autopilot-features-from-customer-1841472617>

- Should the police be able to use V2X communications to send a direct command to an ADS to initiate a Minimal Risk Manoeuvre (MRM)?
- Should the police be required to contact the Licensed Fleet Operator to request a remote MRM command?
  - Does the Licensed Fleet Operator then have a legal responsibility to comply?
- Should the police be required to contact the ADSE to request a remote MRM command?

It was noted in the UK ALKS Call for Evidence that<sup>78</sup>;

*Whilst an ALKS-capable vehicle will not be able to pull over, it may be able to issue a transition demand to the driver. If listed under AEVA (as an automated vehicle), the driver would not be responsible for responding to the signalling of the police vehicle, only to a transition demand. However, there is no explicit requirement in the ALKS Regulation for the vehicle to possess rear-facing sensors. The vehicle may therefore struggle to 'know' to make a transition demand if it is being requested to stop by the police.*

*ALKS must also be able to respond to other signals from Police / DVSA / Traffic Officers, which may include flashing amber lights, flashing red lights, or flashing headlamps.*

The lack of rearward facing perception required for ALKS was mentioned earlier in this paper is a major limitation and a contravention of the 1968 Convention on Road Traffic requirement to have rear facing mirror (for the purposes of rearward perception).

***It is suggested that the proposed regulatory framework makes clear the powers that the police, Driver and Vehicle Standards Agency officers or Traffic Officers will require in relation to stop requests, whether these require new communication procedures and to which legal entity they apply.***

#### *6.2.6.1. Sleeping & self-driving, how can police identify permitted behaviours?*

The proposed regulatory framework is unclear as to whether sleeping is permitted as a UIC or excluded under the "fit to drive" requirement. If the proposal is that UIC enables use of ALKS then these regulations mandate the use of a Driver Availability Recognition System which would detect eye closure as unavailable e.g. sleeping is not permitted. However, this is not made explicit in the current regulatory framework and the potential to have a UIC

---

<sup>78</sup> Safe Use of Automated Lane Keeping System ALKS Call for Evidence page 25 para 3.36 and 3.27  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/921409/Safe-Use-of-Automated-Lane-Keeping-System-ALKS-Call-for-Evidence-FINAL-accessible.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/921409/Safe-Use-of-Automated-Lane-Keeping-System-ALKS-Call-for-Evidence-FINAL-accessible.pdf)

operate in a mode which requires “*transition information*” rather than “*transition demand*” still remains.

So, consider the following scenario;

*A car is driving on the motorway. The DDT is being executed by an automated driving function and the occupant is asleep in the driver's seat.*

This simple scenario raises a number of important questions which do not appear to be addressed by the proposed regulatory framework;

- How will the police know if this is a permitted behaviour?
- Would a police ANPR check extend it's query the Secretary of State's list of “*all vehicles designed or adapted to be capable, at least in some circumstances or situations, of safely driving themselves*”?
  - Is the intent to permit a vehicle to only be categorised as UIC or NUIC?
    - Even if categorised at UIC there may still be times when the vehicle is being used in a not-self-driving mode.
  - If the vehicle is fleet owned, categorised as NUIC and insured;
    - The Licensed Fleet Operator must be operating remote supervision. How can the police officer check the ADS is being actively supervised - without which they cannot ensure the safety of the passenger?
      - Would this require a request to stop? Followed by verbal communication between the police and the remote operator?
        - Could this verbal communication occur without a request to stop?
      - How can the police be sure that remote supervision was active at the time the request was made and has remained active for the entire journey?
  - If the vehicle is privately owned, insured and listed as both not self-driving and UIC;
    - The police need clear guidance as to whether sleeping in a UIC mode is permitted.
      - If sleeping is permitted for a UIC “transition information” mode how will the police know if the vehicle is operating in this approved mode rather than a “transition demand” or not-self-driving mode (such as an SAE Level 2 Autopilot).
      - How will the police know that the UIC mode is in-use for an approved ODD?
  - If the vehicle is privately owned, and listed as having not self-driving, UIC and NUIC the modes of operation;
    - How can the police tell which mode of operation the vehicle is in?
    - If they stop the vehicle, how can they tell which mode of operation the vehicle was in at the time of the request?

- If the vehicle was in NUIC mode how will the police know if the owner has a valid contract with the Licensed Fleet Operator?
  - And that the operator was actively supervising the system use?
- If the vehicle was in a UIC mode how can the police know if the system is approved for use within the current ODD?

*It is suggested that the regulatory framework specifically addresses the issues surrounding sleeping in the drivers seat and how it would be possible for the police to determine whether this is a permitted behaviour for the specific mode in operation for the given ODD.*

6.2.6.1. Causing death by careless driving when under the influence of drink – is the ADSE responsible when the UIC is drunk in charge?

As consultation paper one points out<sup>79</sup>;

*Under the current law it is a criminal offence to be in charge of a motor vehicle while unfit to drive through drink or drugs. It is also an offence to be in charge of a motor vehicle with alcohol levels over the prescribed limits.*

It was also noted that it is highly likely that these offences would already apply to a user-in-charge.

Earlier background analysis in the consultation covered offences of causing death or injury by driving which currently apply to human driver<sup>80</sup>. In relation to these offences with a User-In-Charge the following conclusion was drawn;

*...if an uninsured or disqualified driver was monitoring a vehicle at SAE Level 3 which malfunctioned, causing V's death. The driver would be guilty of causing death if they did something which could properly be criticised and which contributed in more than a minimal way to the death. If, however the fault was entirely with the automated system, the driver would be guilty of driving while uninsured or disqualified, but not of causing death.*

*It is suggested that the final regulatory framework clarifies if a user-in-charge being unfit to drive due to drink or drugs would be treated the same as an uninsured or disqualified driver in the case of a fatal collision. While also additionally clarifying whether if a fatal collision occurs while the ADS was executing a minimal risk manoeuvre (MRM) after a failed transition demand they would remain liable for their "inaction".*

---

<sup>79</sup> Consultation Paper 1 page 134 para 7.40

<sup>80</sup> Compiled background paper 2 - Offences of causing death or serious injury on the roads page 33



In a no-blame safety culture, it is acceptable that any user-in-charge is blamed, and faces 14 year in prison, for a fatal collision resulting during the ADS execution of the MRM, whilst the ADSE faces zero liability?

In a situation similar to the “soft-brush” example a drunk driver may engage the automated driving system immediately after the collision in an attempt to avoid liability. This may again point to the need to record both the time and location at which the system was activated – which in this case could be the difference between 3 months and 14 years in prison.

However, as described earlier this late activation would place them at the scene, where otherwise this would not have been clear. In fact, it's a more likely scenario that the drunk driver would activate the ADS as early as possible so that they are liable any collision for the smallest amount of journey time possible.

It seems clear that the ability for a drunk driver's ability to shift the liability towards the ADSE while the ADS is engaged would encourage, rather than discourage, drink driving.

*It is suggested that the final regulatory framework includes an impact study on the likely effects of user-in-charge automated driving on the expected change in drink driving rates and related safety consequences of a drunk driver attempting, or failing, to responding to a transition demand. It is suggested that the impact study compares these findings against “passenger only” or “transition information” modes where safe execution of the MRM remains the responsibility of the ADS.*

#### *6.2.6.2. Drink driving, drunk in charge – how can police identify offenses?*

The penalties for drunk in charge or drink driving also create an incentive to have the ADS active as often as possible;

##### **Being in charge of a vehicle while above the legal limit or unfit through drink**

You may get:

- 3 months' imprisonment
- up to £2,500 fine
- a possible driving ban

##### **Driving or attempting to drive while above the legal limit or unfit through drink**

You may get:

- 6 months' imprisonment
- an unlimited fine
- a driving ban for at least 1 year (3 years if convicted twice in 10 years)

If the vehicle is being operated in a “passenger only” mode for the entire journey, then neither of these offences apply. However, it would be useful to clarify if “attempting to drive” would apply if at some stage in the journey the passenger is expected to become a driver.

Where the framework permits a single vehicle model to be categorised as not-self-driving, UIC and NUIC this raises a number of important questions in relation to unfit to drive;

- If the police request a vehicle stop how will they know in which mode the vehicle was being operated?
- If the occupant in the driver's seat claims the privately owned vehicle was in a "passenger only" or NUIC mode, how can the police verify this statement?
  - Would the police be given the powers to interrogate the onboard ADS/DSSAD on the spot?
  - Would the police be required to contact the ADSE or Licenced Fleet Operator for this information?
  - Would the police have the powers to breathalyse a "passenger" if they are unable to identify the mode of operation?
- If the occupant in the driver's seat claims the privately owned vehicle was in a UIC mode, how can the police verify this statement to identify if they were driving or a just "in-charge".

*It is suggested that the proposed legal framework clarifies if there is a need for additional police powers to enable them to access onboard ADS data immediately at the scene to confirm driving modes in operation.*

### 6.3. Limitations of the Safety Case for of the ADSE Criminal Liability

The Safety Case<sup>81</sup>;

*"...is a document, or a set of documents, which present a clear, comprehensive and defensible argument for the safety of a given system in a given context". The Ministry of Defence describes a safety case as: a structured argument, supported by a body of evidence that provides a compelling, comprehensible and valid case that a system is safe for a given application in a given operating environment."*

Three key elements are the *claims*, *arguments* and *evidence*.

*Claims* are the safety objectives, and the *arguments* use the *evidence* to justify how the *claims* are met.

The safety case is often considered to be a pre-deployment exercise carried out by a manufacturer or developer in preparation for regulatory review and approval. However, this viewpoint can result in bias towards pre-deployment *evidence* for the safety case and neglect the value of *evidence* gathered through post-deployment in-use monitoring.

---

<sup>81</sup> Consultation Paper 3 page 116 para 7.65

The fact that the safety case is covered in Section 7 entitled "Assessing safety pre-deployment" is an indication of that bias which is carried over to the section conclusion<sup>82</sup>;

*"Assessment methods are still developing, and best practice would suggest simulations, track tests and road tests are all required. These need to be carried out by the developer during the development process, with additional checks by the regulator at the end of the process. The exact combination should be constantly evaluated and will need to be adjusted in line with best practice as it emerges."*

This states that the "assessments" are only to be carried out "during" and "at the end" of the development process. While the assessments themselves are limited to pre-deployment V&V techniques of *simulation, track tests and road tests*.

However, the preceding paragraph recognises that;

*...safety is not a one-off assessment but an ongoing process, in which manufacturers and assessors are continually learning from experience.*

Post-deployment in-use monitoring is the also the only way to gather real-world *evidence* that support the *arguments* that the safety case's *claims* – as such it is essential for *continually learning from experience*.

*It is suggested that the safety assurance regulator be given the powers to implement an ongoing process for continual monitoring and learning of ADS performance that provides evidence that the arguments within the safety case meet the safety claims in 'real-life'.*

#### 6.3.1. New offences based upon the Safety Case

Following on from the above gathering '*real-life*' *evidence* is a fundamental requirement for the new offences based around the safety case.

As described in the consultation paper<sup>83</sup>;

*...the regulatory system depends crucially on the ADSE's safety case, the system would be vulnerable to dishonest behaviour. It would be extremely concerning if an ADSE:*

- 1. misreported test results;*
- 2. suppressed poor test results;*
- 3. claimed to have carried out tests it did not carry out;*
- 4. installed defeat devices, so that the system responded differently in tests than in real life; or*

---

<sup>82</sup> Consultation Paper 3 – Page 122 para 7.98

<sup>83</sup> Summary Consultation Paper 3 page 42 para 7.2

5. *obtained confidential information about test scenarios, and then gamed the system by training only for the test and not for real life.*

Both (4) and (5) make reference to differences between pre-deployment tests and 'real life' performance. So, two in five of the extremely concerning dishonest behaviours related to an ADSE safety case require 'real-life' comparison.

Consider the situation posed by The Molly Problem in which an unoccupied self-driving vehicle (NUIC) is involved in a collision with a child pedestrian, without the presence of other witnesses.

The consultation paper cites a related example to safety cases and AV standard, such as UL4600;

*...if the manufacturer of an AV were to claim that their AV "will not hit pedestrians", this must be supported by arguments such as "the AV will detect all pedestrians" or the "AV will stop or avoid detected pedestrians". This would then need to be backed by evidence, such as detection tests performed on the AV<sup>84</sup>. The aim is get the manufacturer to explain the specifics of their claims so that an independent assessor can analyse whether the product is safe.*

In the case of The Molly Problem what happens after the pedestrian collision during the investigation?

It's clear that a pedestrian collision has occurred. Yet the ADSE produced a safety case with a "claim" that it will not hit pedestrians supported by "arguments" based on pedestrian detection and supported by "evidence" from testing, all of which was analysed by an independent assessor that validated the product as safe for deployment.

Did the ADS perform in 'real-life' as claimed by the ADSE in the safety case? It's impossible to answer this question without access to 'real-life' evidence.

The consultation suggests that "the safety case is central"<sup>85</sup> to the regulatory framework. However, this is not sufficient to assure safety or to determine if offences have been committed by the ADSE.

*It is suggested that within the regulatory framework the "continual 'real-life' assurance of the approved safety case" should be seen as central, not simply pre-deployment approval of the safety case.*

---

<sup>84</sup> Consultation Paper 3 page 119 with reference to  
[https://users.ece.cmu.edu/~koopman/talks/191010\\_UL4600\\_TECH\\_Webinar.pdf](https://users.ece.cmu.edu/~koopman/talks/191010_UL4600_TECH_Webinar.pdf).

<sup>85</sup> Consultation Paper 3 page 122

Without sufficient powers there is a risk that the police, collision investigators and the safety assurance regulator will have insufficient *evidence* to assess even the most critical safety infringements that results in death or serious injury.

*It is suggested, that to assure public safety, the framework grant necessary powers to the safety assurance regulator to ensure the in-use performance of the ADS meets, or exceeds, the performance specified in the ASDE safety case.*

#### 6.3.2. ADSE - aggravated offence for death or serious injury

The consultation states that<sup>86</sup>;

*...regulators will rely heavily on a safety case submitted by the ADSE. It is imperative that this safety case provides an honest account of the tests done and the results obtained. Any misrepresentations or non-disclosures within the safety case would be a breach of the trust placed in the ADSE. And if these failures result in a death or serious injury, we think that would justify serious criminal penalties.*

*...higher penalties should be available when death or serious injury has occurred. This would avoid creating a “radical asymmetry” between the treatment of human drivers and ADSEs.*

*...if the directors of an ADSE were shown to have suppressed safety-critical information, and someone died as a result, the public would demand greater penalties than the maximum two-year sentence available for breach of Regulation 95 of the Human Medicines Regulations 2012.*

The intent behind this criminal liability is right. However, the intent is undermined by the narrow focus on the safety case within the proposed framework.

Consider again The Molly Problem. The severity of the collision was not explicitly stated, so it could equally be a minor, serious or fatal collision. This implies that the public's expressed behavioural expectations for the ADS are applicable to all collision situations, that is to; detect the collision, stop the vehicle, alert other drivers to the hazard and contact the emergency services.

Now consider a situation in which the self-driving vehicle was occupied with a user-in-charge and collision resulted in fatal collision with a child pedestrian for which the ADS failed to stop.

What happens within the currently proposed regulatory framework?

---

<sup>86</sup> Consultation Paper 3 page 241 para 14.15 and 256 para 14.89 & 14.90

- a) The user-in-charge would face criminal charges and imprisonment for failing to report.
- b) The ADSE faces no criminal charges for the fatal collision, the failure to stop or the failure to report.
- c) However, the ADSE could face criminal charges but only if an investigation proves that there were *misrepresentations or non-disclosures* within the *safety case*.

So a bad-actor ADSE, who's ADS detects the collision, identifies the object as a child, deduces the collision is likely to be fatal from the impact velocity, can take the following actions and avoid legal liability;

- they can execute a hit-and-run collision to obstruct the police or witnesses from identifying the vehicle at the scene
- they are under no legal obligation to record time and location data throughout the journey so can obstruct the police from placing the vehicle at the scene of the collision
- they can obstruct the user-in-charge (or Licensed Fleet Operator) from completing their legal duty to report by withholding collision detection information
- they are under no legal obligation to record any evidence related to the collision event so can obstruct any '*real-life*' comparison to the safety case claims.

As long as there are no *misrepresentations or non-disclosures* found within the *safety case* then the ADSE can face no criminal charges for the actions taken in relation to the specific incident hit and run incident with the child.

This appears to perpetuate concerns of "*radical asymmetry*" between the treatment of human drivers and ADSEs. It also prevents the allocation of moral blame, which is a key purpose of criminal penalties.

#### 6.3.3. The significance of *misrepresentations or non-disclosures* related to '*real-life*' safety performance

A regulatory framework in which the "*digital commentary driving is central*" ignores the significance of *misrepresentations or non-disclosures* related to '*real-life*' safety performance.

As an example of *misrepresentations or non-disclosures* related to '*real-life*' we can refer back to the earlier discussion related to safe passing distances to vulnerable road users and the need to introduce active in-use monitoring to capture violations of these leading metrics<sup>87</sup>;

- Do these leading metrics provide a '*real-life*' evidence base for the performance ADS pedestrian detection and avoidance? **Yes.**

---

<sup>87</sup> [See Section 4.5. Leading Metric Example – Safe Passing of Vulnerable Road Users](#)

- Does this '*real-life*' evidence base enable *claims, evidence* and *arguments* made in the safety case to be substantiated or disproved. **Yes.**
- Are the number of pedestrian near-miss violations relevant? **Yes.**
- Is the similarities in circumstances of these near-miss violations and the fatal collision relevant. **Yes.**
- If safety assurance regulator has imposed has faced previous sanctions in the ADSE for VRU near-miss violations is this relevant? **Yes.**
- Is the comparison of '*real-life*' VRU safety performance before and after any sanctions relevant? **Yes.**
- Are the nature of those sanctions relevant? e.g. informal and formal warnings; fines; redress orders; compliance orders; suspension of authorisation; withdrawal of authorisation. **Yes.**
- If the ADSE published a revised safety case, addressing the VRU safety performance, to remove the suspension of authorisation is this relevant? **Yes.**
- Is evidence that this revised safety case leads to the stated improvements '*real-life*' VRU safety performance relevant? **Yes.**
- Would it be appropriate for the safety assurance regulator to impose a period of enhanced '*real-life*' in-use monitoring, for VRU safety performance, as a condition of removing the suspension. **Conceivably, yes.** If this condition was in place at the time of the fatal collision is this relevant? **Yes.**

It seems clear from the above that the "continual '*real-life*' assurance of the approved safety case" is critical to the foundation to a no-blame regulatory framework.

In fact, it's reasonable to expect the public would view *misrepresentations or non-disclosures* related to '*real-life*' performance are more morally objectional than those related only to, the more abstract, safety case.

This public expectation of *disclosure* is already evident in the The Molly Problem responses in which the public expect explainability for '*real-life*' ADS situational awareness, judgement of risk and decision making in regard to appropriate safe actions.

The "continual '*real-life*' assurance of the approved safety case" is also beneficial to resolving the challenges raised in establishing a causal relationship between false information and harm<sup>88</sup>;

*One difficult issue is how to establish a causal relationship between the wrong (the misrepresentation or non-disclosure) and the harm (the serious injury or death). The wrong may be committed many years before the harm occurs and be carried out in a quite different setting.*

---

<sup>88</sup> Consultation Paper 3, page 256, para 14.92 to 14.96



The statement is written with an implicit assumption that the safety case is a static document in which the suppression of a poor test result could remain unknown and undetected for '*many years*'.

The purpose of "continual 'real-life' assurance of the approved safety case" is to ensure that such wrongs can be detected *before the harm occurs* by the use of appropriate leading metrics.

The consultation raises concerns over the breadth of causal relationship and the need to avoid definitions that are "*over-inclusive*" or "*too narrow*".

Again in-use monitoring can provide an *evidence* base in which the strength of causal relationships can be defined.

In the case where an ADS approved as self-driving but the ADSE had suppressed poor night-time test results it was suggested that it would be "*over-inclusive*" to hold as ADSE responsible for this non-disclosure should a death occur in the day-time. The role of in-use monitoring in this scenario is two-fold;

- Firstly, leading metrics should identify the discrepancy between the safety case and the '*real-life*' poor night-time performance before harm occurs enabling the safety assurance regulator to act accordingly
- Secondly, leading metrics provides '*real-life*' *evidence* which may identify a link between poor night-time performance and specific instances of poor day-time performance e.g. driving in a street lit tunnel, multi-story car park or more general issues with traffic movement prediction.

At the other end of the spectrum it was suggested in the consultation that "*it would be too narrow to require that the suppressed test result deal with exactly the same scenario as the one that caused the death.*" So, for example, the aggravated offence could be applied if a vehicle struck a tree at night and it was discovered that results had been suppressed for poor VRU safety performance at night (the implication being that there is more general problem with night driving performance).

Once again, in-use monitoring provide highly valuable *evidence* both in identifying the poor night-time VRU safety performance and establishing a broader link to more general problem with night driving.

Waiting for serious injury or death to occur before assuring '*real-life*' performance matches the safety case is not the best foundation culture safety culture built around continual learning.

Finally, reference was made in the consultation to the causation approach taken in the legislation used within the US aviation industry<sup>89</sup>;

*If, by reason of the failure of the part to operate as represented, the part to which the offense is related is the proximate cause of a malfunction or failure that results in the death of any person.*

In the context of automated driving the “part” may relate to “software”, such as a module, function or algorithm.

The need to identify if a “part” failed to “operate as represented” is exactly the purpose of “continual ‘real-life’ assurance of the approved safety case”.

The Molly Problem results highlight the public expectation that the “part” and “operation” can be identified in a granular manner with roughly three levels covered earlier in this paper;

1. *Reaction explainability* - Failure in the ADS module responsible for selection and execution of the appropriate mitigating action.
2. *Decision explainability* - Failure in the ADS module responsible for hazard awareness and risk evaluation of the road traffic situation (which may additionally result in mitigating action failure)
3. *Perception explainability* - Failure in the ADS module responsible for accurate situational and environmental awareness (which may additionally result in hazard awareness failure and mitigating action failure)

*It is suggested that the proposed regulatory framework places the “continual ‘real-life’ assurance of the approved safety case” as the central foundation and that criminal liabilities for an ADSE are extended to include misrepresentations or non-disclosures in ‘real-life’ evidence.*

#### 6.4. Limitations in scaling investigations for both collisions and near-misses

Consultation question 22 covers the provisional proposal that<sup>90</sup>;

*...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 traffic infractions, including those subject to penalty charge notices;*

---

<sup>89</sup> Consultation Paper 3, page 256 para 14.94

<sup>90</sup> Consultation Paper 3, page 179

*It is suggested that statutory scheme does include these provisions. In doing so the scheme must address the issue of scaling the number and types of investigations to include collisions, near-misses and other, safety or non-safety related, infractions.*

Consultation Question 25 covers the provisional proposal that<sup>91</sup>;

*...a specialist collision investigation unit should be established:*

- (1) to analyse data on collisions involving automated vehicles;*
- (2) to investigate the most serious, complex or high-profile collisions; and*
- (3) to make recommendations to improve safety without allocating blame.*

The joint workshop poll respondents agreed;

- 90.4% *to analyse data on collisions involving automated vehicles*
- 80.4% *to investigate the most serious, complex or high-profile collisions*
- 89.1% *to make recommendations to improve safety without allocating blame*

#### 6.4.1. Who investigates collisions and safety relevant infractions?

The consultation cites two reasons why specialist collision investigation could be desirable;

1. *Firstly, automated vehicles will be complex systems with many parties involved.*
2. *The second reason is that specialist collision investigation could help monitor the safety standard for AVs.*

If the role of the specialist collision investigation unit is only to “analyse data on collisions” and only “investigate most serious, complex or high-profile collisions”, then which organisation is responsible for the analysis and investigation of “safety related traffic infractions such as careless or dangerous driving” and reported “near-misses”?

Do these additional responsibilities also fall within the scope of the specialist collision investigation unit as part of the function to “monitor the safety standards for AVs”?

*It is suggested that the entity responsible for the analysis and investigation of “safety related traffic infractions such as careless or dangerous driving”, reported “near-misses” and “other traffic infractions” is identified within the final regulatory framework.*

#### 6.4.2. Moving beyond criminal liability to a culture of safety

The NTSB investigation into the Elaine Herzberg fatal collision on the 18<sup>th</sup> March 2018, involving an Uber ATG automated test vehicle, was launched on the 19<sup>th</sup> March 2018<sup>92</sup>. The

---

<sup>91</sup> Consultation Paper 3 page 189 para 11.69

<sup>92</sup> <https://www.nts.gov/investigations/Pages/HWY18FH010.aspx>

NTSB filed a provisional report on the 24<sup>th</sup> May 2018 (2 months later) and a final report on 19<sup>th</sup> November 2019 (20 months later).

The final report covered vehicle factors, highway factors, operations factors, human performance factors. The investigation included submissions of evidence from the ADS developer (Uber ATG) and the vehicle manufacturer (Volvo), as well as from independent safety research (Thatcham Research).

The investigation went beyond the requirements for criminal prosecution with the purpose of generating safety culture learnings, shared in the form of formal recommendations.

There is certainly a role for these detailed collision investigations, and this seems to be the role envisaged for the proposed specialist collision investigation unit who's proposed scope is limited to *"the most serious, complex or high-profile collisions"*.

Collisions are a lagging metric, which 92.6% of workshop poll respondents agreed should be collected by the safety assurance regulator. However, leading metrics, which can be captured before harm, are also a foundation to a culture of safety with 83.3% of respondents in agreement should be collected.

It should be noted while leading metrics are often discussed in terms of negative outcomes, such as traffic infractions, they can also be used to collect positive outcomes e.g. how many cyclists or how many traffic light intersections have been passed safely.

*It is suggested that the proposed regulatory framework ensures a balance in the safety analysis between lagging metrics (collisions) and leading metrics (both negative and positive outcomes).*

#### 6.4.3. The need to address the scale, rate and duration of investigations

While the Elaine Herzberg NTSB investigation was thorough, in the age of software and digital technology it would be considered slow. In fact, at 20 months the investigation is 20-40x slower than the rate over the air (OTA) software updates in some currently deployed SAE Level 2 vehicles.

For self-driving vehicles, one might expect the same, or higher, rate of software updates. This might be especially true in early deployment phases and expansion of the operational design domains.

In order for the no-blame safety culture to function effectively, from the perspective of public safety assurance and innovation, as well as commercially for developers and service providers, then the issues of scale, rate and duration of safety critical investigations needs to be addressed.

Access to data will be critical to achieving these requirements and the overall objective of safety improvement through continual learning.

The quality of the *evidence* that can be extracted from the data is key to these efficiencies e.g. higher levels of *explainability* should enable more investigations to be completed in a shorter time and a higher capacity to process more investigations in parallel (critical for multiple vehicle events which likely to be *serious, complex and high-profile*).

*It is suggested that the safety assurance regulator should be granted the powers to access the data that provides evidence to meet the explainability levels required for safety assurance to ensure cost, time and human resource efficient processing of incident analysis and investigations.*

#### 6.4.4. The value of 'real-life' evidence to incident investigation efficiency and proportional use of suspension or withdrawal of authorisation

Efficient investigations are the foundation to a safe, innovative and efficient CAV mobility sector.

For example, consider a future scenario of a collision investigation, such as the Elaine Herzberg night-time fatality, which involves one of a fleet of self-driving vehicles operated by Uber in London. As of July 2020, there are 45,000 licensed drivers and 3.5 million riders who depend upon Uber in the city<sup>93</sup>. What happens after the *during* the collision investigation?

The safety assurance regulator should only be concerned with public safety, so immediate suspension of authorisation to operate would be reasonable to impose.

However, it's also clear that suspension will have significant financial impact on the License Fleet Operator as well as on the mobility options available to 3.5 million riders. If the suspension was to last for the 20-month duration it took for the NTSB investigation such a suspension might be infeasible commercially.

*It is suggested that the proposed regulatory framework clearly addresses the potential conflict of interest between the "safety first" role of the safety assurance regulator in imposing authorisation suspensions and resulting pressures created by the needs to restart commercial operations as well as resume service for dependent mobility users. It is suggested that consideration be given to whether the safety assurance regulator is best placed to manage this judgement or if additional bodies are required?*

---

<sup>93</sup> <https://www.forbes.com/sites/daviddawkins/2020/07/30/uber-vs-london--the-courtroom-battle-the-world-is-watching>

Ultimately, the decision to both suspend, and later resume, authorised operations is a judgement based upon public risk and public value.

'Real-life' data is essential in making this risk-based judgement using evidence to determine the "*circumstance and situation*" in which collision occurred;

Should operations of the entire fleet be suspended?

Or does the data provide evidence that;

- the collision was not preventable, meaning appropriate mitigating action was taken at the appropriate time and executed correctly. The hazard awareness and situational awareness software modules operated as represented and as required?
  - therefore, fleet operations can continue but under enhanced in-use monitoring for a defined period.
- the collision was related only to VRU detection and avoidance module?
  - therefore, operations could be reasonably suspended in ODDs with high VRU densities until the issue is resolved
- the collision was related only to night-time performance of the VRU detection and avoidance module?
  - therefore, operations could be reasonably suspended at night in ODDs with high VRU densities until the issue is resolved
- the collision was related to a specific ADS software release;
  - therefore, operations could be reasonably suspended for the vehicles operating this software version

It should be clear from the above that data captured using the "*perception, decision, reaction and outcome explainability model*<sup>94</sup>" about the collision can provide *evidence* for the judgement of public risk and the levels of sanctions imposed. While also managing the level of public service disruption and resulting financial impact on the ADSE or Licensed Fleet Operator in a reasonable manner.

In this regard, there is a clear financial incentive for the ADSE, manufacturer and Licensed Fleet Operator to provide access to the necessary data for explainability in relation to collisions and safety relevant infractions. Whilst also engaging in continual in-use monitoring which provides further historical '*real-life*' evidence to support *claims* of safe behaviour in related circumstances and situations.

*It is suggested that the regulatory framework continues to develop its proposal for in-use monitoring and that further consideration is given to the most appropriate incentive structure to create positive and active engagement within the no-blame safety culture. It is suggested that for suspension of authorisation to be proportional to risk it is in the*

---

<sup>94</sup> [Perception, decision, reaction and outcome explainability model](#)

*interests of the ADSE to enable access to 'real-life' data that provides evidence and explainability of performance of perception, decision, reaction and outcome.*

#### 6.4.5. "Data-dumps" are not just as safety case issue

While the discussion above highlights the incentives for an ADSE to engage in the sharing of high-quality information regarding safety performance, this may not be guaranteed.

The consultation paper raises issue of "data-dumps" in relation to the safe case<sup>95</sup>;

*Correcting the imbalance of information held by the ADSE and the safety assurance scheme will be important to the efficiency of the scheme. However, we also want to deter ADSEs submitting "data dumps" that make it difficult to properly assess the safety case and to identify non-disclosures and misrepresentations.*

Consultation Question 46 seeks;

*...views on whether an ADSE should be under a duty to present information in a clear and accessible form, in which safety-critical information is indexed and signposted.*

*It is suggested that this requirement is enforced for the safety case but that the same principles are also extended to include the safety critical data required for 'real-life' assurance of the safety case as well as investigation of collisions and safety relevant infractions.*

#### 6.4.6. How fast is safe enough?

In the context of automated vehicles, the question "how safe is safe enough?" is often asked. In fact, the consultation paper devotes a whole chapter to the topic<sup>96</sup>.

However, in the context of investigations, perhaps the more appropriate question is "how fast is safe enough?". If an investigation is;

- too slow lessons may not be learned quickly enough to avoid further exposure to harm.
- too fast may not enable a detailed enough investigation to identify the safety critical issues required to inform the learning.
- too slow and mobility services may be disrupted for extended periods of time.
- too fast and services might be resumed too early while safety critical flaws remain present.

---

<sup>95</sup> Consultation Paper 3, page 258, para 14.101 to 14.105

<sup>96</sup> Consultation Paper 3, Chapter 5 How safe is safe enough? Page 61



It's a complex balance.

*It is suggested that this complex balance could be addressed by a continued and ongoing safety dialogue between the safety assurance regulatory, the ADSEs, manufacturers and licensed fleet operators. Regular aggregate reporting of in-use leading metrics and ad hoc safety bulletins would be one valuable approach – which again demonstrates the value of in-use monitoring.*

However, the safety assurance regulatory should performance targets for “*how fast is fast enough?*”.

Similar to hospital waiting times<sup>97</sup> these investigation duration targets could be a way to measure how the safety assurance regulator and/or the specialist collision investigation unit are performing.

For example, after a fatal collision the safety assurance regulator may be set targets to;

- a) within 24hrs launch an investigation
- b) within 48hrs impose temporary suspension on authorisation within
- c) within 120 hours publish provisional findings on the ADS performance including confirmation of the situational awareness, hazard assessment and mitigating action

In order to achieve those targets the safety assurance regulator may use its power to impose related mandatory targets on the ADSE for the sharing of all relevant safety critical information.

*It is suggested that the regulatory framework considers the need for performance targets on the safety assurance regulator and any related powers they might require to impose obligations on other actors in order that those targets can be met.*

*It is suggested that, where feasible, safety focused investigations should be carried out in parallel to any criminal proceedings carried out by the police.*

#### 6.4.1. Police - immediate on the spot access to ADS in-use performance data

For a number of reasons, the police may require immediate and on the spot access to data related to ADS in-use performance.

---

<sup>97</sup> <https://www.nhs.uk/nhs-services/hospitals/guide-to-nhs-waiting-times-in-england/>

This may be in the case of a traffic infraction such as speeding in which the driver/UIC claims the ADS was engaged, or a drunk occupant who claims the vehicle operating in a NUIC mode.

Feedback received in an earlier consultation suggested<sup>98</sup>;

*...there was widespread agreement that if an offence appears to have been caused by an ADS, the police should refer the matter to a regulatory authority for investigation.*

The time it takes the police to refer the matter will impact the efficiency in the operations of the safety assurance regulators and specialist collision investigation unit. Ensuring the police have access to the necessary data required to refer the matter is a key part to the way the regulatory framework will function.

Establishing the cause at the scene of motorway collision may involve multiple lane closures and significant disruption to traffic. The speed at which sufficient evidence can be gathered is therefore similar to the “*how fast is safe enough?*” question.

Having direct, immediate and on the spot access to data related to ADS in-use performance would be a crucial source of evidence and may significantly speed up the onsite investigation process.

One possible target in this regard could be that the police are granted powers to access to a single diagnostics port on the vehicle which enables access to all data required to reconstruct the road traffic situation that led to the collision, the hazard evaluation completed by the ADS, and a record of the mitigating actions taken by the ADS.

This is exactly the same “*perception, decision, reaction and outcome explainability*” data suggested for the in-use assurance of the ADS safety case, so should be readily available if these suggestions are adopted.

*It is suggested that further consultation with the police and collision investigators should be conducted to confirm the minimum set of data required for the analysis and the requirement for the most appropriate analysis tools. It is suggested that a 3D visualisation tool in which the investigator can playback the collision, view the collision from different angles and have clear onscreen notifications of significant events related to the safety performance of the system would be very powerful.*

It's conceivable that in a multiple CAV collision data from each involved vehicle, including “witness vehicles”, could be downloaded and viewed within this single 3D visualisation tool for the purposes of comparison.

---

<sup>98</sup> Summary Consultation Paper 3 page 27 para 4.34

Additionally, if the collision occurs upon in the vicinity of a smart intersection or roadway then this infrastructure data should also be accessible within the 3D visualisation tool.

Access to this same data and tools should also be available for near-miss and other safety related traffic infractions.

*It is suggested that the proposed regulatory framework considers the significant impact direct, immediate on the spot access to ADS safety critical performance data could have enhancing investigation quality and efficiency.*

In addition, it is also conceivable that ADS safety critical performance data could be accessed remotely from an investigation centre with cloud hosted 3D visualisation tools used for event reconstruction and analysis.

*It is suggested that the necessary powers are granted to the necessary authorities to enable this cloud-centric form of investigation which has the potential to radically impact the scale, rate and duration required for the future of safe CAV mobility.*

## 6.5. International safety assurance

While the regulatory framework is specifically focused upon domestic matters, the self-driving industry, like mobility itself, is global and creates a need for international harmonisation – the impact of which is explored below.

### 6.5.1. Leading metrics and collisions investigations at an international scale

If a fatal collision occurs in another country, leading to suspension of authorisation within that country or criminal investigation, then what steps should the domestic safety assurance regulator take to ensure public safety?

In the opposite scenario, where a fatal collision and suspension occurs domestically, should the safety assurance regulator share information with other countries to assist in their public safety assurance?

It would seem logical that there should be some level of international cooperation between the domestic safety assurance regulators.

The OECD recognised a similar need for global collaboration in relation to product recalls which led them to create<sup>99</sup>;

---

<sup>99</sup> <https://www.oecd.org/about/impact/global-recalls-portal.htm>

*...a global platform on recalls {which} has enabled domestic authorities to take action to protect people from possible harm, while also saving time, money and resources.*

As of March 2021, the OECD Global Portal on Product Recalls, list 4838 automotive recalls which is the second highest amount with toys at 5291 and 2927.

In the digital world of automated driving software and data communications it may be more appropriate to align the international safety cooperation efforts with those already in place for cybercrime.

For example, the Budapest Convention on Cybercrime, to which the UK is one of 67 signatories, states signatories<sup>100</sup>;

- *Convinced of the need to pursue, as a matter of priority, a common criminal policy aimed at the protection of society against cybercrime, inter alia, by adopting appropriate legislation and fostering **international co-operation***
- *Believing that an effective fight against cybercrime **requires increased, rapid and well-functioning international co-operation** in criminal matters;*
- *the adoption of powers sufficient for effectively combating such criminal offences, by facilitating their detection, investigation and prosecution at both the domestic and international levels and by providing arrangements **for fast and reliable international co-operation**;*
- *Article 25 – General principles relating to mutual assistance*
  - *The Parties shall **afford one another mutual assistance to the widest extent possible for the purpose of investigations** or proceedings concerning criminal offences related to computer systems and data, or for the collection of evidence in electronic form of a criminal offence.*
- *Article 26 – Spontaneous information*
  - *Party may, within the limits of its domestic law and without prior request, **forward to another Party information obtained within the framework of its own investigations when it considers that the disclosure of such information might assist the receiving Party in initiating or carrying out investigations** or proceedings concerning criminal offences established in accordance with this Convention or might lead to a request for co-operation by that Party under this chapter.*

While the Budapest Convention provides a template for international cooperation on self-driving safety assurance, a number of the cybercrime offences are directly relevant to automated driving systems, including; Illegal access, Illegal interception, data interference,

---

<sup>100</sup> Convention on Cybercrime Budapest, 23.XI.2001 <https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/0900001680081561>

system interference, misuse of devices, computer-related forgery and computer-related fraud.

#### 6.5.2. Great Britain, Northern Ireland, Ireland & Europe Cooperation

This issue of cross-border collaboration, in relation to self-driving safety assurance, becomes particularly important even within the United Kingdom. The consultation paper explicitly states that the remit of the Law Commissions review<sup>101</sup>;

*...does not extend to Northern Ireland: we can only make recommendations in respect of England, Wales and Scotland. Northern Ireland is subject to different constraints under the Ireland-Northern Ireland protocol and will require a different system of regulation.*

*It is suggested that, while it may not be possible to comment on the specific regulations in Northern Ireland, the framework should provide guidance on the proposed approach to cross-border CAV safety assurance.*

*It is suggested that developing a cross border CAV safety assurance strategy between Great Britain, Northern Ireland, Ireland and therefore the European Union would be a great way to extend and continue the leadership demonstrated by the Law Commission of England and Wales and the Scottish Law Commission during this public consultation.*

#### 6.5.3. 1968 Convention on Road Traffic - Article 34 bis & global fragmentation in self-driving regulation

According to an explanatory memorandum prepared by The Department for Transport in 2019<sup>102</sup>;

*As part of the government's preparations for leaving the EU, the UK has ratified the 1968 Convention, which is a UN international treaty designed to facilitate **international road traffic and increase road safety by establishing uniform traffic rules**. The UK originally signed the 1968 Convention in November 1968, but only ratified it in 2018 because of the decision to leave the EU and the need to ensure the continued recognition of UK driving licences.*

During the 81<sup>st</sup> Session of the Global Forum for Road Traffic Safety contracting parties agreed to adopt amendments to "ensure greater levels of legal certainty due to increasing vehicle automation"<sup>103</sup>.

---

<sup>101</sup> Consultation Paper 3, page 6, para 1.41

<sup>102</sup> [https://www.legislation.gov.uk/ukxi/2019/563/pdfs/ukxiem\\_20190563\\_en.pdf](https://www.legislation.gov.uk/ukxi/2019/563/pdfs/ukxiem_20190563_en.pdf)

<sup>103</sup> ECE/TRANS/WP.1/2020/1/Rev.1 <https://undocs.org/ECE/TRANS/WP.1/2020/1/REV.1>

Specifically, these were two new definitions under Article 1 and a new Article 34bis. For completeness these are shown in full below;

#### *Article 1 - Definitions*

*(ab) "Automated driving system" refers to a vehicle system that uses both hardware and software to exercise dynamic control of a vehicle on a sustained basis.*

*(ac) "Dynamic control" refers to carrying out all the real-time operational and tactical functions required to move the vehicle. This includes controlling the vehicle's lateral and longitudinal motion, monitoring the road, responding to events in the road traffic, and planning and signalling for manoeuvres."*

#### *Article 34bis - Automated driving*

*The requirement that every moving vehicle or combination of vehicles shall have a driver is deemed to be satisfied while the vehicle is using an automated driving system which complies with:*

*(a) domestic technical regulations, and any applicable international legal instrument, concerning wheeled vehicles, equipment and parts which can be fitted and/or be used on wheeled vehicles, and*

*(b) domestic legislation governing operation.*

*The effect of this article is limited to the territory of the contracting party where the relevant domestic technical regulations and legislation governing operation apply.*

The two-step process of (1) type approval and (2) categorisation, proposed for the GB regulatory framework<sup>104</sup> is aligned with the Article 34bis requirement for automated driving systems to comply with (a) technical regulations (domestic or applicable international) and (b) operation legislation (domestic only).

The consultation describes Step 2 as follows;

*Step 2 is categorisation. While national or international type approval is largely concerned with technical issues, the **categorisation** decision is wider and more holistic. Rather than focus on a single system, it considers the **safety of the vehicle as a whole**, as embedded within safe processes to update, maintain and repair that vehicle and inform users about their responsibilities.*

As described in the Article 34bis explanatory memorandum'

---

<sup>104</sup> Consultation Paper 3 page 125-126 para 8.5-8.8

*...the automated driving system must comply with “domestic legislation governing operation” ...Under this provision a party could also impose additional requirements for the operation of automated driving systems.*

The domestic powers contained in Article 34bis (a) & (b) will enable contracting parties to define their own technical and operation regulations and legislation. The consultation paper suggested domestic regulation is required as<sup>105</sup>;

*“...there is no international agreement on what constitutes self-driving.”*

The domestic powers under Article 34bis means that contracting parties of the 1968 Convention on Road Traffic will never have to reach international agreement which will undoubtably cause in fragmentation in the global regulation of self-driving mobility; as each state creates their own definition for “self-driving”, “safely driving itself” and “circumstances or situations”.

***It is suggested that, in addressing domestic needs, the proposed regulatory framework should also provide a global benchmark for self-driving safety and a foundation for reaching international agreement on automated driving regulation.***

#### *6.5.3.1. 1968 Convention on Road Traffic – importance for Article 34 bis, ALKS & safety assurance*

While the consultation paper suggests “there is no international agreement on what constitutes self-driving” there is already international agreement on safe driver behaviour. The 1968 Convention on Road Traffic sets behavioural expectations for drivers and road users.

If “self-driving” means “safely driving itself” under certain “circumstances or situations” to the level of a “competent and careful human driver” then it’s clear that the self-driving system (a *non-human driver*) that replaces the human must also meet the definition of a “driver” within the convention.

In September 2020 FG-AI4AD submitted, to the UNECE Global Forum on Road Traffic Safety (WP.1), an analysis of the Convention highlighting 32 Article and 132 Clauses that are relevant for both human and *non-human drivers*<sup>106</sup>.

The references to expected driver behaviours included;

#### *Article 7 – General Rules*

---

<sup>105</sup> Consultation Paper 3 page 135 para 8.53

<sup>106</sup> <https://unece.org/fileadmin/DAM/trans/doc/2020/wp1/ECE-TRANS-WP1-SEPT-2020-Informal-6e.pdf>



- Road-users shall **avoid any behaviour likely to endanger or obstruct traffic**, to endanger persons, or to cause damage to public or private property.
- Drivers shall show **extra care in relation to the most vulnerable road users**, such as pedestrians and cyclists and in particular children, elderly persons and the disabled.

#### Article 8 – Drivers

- Every driver shall possess the necessary **physical and mental ability** and be in **a fit physical and mental condition** to drive.
- Every driver of a power-driven vehicle shall possess the **knowledge and skill necessary for driving the vehicle**
- Every driver shall at all times be able to **control his vehicle**

#### Article 17 – Slowing Down

- No driver of a vehicle shall brake abruptly unless it is necessary to do so for safety reasons.
- Every driver intending to slow down to an appreciable extent shall, except where his slowing down is in response to an imminent danger, first **make sure that he can do so without danger or undue inconvenience to other drivers**.

#### Article 21 - Behaviour of drivers towards pedestrians.

- Every driver shall **avoid behaviour likely to endanger pedestrians**.

#### Article 25 – Motorways and similar roads

- Drivers shall be **forbidden**:
  - o To have their vehicles standing or parked elsewhere than at marked parking sites; if a vehicle is compelled to stop, its driver shall endeavour to move it off the carriageway and also off the flush verge and, if he is unable to do so, immediately signal the presence of the vehicle at a distance so as to warn approaching drivers in time;

#### Article 31 - Behaviour in case of accident

- every driver or other road-user involved in a traffic accident shall:
  - (a) **Stop as soon as he can do so without causing an additional danger to traffic;**
  - (b) Endeavour to **ensure traffic safety at the site** of the accident and, if a person has been killed or seriously injured in the accident, to prevent, in so far as such action does not affect traffic safety, any change in conditions at the site, including the disappearance of traces which might be useful for determining responsibilities;
  - (c) If so requested by other persons involved in the accident, identify himself to them;
  - (d) If a person has been injured or killed in the accident, **notify the police** and remain on the scene of the accident or return to it and wait there until the

*arrival of the police, unless he has been authorized by the police to leave or has to assist the injured or to receive attention himself.*

It should be immediately clear that these articles (just 6 from 32) relate to the circumstances situations of the The Molly Problem and ALKS.

In fact, there is an additional article, not originally referenced in the FG-AI4AD analysis, that relates to ALKS sensor hardware;

#### *Annex 5 Technical conditions concerning Motor Vehicles and Trailers*

##### *47. Driving (rear-view) mirror*

- *Every motor vehicle shall be equipped with one or more driving (rear-view) mirrors; the number, dimensions and arrangement of these mirrors shall be such as **to enable the driver to see the traffic to the rear of his vehicle.***

The UNECE regulations for ALKS were developed within the World Forum for Harmonization of Vehicle Regulations (WP.29) and no references were made to these expectations within the Convention on Road Traffic administered by the Global Forum for Road Traffic Safety (WP.1).

It seems clear that ALKS would fall short of the expectations for a *driver*. In particular it contravenes the following aspects of the Convention;

- Annex 5.41 - enable the driver to see the traffic to his rear of his vehicle
  - ALKS has zero rearward perception
- Article 31 - behaviour in the case of accident
  - ALKS has no collision detection system
  - ALKS has no interface requirement to eCall
- Article 25 - forbidden from stopping on a motorway, if compelled to move off the carriageway, if he is unable to do so immediately signal the presence of the vehicle at a distance so as to warn approaching drivers in time
  - ALKS considers stopping in lane a safe behaviour as part of a minimal risk manoeuvre (MRM)
  - ALKS will activate the hazards lights of the vehicles, however, is this sufficient to “warn approaching drivers in time” where normally a warning triangle would be required at a minimum distance of 45 metres<sup>107</sup>
- Article 17 - Slowing Down must not do so abruptly unless for a safety reason, if intending to slow down must make sure that he can do so without danger or undue inconvenience to other drivers
  - The ALKS regulation states “A transition demand shall not endanger the safety of the vehicle occupants or other road users.”

---

<sup>107</sup> Highway Code Breakdowns (274) <https://www.highwaycodeuk.co.uk/breakdowns-and-incidents.html>

- ALKS can still execute an MRM without a safety reason at the end of a transition demand, do so in an unsafe manner and the user-in-charge be held responsible
- ALKS cannot make sure the MRM is executed “without danger or undue inconvenience to other drivers” because it has no rearward perception capability
  - This was even acknowledged in the consultation paper<sup>108</sup> “*we would hope that transition demands would not be necessary: the vehicle would be able to achieve a minimal risk condition in all circumstances without endangering road safety or obstructing traffic flow.*”
- Article 8 - knowledge and skill necessary for driving, possess necessary *physical and mental ability* as well as *fit physical and mental condition, control his vehicle*.
  - ALKS requires a Driver Availability Recognition System to be active at all times because;
    - It does not possess the necessary skills and knowledge to meet Article 17, 21, 31 (and others). In particular it cannot execute a safe MRM on a motorway and requires a transition demand to a “*the driver*”.
    - Fit physical and mental condition (of sensors, computers and software) cannot be guaranteed so requires transition demand to “*the driver*”
    - Physical and mental ability (of actuators, sensors, computers and software) cannot be guaranteed for an MRM so human driver remains responsible (under the proposed framework)
      - *It's unclear how ALKS meets the fitness to drive requirements for humans, for which safe driving requires involvement of: vision, visuospatial perception, attention and concentration, memory, insight and understanding, judgement, adaptive strategies, good reaction time, planning and organisation, ability to self-monitor, sensation, muscle power and control, coordination*<sup>109</sup>.
- Article 7 - avoid any behaviour likely to endanger or obstruct traffic
  - ALKS will obstruct traffic during a 10 second transition demand by maintaining a 37mph speed while surrounding traffic reaches 69mph before then
  - ALKS regulations permit execution of an MRM likely to endanger and obstruct traffic by stopping in lane from a speed of 37mph, while surrounding traffic is moving at 69mph

<sup>108</sup> Consultation Paper 3 page 199 para 12.36

<sup>109</sup> [Assessing fitness to drive – a guide for medical professionals](#)

If ALKS contravenes the 1968 Convention on Road Traffic in relation to the expectations of all *drivers* on the international roads of contracting parties then does it really meet the requirements of a “*driver*”, let alone a “*competent and careful driver*” as it purports to be?

A critical point in relation to Article 34bis is the use of a “deeming” provision for the driver requirement of Article 8.1 of the Convention, which is described in the explanatory memorandum<sup>110</sup>;

*....“deeming” the driver requirement, i.e. **that every moving vehicle shall have a driver**, to be satisfied under the conditions explained below. This approach respects the Convention’s substantive requirements.*

The draft FG-AI4AD Technical Report on the Public safety benefits of continual monitoring makes the following observation in relation in relation to Article 34 bis<sup>111</sup>;

*Article 34bis states that the requirement of art. 8(1) is “deemed to be satisfied” when an automated driving system is being used. This can be interpreted in two ways:*

- 1. The automated driving system should be considered to be the driver of the vehicle; or*
- 2. The requirement of art. 8(1) is satisfied, but the automated driving system is not the driver of the vehicle.*

*Option 1 appears to be an incorrect interpretation of the phrase “deemed to be satisfied” as this contradicts the definition of “driver” given in art. 1 (v): ““Driver” means any person who drives a motor vehicle or other vehicle (including a cycle), (...)”. Other wording would be more suitable for interpretation-option 1.*

*Therefore, option 2 seems to be the correct interpretation of art. 34bis.*

*Given art. 34bis, all traffic rules from the 1968 Convention directed at a driver (such as art. 11 on overtaking) would not apply to automated driving systems, as a vehicle with engaged automated driving system does not have a driver.*

The fact that the ALKS regulations, excluded and contravene of the driver expectations detailed in the Convention also suggests that the automated driving system is not a “driver” (option 2). Therefore, that ALKS, and all automated driving systems, are not subject to the same driving conventions as all other drivers. Instead, ALKS/ADS safety performance is defined in isolation and solely by “*domestic technical regulations, and any applicable international legal instrument*”.

---

<sup>110</sup> <https://undocs.org/ECE/TRANS/WP.1/2020/1/REV.1>

<sup>111</sup> FGAI4AD-02 Automated driving safety data protocol – Public safety benefits of continual monitoring

This opens an important and fundamental question which was not addressed in the proposed regulatory framework;

- As a signatory to the 1968 Convention on Road Traffic do the driver behavioural expectations and driver related technical conditions apply to the proposed regulatory framework for “self-driving” vehicle to “safely drive itself” using an “automated driving system”?

The only reference to the Convention within the consultation paper appears to Article 34bis itself which is limited to discussion of the two-step authorisation process.

The consultation paper makes no reference to the 32 articles and 132 clauses of the Convention identified by FG-AI4AD as relevant to the definition of “self-driving” and “safely drive itself”, and because of which, it makes no reference to whether these should be addressed within type-approval or categorisation step.

The benefit of the two-step process in the proposed regulatory framework is that if it is felt that these requirements are not adequately dealt within the type-approval step, through domestic or international technical regulations, then they can be addressed in the categorisation step.

*It is suggested, as essential, that the proposed regulatory framework must make it explicit as to whether the articles and clauses of the 1968 Convention on Road Traffic relevant to drivers are applicable to the definitions of “self-driving”, “safely driving” itself and “circumstances or situations” and if they are not, provide a reasoned explanation as to why they should not apply.*

In the workshop poll participants were asked to respond to the following question;

*Do you think the in-use regulatory authority should be separate from type approval authorities (as is already the case? In the UK with VCA and DVSA respectively?*

Which relates closely to Consultation Question 20<sup>112</sup>;

*Should the authority administering the scheme to assure safety while automated vehicles are in use be kept separate from type approval authorities (as is already the case)? Alternatively, should both functions be combined in a single body?*

In the workshop poll responses 70.5% believed that the safety assurance regulator should be from the type-approval authority.

---

<sup>112</sup> Consultation Paper 3 page 173 para 10.100

*It is suggested that the type-approval authority remains closely aligned to the activities of the World Forum for Harmonization of Vehicle Regulations (WP.29) and focus on the technical approval of vehicle systems (including self-driving). Additionally, the proposed safety assurance regulator should align to the activities of the Global Forum for Road Traffic Safety (WP.1) and focus on safe, equitable mobility.*

#### 6.5.4. Safety Assurance Regulator – extended responsibility for CAV Ethics

In June 2020 an independent expert group, advising the European Commission issued a report entitled<sup>113</sup>;

*Ethics of Connected and Automated Vehicles  
Recommendations on road safety, privacy, fairness, explainability and responsibility*

The report lays out 20 recommendations including the need to;

- **Recommendation 1 - Ensure that CAVs reduce physical harm to persons.**
  - ...establish an objective baseline and coherent metrics of road safety that enable a fair assessment of CAVs' performance relative to non- CAVs and thereby publicly demonstrate CAVs' societal benefit.
    - This should be accompanied by new methods for continuously monitoring CAV safety and for improving their safety performance where possible.
- **Recommendation 5 - Redress inequalities in vulnerability among road users.**
  - *address current and historic inequalities of road safety, CAVs may be required to behave differently around some categories of road users, e.g. pedestrians or cyclists, so as to grant them the same level of protection as other road users.*
- **Recommendation 6 – Manage dilemmas by principles of risk distribution and shared ethical principles.**
  - *...it may be impossible to regulate the exact behaviour of CAVs in unavoidable crash situations, CAV behaviour may be considered ethical in these situations provided it emerges organically from a continuous statistical distribution of risk by the CAV in the pursuit of improved road safety and equality between categories of road users.*
- **Recommendation 12 – Audit CAV Algorithms**
  - Investments in developing algorithmic auditing tools and resources specifically adapted to and targeting the detection of unwanted consequences of algorithmic system designs and operations of CAVs are recommended.

---

<sup>113</sup> Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659). Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility. 2020. Publication Office of the European Union: Luxembourg.



- **Recommendation 13 - Identify and protect CAV relevant high-value datasets as public and open infrastructural resources.**
  - ... data on crash or near-crash situations should be identified and kept free and open, insofar as they can be likened to infrastructural resources that support free innovation, competition and fair market conditions in CAV related sectors.
- **Recommendation 14 - Reduce opacity in algorithmic decisions**
  - User-centred methods and interfaces for the explainability of AI-based forms of CAV decision-making should be developed. The methods and vocabulary used to explain the functioning of CAV technology should be transparent and cognitively accessible, the capabilities and purposes of CAV systems should be openly communicated, and the outcomes should be traceable.
- **Recommendation 17 - Promote a culture of responsibility with respect to the obligations associated with CAVs.**
  - Similar to what happened, for instance, in aviation in relation to the creation of a culture of safety or in the medical profession in relation to the creation of a culture of care, a new culture of responsibility should be fostered in relation to the design and use of CAVs.
- **Recommendation 18 – Ensure accountability for the behaviour of CAVs (duty to explain)**
  - “Accountability” is here defined as a specific form of responsibility arising from the obligation to explain something that has happened and one's role in that happening. A fair system of accountability requires that: (a) formal and informal fora and mechanisms of accountability are created with respect to CAVs; (b) different actors are sufficiently aware of and able to discharge their duty to justify the operation of the system to the relevant fora; (c) and the system of which CAVs are a part is not too complex, opaque, or unpredictable.
- **Recommendation 19 – Promote a fair system for the attribution of moral and legal culpability for the behaviour of CAVs.**
  - The development of fair criteria for culpability attribution is key to reasonable moral and social practices of blame and punishment - e.g. social pressure or public shaming on the agents responsible for avoidable collisions involving CAVs – as well as fair and effective mechanisms of attribution of legal liability for crashes involving CAVs. In line with the principles of fairness and responsibility, we should prevent both impunity for avoidable harm and scapegoating.
- **Recommendation 20 - Create fair and effective mechanisms for granting compensation to victims of crashes or other accidents involving CAVs.**
  - Clear and fair legal rules for assigning liability in the event that something goes wrong with CAVs should be created. This could include the creation of new insurance systems. These rules should balance the need for corrective justice, i.e. giving fair compensation to victims, with the desire to encourage innovation.



Many of these recommendations, and others detailed within the report, are directly relevant to the proposed regulatory framework, the no blame safety culture, the need for continual learning and comparability of safety performance of CAVs and non-CAVS.

They also align well with the suggestions put forward in this paper regarding;

- the need for CAV explainability, transparency and outcome traceability
- the need to identify high-value datasets including crash and near-crash situations in support of free innovation, competition and fair market conditions
- the need to redress inequalities in vulnerability among road users
- the need for new methods for continuously monitoring CAV safety

*It is suggested that the proposed regulatory framework makes reference to the ethics of connected and automated vehicles, identifies the appropriate legal entity responsible and considers expanding the role of the “safety assurance regulator” to include these important topics on road safety, privacy, fairness, explainability and responsibility.*

#### 6.5.5. Safety Assurance Regulator – extended responsibility for UN Improving Global Road Safety resolution

In August 2020 the United Nations adopted a new resolution for “*Improving global road safety*”<sup>114</sup> containing the following important statements;

- *Reaffirming its resolution 70/1 of 25 September 2015, entitled “Transforming our world: the 2030 Agenda for Sustainable Development”*
- *Recalling that the Sustainable Development Goals and targets are integrated and indivisible and balance the three dimensions of sustainable development, and acknowledging the importance of reaching the road safety-related targets of the 2030 Agenda*
- *Recognizing that human suffering, combined with costs to some countries of up to 5 per cent of their gross domestic product a year, makes reducing road traffic deaths and injuries both an economic and a social priority...*
- *Taking into account that road traffic deaths and injuries are also a social equity issue, as the poor and the vulnerable are most frequently also vulnerable road users, namely, pedestrians, cyclists, users of motorized two- and three-wheeled vehicles and passengers of unsafe public transport, who are disproportionately affected and exposed to risks and road traffic crashes...*
- *Recognizing that road safety requires addressing broader issues of equitable access to mobility and that the promotion of sustainable modes of transport,*
- *Encouraging Member States to promote multi-stakeholder partnerships to address the safety of vulnerable road users, the delivery of first aid to victims of*

---

<sup>114</sup> Resolution A/RES/74/299 74/299. Improving global road safety adopted by the General Assembly on 31 August 2020

*road traffic accidents or crashes, training and education, notably in developing and least developed countries,*

- *Noting that continuous progress of automotive and digital technologies could improve road safety, including through the progressive development of highly and fully automated vehicles in road traffic, and in this regard noting with appreciation the adoption under the auspices of the Economic Commission for Europe of a Global Forum for Road Traffic Safety resolution on the deployment of highly and fully automated vehicles in road traffic,*
- *Invites Member States to encourage and incentivize the development, application and deployment of existing and future technologies and other innovations to improve accessibility and all aspects of road safety, from crash prevention to emergency response and trauma care, with special attention given to the safety needs of those road users who are the most vulnerable, including pedestrians, cyclists, motorcyclists and users of public transport;*

*It is suggested that the proposed regulatory framework make specific reference to UN resolution on Improving Global Road Safety and identifies the appropriate legal entity responsible and considers expanding the role of the “safety assurance regulator” to include delivery against the specified goals and targets of the resolution and referenced United Nations Sustainable Development Goals.*

#### 6.5.6. The monitoring test and control test – revisited and revised

The consultation paper makes reference to the ALKS Call for Evidence in which;

...the Government sets out two tests which it proposes to use to decide whether a vehicle or type of vehicle is capable of driving itself safely: the *monitoring test* and the *control test*<sup>115</sup>.

(1) The *monitoring test* states that an individual does not need to monitor the vehicle if the vehicle can safely achieve the following without human monitoring:

- (a) comply with relevant road traffic rules;
- (b) avoid collisions which a competent and careful driver could avoid;
- (c) treat other road users with reasonable consideration;
- (d) avoid putting itself in a position where it would be the cause of a collision;
- (e) recognise when it is operating outside of its operational design domain.

(2) The *control test* states that a vehicle is not being ‘controlled’ by an individual if the individual controls none of the following:

---

<sup>115</sup> Safe Use of Automated Lane Keeping System ALKS Call for Evidence page 25 para 3.36 and 3.27  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/921409/Safe-Use-of-Automated-Lane-Keeping-System-ALKS-Call-for-Evidence-FINAL-accessible.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/921409/Safe-Use-of-Automated-Lane-Keeping-System-ALKS-Call-for-Evidence-FINAL-accessible.pdf)

- (a) longitudinal dynamics (speed, acceleration, braking, gear selection);
- (b) lateral dynamics (steering).

It can be argued that neither test is valid and that both look at the categorisation from the wrong direction. The most important thing is not *what individual no longer needs to do* but *what the ADS requires of the individual*.

In order to “*decide whether a vehicle or type of vehicle is capable of driving itself safely*” there are two very simple alternatives, the *driver license test* and the *driver monitoring test*.

1. The “*driver license test*” would state that if a licensed driver is required during ADS operation the vehicle is not self-driving.
2. The “*driver monitoring test*” would state that if a driver availability monitoring system is required during ADS operation the vehicle is not self-driving.

It's clear that a SAE Level 4 High Driving Automation would meet this requirement due to its ability to safely execute the entire dynamic driving task, including minimal risk manoeuvres, in a passenger only or unoccupied mode.

It's also clear that ALKS, and an SAE Level 3 Conditional Driving Automation would fail both tests due to their requirement for a licensed human driver to be receptive to transition demand and for the licenced driver to detect and response to DDT performance-relevant system failures in vehicle systems. As such neither can logically be categorised as self-driving under any circumstances or conditions.

It also follows that a User-In-Charge definition that requires a human driver who is “qualified and fit to drive” and “is receptive to a transition demand” would also fails these two tests and cannot be “in charge” of a self-driving vehicle.

However, the consultation paper does describe the potential for a User-In-Charge “transition information” mode of operation which conceivable be extended to include passenger only modes that do not require the occupant to have a driver's license.

Using *driver license and driver monitoring tests* above as sufficient for categorising a vehicle as self-driving without the need for the additional “*control test*”. However, that test does have some benefits that are explored further below.

#### 6.5.6.1. *The AD control test to determine use only with Human Driver Supervision (HDS)*

What should happen to automated vehicles modes that fail the *driver license and driver monitoring tests*? Is categorisation as “not self-driving” sufficient for safety? How can the categorisation of SAE Level 2, SAE Level 3 and ALKS be determined?

Firstly, all these systems fail the *driver license and driver monitoring tests*, so all that remains is how they are differentiated from SAE Level 0 and SAE Level 2 systems. For that it is suggested that the “control test” be refocused from “*what is not being controlled by individual*” to “*what is being controlled by the automated driving system*” i.e. the *AD Control Test*;

1. The “*AD control test*” would state that a vehicle is being ‘controlled’ by automated driving system if the software controls both of the following;
  - (a) longitudinal dynamics (speed, acceleration, braking, gear selection);
  - (b) lateral dynamics (steering).

SAE Level 0 fails the test as the human driver controls both axes. SAE Level 1 fails the test as the human driver controls one or other of the axes. SAE Level 2, SAE Level 3 and ALKS all pass the test as the software is in control of both longitudinal and lateral dynamics.

The language used within the consultation paper is;

- AVs that can only be used with a User-In-Charge
- AVs that can be used without a User-In-Charge

It is suggested that an AV that fails the *driver license and driver monitoring tests* but passes the *AD control test* could be described as;

- AVs for use only with Human Driver Supervision (HDS)

*It is suggested that this new categorisation of Human Driving Supervision (HDS) would acknowledge that the requirements for “self-driving” have not been met whilst also acknowledging that control is being executed by an automated driving system which must have a licensed human driver to be present, fit to drive and monitored.*

*It is suggested that the regulatory framework explicitly defines what happens in the case that an ADS, such as Tesla Full Self Driving or ALKS, which would fall under the Human Driver Supervision category, has already been categorised as driver assistance by another countries type-approval authority. Does ADS immediately fall outside of the proposed safety assurance scheme?*

#### 6.5.1. ADSE regulatory sanctions—Penalty Points (Endorsements)

Consultation Question 23 asks for feedback on the range of regulatory sanctions<sup>116</sup>;

---

<sup>116</sup> Consultation Paper 3 – page 185 para 11.53

*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.*

For human drivers there is a system for fines and endorsements (penalty points)<sup>117</sup>;

- The courts can fine you and 'endorse' your driving record with penalty points if you're convicted of a motoring offence.*
- Endorsements must stay on your driving record for 4 or 11 years, depending on the offence.*
- You can be disqualified from driving if you build up 12 or more penalty points within a period of 3 years.*

There is already a comprehensive list of endorsement codes and penalty points;

- Each endorsement has a special code and is given 'penalty points' on a scale from 1 to 11. You get more points for more serious offences.*
- Some offences may also involve a disqualification.*
- Offence codes and penalty points must stay on your driving record for 4 or 11 years depending on the offence.*

Relevant endorsement codes for "self-driving" appear in the following categories;

<b>Code</b>	<b>Offence</b>	<b>Penalty points</b>
<b>AC10</b>	Failing to stop after an accident	5 to 10
<b>AC20</b>	Failing to give particulars or report an accident within 24 hours	5 to 10
<b>AC30</b>	Undefined accident offences	4 to 9
<b>BA10</b>	Driving while disqualified by order of court	6
<b>BA30</b>	Attempting to drive while disqualified by order of court	6

<sup>117</sup> <https://www.gov.uk/penalty-points-endorsements>

<b>Code</b>	<b>Offence</b>	<b>Penalty points</b>
<b>BA40</b>	Causing death by driving while disqualified	3 to 11
<b>BA60</b>	Causing serious injury by driving while disqualified	3 to 11
<b>CD10</b>	Driving without due care and attention	3 to 9
<b>CD20</b>	Driving without reasonable consideration for other road users	3 to 9
<b>CD30</b>	Driving without due care and attention or without reasonable consideration for other road users	3 to 9
<b>CD80</b>	Causing death by careless, or inconsiderate, driving	3 to 11
<b>CD90</b>	Causing death by driving: unlicensed, disqualified or uninsured drivers	3 to 11
<b>DD10</b>	Causing serious injury by dangerous driving	3 to 11
<b>DD40</b>	Dangerous driving	3 to 11
<b>DD60</b>	Manslaughter or culpable homicide while driving a vehicle	3 to 11
<b>DD80</b>	Causing death by dangerous driving	3 to 11
<b>DD90</b>	Furious driving	3 to 9
<b>MS10</b>	Leaving a vehicle in a dangerous position	3
<b>MS30</b>	Play street offences	2
<b>MS50</b>	Motor racing on the highway	3 to 11
<b>MS60</b>	Offences not covered by other codes (including offences relating to breach of requirements as to control of vehicle)	3
<b>MW10</b>	Contravention of special roads regulations (excluding speed limits)	3
<b>PC10</b>	Undefined contravention of pedestrian crossing regulations	3
<b>PC20</b>	Contravention of pedestrian crossing regulations with moving vehicle	3

Code	Offence	Penalty points
PC30	Contravention of pedestrian crossing regulations with stationary vehicle	3
SP10	Exceeding goods vehicle speed limits	3 to 6
SP20	Exceeding speed limit for type of vehicle (excluding goods or passenger vehicles)	3 to 6
SP30	Exceeding statutory speed limit on a public road	3 to 6
SP40	Exceeding passenger vehicle speed limit	3 to 6
SP50	Exceeding speed limit on a motorway	3 to 6
TS10	Failing to comply with traffic light signals	3
TS20	Failing to comply with double white lines	3
TS30	Failing to comply with 'stop' sign	3
TS40	Failing to comply with direction of a constable/warden	3
TS50	Failing to comply with traffic sign (excluding 'stop' signs, traffic lights or double white lines)	3
TS60	Failing to comply with a school crossing patrol sign	3
TS70	Undefined failure to comply with a traffic direction sign	3

Consultation Question 18 provisionally proposed<sup>118</sup>;

*...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;*

It would seem clear that these offenses, and the related codes, would provide a suitable basis for a comparative scheme between automated driving systems and human drivers.

<sup>118</sup> Consultation 3 page 169 para 10.83



This would complement the value of the OEDR in-use monitoring scheme which would form the foundation for detecting the offences in the first place.

It might also be expected by the public that an ADSE is granted a license under the condition that the operate under a similar endorsement-based system, albeit one that might have a different points structure.

*It is suggested that the proposed regulatory framework considers the value of creating an endorsement-based scheme for ADSE licensing, that the scheme references the same offenses and codes of the human scheme but remains flexible for the safety assurance regulator to set the penalty points structure and related thresholds for suspension and withdrawal of authorisation.*



Chair ITU FG-AI4AD Bryn Balcombe: [REDACTED]

General mailing list: [fgai4ad@lists.itu.int](mailto:fgai4ad@lists.itu.int)

Dedicated secretariat email: [tsbfgai4ad@itu.int](mailto:tsbfgai4ad@itu.int)

Dedicated webpage: [www.itu.int/en/ITU-T/focusgroups/ai4ad](http://www.itu.int/en/ITU-T/focusgroups/ai4ad)

In memory of Elaine Herzberg  
18<sup>th</sup> March 2018