

June 2021

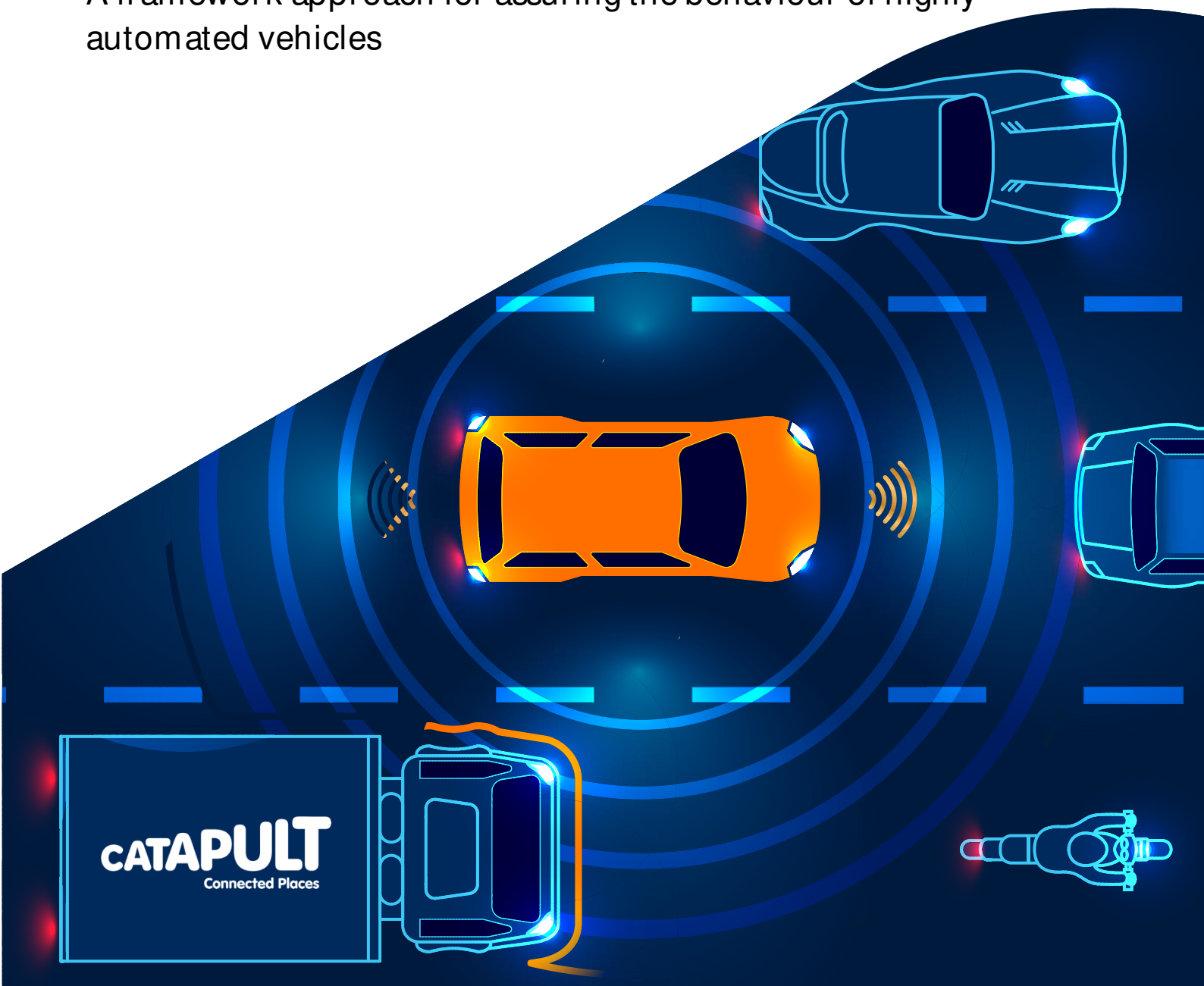
CertiCAV

CATAPULT
Connected Places

WMG
THE UNIVERSITY OF WARWICK

CertiCAV Assurance Paper

A framework approach for assuring the behaviour of highly automated vehicles



CertiCAV Key Findings

The CertiCAV project has produced the following key findings, relevant for informing the future of Highly Automated Vehicle assurance.

These findings are summarised below with links to the relevant sections in the report. In line with the scope and approach of the CertiCAV project (see Section 1.2) these findings:

- Focus on the assurance of the perception and tactical decision making performed by a nominally performing Automated Driving System (ADS) integrated into a Highly Automated Vehicle (HAV)
- Define a relatively high level framework that takes first steps to form the basis of a safety assurance scheme for the deployment to market of HAVs
- May be applicable to trials, but this report does not attempt to directly address trial deployments. However, there is a discussion on application of system approval to trials in Section 5.4. The definition of an advanced trial would need to be clarified before CertiCAV's context to advanced trials can be defined (see Section 5.4.1)

Please refer to the **Terminology** table of this document for a definition of terms and abbreviations.

Context, system definition and specification

- The framework intends to cover the assurance for the new entity of a 'HAV' being introduced to the road traffic environment. The introduction of this entity must draw from and contribute to existing regulation and practice (see Section 2.1.1)
- Safety is important for 'assurance', however other factors including road user co-operation and transport efficiency are also concerns (see Section 2.1.2). This contrasts with 'safety assurance' which is limited to safety concerns (see Section 2.1.2.1)
- There are several stakeholders who would have to support the deployment of HAV on roads. Key assurance stakeholders are the Automated Driving System Entity (ADSE) who develops the system, the regulator/Type Approval Authority (TAA) who approves its use on roads, and the technical service who provides independent assessments to support approvals (see Section 2.2)
- We have developed a common system specification, including a definition of a HAV, and a breakdown of items of an ADS to support common specification and understanding amongst stakeholders. Wirelessly connected systems may impact HAV behaviour, and to capture this we have defined the Highly Automated SuperSystem (HASS) (see Section 2.3)
- We have identified the importance of having a well-developed HASS specification to provide explanation between stakeholders and support their activities. We present a structure and definition for a HASS specification with the intent to facilitate the development of a common system of specification that supports good relationships and co-operation between stakeholders (see Section 2.4)
- The Operational Design Domain (ODD) definition language used should be agreed between various stakeholders, including manufacturers, regulators, and local authorities

Common Assurance practice

- There is a need to develop performance indicators as a common reference that supports the definition of: requirements, assessments, management of system safety/quality, and ongoing monitoring. We have identified four categories of performance indicator, each of which may perform different roles in an assurance framework (see Section 3.2)
- The development and use of formalised performance indicators enables auditing of the system developer's quality/safety case argumentation and its supporting evidence by providing traceability to requirements
- There is a need to develop regulatory requirements covering HASS performance standard conformance, in particular for safety (see Section 3.3)
 - Compared to the existing practice of vehicle type approval which predominately uses relatively prescriptive specifications, needs-based requirements may play an increased role for regulating HAVs. This is due to the relative complexity and wide responsibilities applied to driving tasks performed by HAV driving functions compared to existing ADAS driving functions
 - We present an approach for defining needs-based requirements for HAVs specified with components of conditions, limitations, and certainty. Driving performance criteria could be used to help identify and verify appropriate requirements (see Section 3.3.3). This significant work area would address the "how safe is safe enough" question and is interdependent on other work areas such as the development of a codified Highway Code
- We discuss the state of existing practice relevant for HAV assurance and identify gaps. There are needs for best practice development for HAV assurance both to support the development of regulatory performance requirements, and to facilitate the assurance of their achievement (see Section 3.4)
- The intended behaviour of HAVs is a complex topic, and there is a significant need to develop practice to cover both creation and validation of intended behaviour. This is required to ensure a HAV's specification is complete and sufficient for taking full responsibility of the driving task. This field of practice we term behavioural safety, and includes:
 - The handling of road traffic rules and co-operation. Developing a codified Highway Code would develop an understanding of desired behaviour of HAVs including guidance that is needed on Minimal Risk Manoeuvres and Minimal Risk Conditions (see Section 3.4.5.1)
 - The risk control strategy for HAV is not limited to response/mitigation of hazards, as the HAV's behaviour influences the exposure to hazards in the multi-agent environment it is operating within. Risk modelling should therefore account for this area of tactical safety, in addition to the more established area of active safety (see Section 3.4.5.2)
- Probabilistic systems are typically involved in ADS perception, and may be involved in decision making. Some of this area is covered by current Safety of The Intended Functionality (SOTIF) best practice. However, practice for verification and validation (V&V) of systems trained using machine learning is insufficiently mature, and additional research and development on this topic is needed
- Practice is required to ensure the suitable specification and use of sensors and data systems, to support modelling and definition of physics-based sensors
- Although outside of CertiCAV scope, we have identified some needs to develop harmonised standards and principles for appropriate use of co-operative sensors (see Section 3.4.7.1)

- Validation and testing (see Section **3.4.9**)
 - On-road testing is important for validation; however, risks need to be managed, especially risks to the public
 - Practice is required for the management of complexities of HASS development and management of system safety cases
 - System development processes for HAV have various interrelated inputs and outputs with each other that need to be co-ordinated
 - Developments are required to manage the significant role of virtual testing in HAV assurance. Tooling needs to be V&V'd to have appropriate fidelity for the attributes it is modelling and the role of the evidence it creates in safety argumentation
 - The development of standardised a HAV manoeuvre library (see Section **3.4.9.3**) would help specify ODDs and generate scenarios. Furthermore, it could formalise part of the vocabulary for a codified Highway Code (see Section **3.4.5.1**)
 - Scenario based testing is important for assessment and practice is required for generation of scenarios and mapping these to coverage of safety and performance. Both the ODD and Behaviour of a HAV are important for determining the set of scenarios that are relevant for HAV assessment testing, and system development V&V
 - In order to use safety case evidence based on simulation-based testing, there is a need to create a standard set of metrics and a process to validate the suitability of simulation testbench for the type and role of the evidence it generates
- While BSI (The British Standards Institution, 2018) and the UK government (HM Government, 2017) have created excellent principles for HAV cybersecurity, it may be beneficial to establish a detailed process for assuring the cybersecurity of HAVs (noting this is out of scope for CertiCAV)
- A common challenge is that the performance indicators, requirements, and codes currently applied to human drivers do not quantify performance or support traceability to formal V&V of performance achievement. There is therefore a need to formalise these to support the assurance of HAV

System Development

At least for the time being there is a lack of established good practice and harmonisation to facilitate assurance through demonstration of conformance, therefore the development of a safety case is considered to be important for the management safety and quality during HASS system development. In Section **4.2** we identify a set of principles, objectives and requirements that may be applied to the practice of system safety case. This includes various development processes to be covered in the system safety case, which in turn reference the discussion of HAV assurance practice. This includes that assurance should be achieved by the development of a system safety case that:

- Is formalised, structured and consistent
- Relates to performance indicators to provide traceability to requirements and to evidence
- Covers and co-ordinates system development processes
- Demonstrates achievement of safety requirements and is consistent with defined organisational risk appetite

System Approval

- In Section 5.2 we outline a process of independent assessment associated with the system approval of a HASS
- This includes presenting a high-level guide for audit and inspection processes undertaken by an independent assessor
- It will take time and experience to develop a mature system of independent assessment. The content in this section may be considered as a starting point for engagement and development
 - We outline activities at a high level including audit/inspection, road testing, testbench and physical testing. However, refining and developing a more comprehensive low-level guide may be useful to ensure consistency of the assurance process
- There are alternative viewpoints of focussing on audit/review of process, and focussing on scenario testing. Both viewpoints have been advocated by different industry members we have engaged with. A blend of both approaches, process and evidence review/audit followed by verification/consistency scenario testing is a potential best-of-both option
- We have identified the need to implement processes to capture learnings from current and future consultation/approval activities to facilitate:
 - Continuous improvement of system approval activities
 - Development and improvement of standards/practice
- We have identified three key areas to be covered:
 - Learning: There is a need to develop and manage knowledge relating to HASS related audit and assessment activities. RDW, the Netherlands Vehicle Authority, present an approach for this, including lists of inspection/audit questions in their Vehicle Safety and Security Framework ¹
 - Test Scenarios: There is a need to develop database for logical scenarios and parameter distributions. National/international collaboration regarding scenario databases would deliver value in making assurance processes more efficient, transferable, and scalable
 - Techniques & Tooling: There is a need to develop capabilities to support assessment activities. This includes modelling tooling, and assessment techniques including test selection and performance assessment
- Further observations are that:
 - The behaviour of a HASS is influenced by a very large range of parameters. Assuring HASS behaviour would involve assessing an accordingly large body of evidence demonstrating HASS behaviour across combinations of parameter conditions. However, section 5.3.2 finds that assessing evidence demonstrating the behaviour of separate perception and behaviour subsystems of an ADScan substantially reduce the overall quantity of evidence assessment required to assure HASS behaviour to a given level. While the provision of separate evidence for these subsystems may not be necessary, in cases where separate evidence could be provided, standardisation activities in this area would aid assessment practice. For example, a standard interface between the perception and behaviour subsystems would allow assessors to more easily confirm that expected results have been obtained, and that expected scenarios have been tested

¹ (RDW, 2019)

- Guidelines on what independent tests should be carried out as part of the assurance process are needed. A common stakeholder view is that a short public road "test drive" of the HAV would be appropriate, but other tests may also be cost-effective for increasing confidence in the evidence presented by the ADSE (see Sections **5.2.2** and **5.1.3**)
- If independent assurance relies on common, standardised scenarios, then there are advantages to creating these scenarios by an expert group representing a range of stakeholders and by following a best-practice process (see Section **5.2.4.1**)

Legal and through-life issues

While legal and through-life issues are out of scope for CertiCAV, they are critical for HAV safety and so suggested areas for further work are listed here.

- Further development of operational safety guidelines for HAVs is likely to be required
- There are needs for some aspects of in-use safety to be covered by regulations to clarify responsibilities and liabilities. For example, the correct functioning of a HAV's ADS could be tested by a 3rd party on a yearly basis, and vehicle sensors may need to be regularly inspected and cleaned. The Law Commissions' third consultation suggests a User-in-Charge or Licenced Fleet Operator would have some of these responsibilities²
- A mechanism for ensuring the safety of software updates issued by ADSEs should be enacted
- Through-life monitoring of deployed HAVs will allow regulators to understand the safety of HAVs with a much higher degree of certainty than the pre-market processes that are the scope of CertiCAV
- It may be beneficial to treat safety validation as a continual process (especially for some criteria/ metrics), as the driving environment, societal expectations, and body of knowledge on HAV assurance will change over time

² (The Law Commissions, 2020)

Contents

Context, system definition and specification	1
Common Assurance practice	2
System Development	3
System Approval	4
Legal and through-life issues	5
Disclaimer	11
Terminology	11
1 Introduction	17
1.1 Background	17
1.2 Scope and approach	17
1.2.1 From trials to deployment - a process of learning approach	20
1.3 Content	21
2 Context and Assumptions	22
2.1 CertiCAV's use-case	22
2.1.1 Position of CertiCAV in overall system	23
2.1.2 Context of assurance	24
2.1.3 Context of on-road HAV deployment	27
2.2 Stakeholder roles, responsibilities and needs	29
2.3 Common system definition	33
2.3.1 Highly Automated Vehicle (HAV)	33
2.3.2 Automated Driving System (ADS)	33
2.3.3 Highly Automated SuperSystem (HASS)	38
2.4 HASS specification	39
2.4.1 Vehicle operational characteristics	40
2.4.2 Operational Design Domains (ODDs)	41
2.4.3 ADS functionality	42
2.4.4 HASS architecture	43

2.4.5	Operational management	44
2.5	Assurance goals	45
2.5.1	System development	45
2.5.2	System approval	45
3	Common Assurance Practice	46
3.1	Introduction	46
3.2	Performance Indicators	47
3.2.1	Event occurrence	47
3.2.2	Performance models	50
3.2.3	Formal specifications	51
3.2.4	Informal specifications	52
3.3	Developing performance requirements for HAVs	53
3.3.1	Introduction	53
3.3.2	Defining requirements for performance needs	55
3.3.3	Driving performance criteria	57
3.3.4	Application challenges	60
3.3.5	Requirement development options	62
3.4	HAV assurance practice	66
3.4.1	HAV assurance best practice	66
3.4.2	Safety & quality management	68
3.4.3	Operational safety	71
3.4.4	Functional safety (FuSa)	71
3.4.5	Behavioural safety	72
3.4.6	Probabilistic systems	81
3.4.7	Sensors & Co-operative safety	89
3.4.8	Data management	91
3.4.9	Validation and testing	94
3.5	Summary Findings	101
4	System Development	103
4.1	Introduction	103
4.2	System Safety Case	103
4.2.1	Operational safety	106

4.2.2	Functional safety	106
4.2.3	Behavioural safety	107
4.2.4	Probabilistic systems	107
4.2.5	Sensors & co-operative safety	108
4.2.6	Data management	108
4.2.7	Validation & testing	109
4.3	Summary	110

5	System Approval	111
5.1	Introduction	111
5.1.1	Context and Objective	112
5.1.2	Deployment risk classification	112
5.1.3	Change control	113
5.2	Approval process	114
5.2.1	HASS Specification & Safety Check	116
5.2.2	Road test	118
5.2.3	Audit and inspection	120
5.2.4	System Safety case / evidence evaluation	122
5.2.5	Scenario generation	124
5.2.6	Testbench testing	128
5.2.7	Physical testing	129
5.2.8	Fidelity check	130
5.3	Performance Assessment	131
5.3.1	Performance Indicators	131
5.3.2	Assessment techniques	131
5.4	Application to trial deployments	132
5.4.1	Definition of advanced trial	133
5.5	Summary findings	135

6	Acknowledgements	136
----------	-------------------------	------------

7	Works Cited	137
----------	--------------------	------------

8	Appendices	142
A	Workshop & Expert Consultations	142
8.1	Workshop 1: Consultation Webinar	142
8.2	Workshop 2: HAV Assurance, Requirements, Goals, & Strategy	143
8.3	Aviation Interview Study	144
8.3.1	Introduction	144
8.3.2	Survey Outputs	144
8.3.3	Relevant Regulation	146
8.3.4	Conclusion	147
8.3.5	Acknowledgement	147
8.4	WMG Interview Study	148
8.4.1	Method:	148
8.4.2	Participants:	148
8.4.3	Interview design:	149
B	State of the Art	153
8.5	Law Commissions Consultations	153
8.5.1	Consultation 1	153
8.5.2	Consultation 2	158
8.6	159	
8.7	An overview of Systems Theoretic Process Analysis (STPA)	160
8.7.1	Step 1: Define purposes of the analysis	160
8.7.2	Step 2: Model the Control Structure	161
8.7.3	Step 3: Identify Unsafe Control Actions	162
8.7.4	Step 4: Identify Loss Scenario	162
8.7.5	STPA in the SOTIF context	163
8.8	BSI PAS 1880:2020 Guidelines for Developing and Assessing Control Systems for Automated Vehicles	164
8.9	BSI PAS 1881: Assuring the safety of automated vehicle trials and testing – Specification	165
8.10	Draft PAS 1882:2020	166

8.11	Performance Testing for Sensors in Connected and Autonomous Vehicles: Feasibility Study	167
8.11.1	Conclusions	167
8.12	VeriCAV	168
8.13	dRISK	169
8.14	PD ISO/PAS 21448:2019, Road Vehicles – Safety of the intended functionality	169
8.15	UL 4600	172
8.16	TR68: Singapore's technical Reference for AVs	174
8.17	RAND – Approaches to assessing acceptable safety for automated vehicles	175
8.17.1	Safety as a measurement	175
8.17.2	Safety as a process	176
8.17.3	Safety as a threshold	177
8.17.4	Conclusions	178
C	Example HASS specification	179
8.18	Introduction	179
8.19	Parking Valet, example HASS specification	179
8.20	HASS specification relationship to system safety case and system assessment	181

Disclaimer

The CertiCAV project, and by extension this document, is intended to be a thought-piece in line with the aims and scope of the project. The document does not intend to be exhaustive, present authoritative recommendations or final solutions. Where assumptions have been made, for the purpose of project delivery, these have been set without prejudice to alternatives.

This document does not necessarily represent the views of the Department for Transport or any UK government body.

Terminology

This document assumes the terminology defined in the 'BSI Connected and automated vehicles – Vocabulary' (BSI Flex 1890 v3.0:2020-10)³. These are supplemented by further terms related to driving automation systems adopted from SAE J3016_201806⁴ including the driving automation levels. This report adopts terminology defined in the Law Commissions' Automated Vehicle consultations⁵. The below table describes the key terms and abbreviations used in this report, including those from the aforementioned sources, and terms defined within this report.

Table 1: Terminology

Abbrev	Term	Definition
ADS	Automated Driving System	An ADS is the collection of hardware and software that are collectively capable of performing the entire DDT on a sustained basis. A HAV will have at least one ADS installed. (SAE International, 2018)
ADSE	Automated Driving System Entity	This document defines the ADSE as the organisation or individual that presents a HASS for approval for use and is responsible for its system safety, relevant non-safety performance, and through life support, including co-ordination with the UIC / LFO and/or service providers involved in the deployment operation of the HASS. The ADSE may be liable for dynamic driving task related criminal and civil offences when the ADS is engaged. This is a development of the CAV Vocabulary² definition of the ADSE as 'the organisation or individual that presents a HAV deployment system for authorisation for use and is responsible for its system safety'. Which is in turn based on The Law Commissions' definition of this term⁴. We have adapted these previous definitions to refer to HASS rather than HAV and include non-safety, support and co-ordination responsibilities based on the findings of CertiCAV.

³ (BSI, 2020)

⁴ (SAE International, 2018)

⁵ (The Law Commissions, 2020)

		In conventional automotive type approval, this organisation is referred to as the "manufacturer". This report uses the term ADSE to align with the BSI/Law Commissions terminology, and to explicitly acknowledge that the ADSE may not be the vehicle manufacturer in the conventional sense and/or may have different organisational responsibilities (see Section 2.2).
AI	Artificial Intelligence	Broad field of intelligence demonstrated by machines
ASAM	Association for Standardization of Automation and Measuring Systems	An automotive standards body based in Germany but operating internationally. Focuses on standards relating to automotive development and testing tool chains.
	Assurance	Within the context of this document 'assurance' relates to managing the operation within the road traffic environment toward achieving outcomes desirable for society. A primary assurance concern is safety, such as minimising the rate of road casualties. However, other concerns are included such as minimising conflicts between users, facilitating smooth and efficient traffic flow, and reducing property damage. Assurance, therefore, contrasts with the term 'safety assurance' which is limited to safety concerns.
	Behavioural Safety	Field of safety concerned with addressing the safety and sufficiency of the intended functionality of driving features
BSI	(The) British Standards Institution	The UK national standards body
CAV	Connected and Automated Vehicle	Used in this report only when referring to vehicles intended to remain below level 4 automation during operation, or when referring to external sources or the automated vehicle industry in general. Otherwise, the term HAV is used.
CHC	Codified Highway Code	A proposed development of The Highway Code which specifies driving rules and codes in a more codified explicit format with the intention of providing a reference for the development, assessment and regulation of HAV behaviour (see Section 3.4.5.1)
	Deployment Domain	The geofenced area within which a HAV is to be deployed. For HAVs which operate on a fixed route, this would be a tight geofence containing the route. For an ADS which can operate on any road conforming to its ODD, for example a motorway chauffeur system, this geofence would be the country of operation.
DDT	Dynamic Driving Task	All of the real-time operational and tactical functions required to operate a vehicle in on-road traffic, excluding the strategic functions such as trip scheduling and selection of destinations and waypoints. (SAE International, 2018).
	Ego vehicle	The HAV (or simulated HAV) that is under test.

FuSa	Functional Safety	Field of safety concerned with hazards arising from malfunctioning behaviour of systems. In automotive context FuSa is often used in reference to ISO 26262:2018 which is a risk-based safety standard aimed at addressing possible hazards caused by the malfunctioning behaviour of electronic and electrical systems in all road vehicles except mopeds.
GNSS	Global Navigation Satellite System	The general term for satellite navigation systems that provide geo-spatial positioning with global coverage. Operational GNSS systems include GPS, GLONASS, Galileo, and BeiDou.
	Harm	Injuries to road users (inside and outside the ego vehicle) who may include humans and livestock protected by legislation ⁶ . Note that this could occur without a collision, especially for the case of a vehicle with standing passengers.
HARPS	Highly Automated Road Passenger Services	A passenger transport service operated using HAVs. This term was coined by the Law Commissions in their second consultation paper. (The Law Commissions, 2019)
HASS	Highly Automated SuperSystem	The collection of systems involved in the conduct of the DDT and execution of the behaviour of a HAV. (see Section 2.3.3)
	HASS Specification	A specification relating to a HASS. (see Section 2.4)
HAV	Highly Automated Vehicle	A Highly Automated Vehicle is a vehicle equipped with an automated driving system that may perform the DDT within a specific operational design domain for some, or all, of the journey, without the need for human involvement in the DDT as a fallback to ensure road safety. In terms of automation level, HAVs correspond to level 4, and for the purpose of this document may also be assumed to include level 5. (see Section 2.3.1) The levels are defined in SAE J3016. (SAE International, 2018)
	HAV behaviour	HAV behaviour relates to any description of a HAVs actions.
	HAV manoeuvre	A HAV manoeuvre is a semantic description given to a set HAV behaviour that may be triggered by a tactical driving decision e.g. a "lane change".
HMI	Human Machine Interface	Systems that provide an input and/or output interface between humans and machines.
	Intended functionality	Defined in ISO/PAS 21448:2019 as "behaviour specified for a system" which in turn specifies intended behaviour as "specified behaviour of the intended functionality including interaction with items". This being 'intended' relates to nominal/fault free performance. In relation to HAV, intended functionality/behaviour usually relates to specification features that execute the dynamic driving task, although

⁶ (Highway Act 1835, Section 78, 1835)

		technically the specification could relate to other layers of abstraction.
LFO	Licenced Fleet Operator	See ' <i>User-In-Charge or Licenced Fleet Operator</i> '
MRC	Minimal Risk Condition	A condition to which a vehicle is brought to after performing an MRM to reduce the risk of harm when the execution of the DDT cannot or should not continue to be executed.
MRM	Minimal Risk Manoeuvre	The manoeuvre performed to bring the vehicle to an MRC when the execution of the DDT cannot or should not continue to be executed.
	Mitigated deployment	Refers to the involvement of safety mitigations in the deployment of an automated vehicle rather than operating the vehicle fully as a HAV. This is usually as part of a trial where involvement of a human driver is required to comply with laws, and/or as part of risk management due to the incomplete safety validation of the HASS. (see Section 2.1.3.3)
NUIC	No User-In-Charge	Relates to a HAV which is designed to operate without needing a human occupant. A NUIC HAV will not depend on a human occupant to take over the DDT at any point, or to perform any non-DDT duties expected of the driver of a conventional vehicle. (The Law Commissions, 2019)
	Nominal Performance	Refers to the performance of the ADS when the ADS and HAV are in a nominal state of operation. Nominal state of operation means free from erroneous conditions, and within specified tolerance for nominal state. Nominal performance therefore excludes performance during erroneous condition of the ADS or HAV, for example due to malfunctioning exteroceptive sensor. It also excludes off-nominal state such as low tyre pressure, unavailable data communications when the nominal state is defined in the HASS specifications. This however does not place any limitations on external conditions, which is managed by ODD.
ODD	Novel vehicle	A vehicle that does not conform to the existing Construction and Use regulations. For example, a vehicle without traditional human controls inside the vehicle.
	On-road deployment	Refers to a HAV operating on public roads subject to road traffic rules.
	Operational Design Domain	Referring to PAS1883 an ODD is the operating conditions under which a given driving automation system or feature thereof is specifically designed to function. (SAE International, 2018)
	Perception	The collection of on-board systems that derives tactical situational awareness. (see Section 2.3.2.1)

PI	Performance Indicator	A metric that quantifies performance of a system to evaluate level of success, determine achievements, and argue the satisfaction of specifications and needs. (see Section 2.3)
	Scenario	For the purpose of this document a Scenario is considered as combination and sequence of events and factors that may be present during a driving trip of the HAV. Factors/variables describing a scenario include: • Nature and behaviour of other actors (e.g. other vehicles, pedestrians etc) • ODD-related attributes, including the road network, infrastructure, and weather conditions. Infrastructure covers physical and digital infrastructure, and may include items not part of the road network (such as buildings and trees) • Any other dynamic elements, for example changing traffic signals, debris being blown into the road, or a change in the weather during the scenario • Initial conditions for the ego vehicle and any settings for the ADS (for example, the goal destination) With respect to the scope of this paper, the scenarios considered herein do not specifically consider: • Events and factors relating to systems performing in off-nominal conditions, such as due to malfunctioning behaviour of electronic and electrical systems (Functional Safety) • Events and factors arising from cyberattack It should be noted that definition above may differ from other document that apply scenarios in a different context. For example, scenarios defined in ISO 21448 SOTIF are the temporal development between several scenes in a sequence of scenes, this is applying scenarios in more specific and limited way than in this document.
	Sensory conditions	For the purpose of this document 'sensory conditions' is used as the term describe the conditions that affect the performance of environmental sensors. Alternative commonly used terms such as 'weather', or 'atmospheric conditions' indicate a scope that excludes several conditions that are relevant for sensor performance, such as: vibrations, pitch and roll, lens dirt, sensor misalignment, and degraded performance of sensor components.

SAE	SAE International, previously known as the Society of Automotive Engineers	An automotive-focused standards body which is based in the USA, however it operates internationally.
	Tactical decision-making	The decisions and control actions required of a vehicle driver (human or automated) to perform the DDT.
TAA	Tactical Safety systems	Systems that act to reduce exposure to hazards. These contrast with active safety systems, that are systems activated in response to hazards.
	Through-life safety	This includes in-use monitoring, OTA software updates, and inspections / physical condition monitoring.
	Type Approval Authority	Refers to the authority with responsibility for the approval of a vehicle, their systems, their components and separate technical units, and for the issuance of approval certificates.
	Unmitigated deployment	Refers to the deployment of a HAV that does not require human involvement in the driving task or being receptive to requests to intervene. Usually related to the market use of a HAV as a service or product. (see Section 2.1.3.3)
UIC / LFO	User-In-Charge or Licenced Fleet Operator	For the context of this document the entity responsible for non-automated driving task aspects of a HAVs operation may be User-In-Charge (UIC) or Licenced Fleet Operator (LFO), referred to as 'UIC / LFO' (see Sections 1.2 and 2.2) Due a HAV being defined as operated at level 4 automation level, the UIC/LFO do not have a responsibility in the DDT in this context. The Law Commissions' consultation paper 3 (CP3⁷) defines: • User-In-Charge (UIC): the human in the vehicle or in sight of the vehicle, with access to the controls • Licensed Fleet Operator (LFO): who would take responsibility for operating vehicles which are authorized for use without a user-in-charge. CP3 provides a detailed and distinct definition of the responsibilities for each, however this level of detail is not applicable to the scope for CertiCAV.
V2X	Vehicle to everything	The data exchange from a vehicle to entities in its surroundings using wireless technology that provides low latency and reliable data exchanges. V2X is an umbrella term incorporating several more specific types of vehicular communication including but not limited to: V2I (vehicle-to-infrastructure), V2N (vehicle-to-network), and V2V (vehicle-to-vehicle).

⁷ (The Law Commissions, 2020)

1 Introduction

1.1 Background

CertiCAV was a project undertaken by the Connected Places Catapult (CPC) and WMG, University of Warwick, and commissioned by the Department for Transport (DfT).

The aims of the CertiCAV project were to:

- Identify key considerations for assuring the safety and suitability of Highly Automated Vehicles (HAVs), discuss how these considerations contribute to assurance both individually and collectively, and suggest approaches in addressing them
- Through consideration of existing type approval practices, present suggestions on how different actors may contribute to HAV assurance
- Highlight some of the practical considerations when implementing certain approaches to assurance
- Draw together and build upon existing expertise and work done elsewhere in the UK to encourage knowledge transfer and the advancement of UK capability in this field
- Advance the knowledge of the Department in the field of assurance of HAVs, establishing some key foundations upon which an assurance scheme may be developed, as part of the CAVPASS⁸ programme, to enable the introduction of HAVs on GB roads.
- Support industry by disseminating key findings and therefore take a step toward the socio-economic benefits offered by deployment of HAVs

In pursuit of these aims, CertiCAV also developed and demonstrated a software toolkit to illustrate the application of the key aspects of this report.⁹ This toolkit is available as open-source software.¹⁰

1.2 Scope and approach

The scope of this document is limited to assurance of the perception and tactical decision making performed by a nominally performing Automated Driving System (ADS) integrated into a HAV in a

⁸ The CAVPASS (Connected and Autonomous Vehicle Process for Assuring Safety and Security) programme was established by the Department's Centre for Connected and Autonomous Vehicles (CCAV) in 2019. CAVPASS is much broader in scope than the CertiCAV project and therefore CAVPASS considers areas of safety and security which fall out of the scope of this project.

⁹ (Connected Places Catapult, 2021a)

¹⁰ <https://gitlab.com/connected-places-catapult/CertiCAV>

range of environmental conditions as appropriate for its ODD. However, the approaches considered herein are intended to be applicable to a range of HAVs and HAV deployments by making requirements proportional to the deployment domain and associated risks. Furthermore, 'novel' vehicle types in the context of providing a specific service (e.g. automated, driverless pods for the transport of goods or people within a specific environment) are considered within scope.

Other than through brief references (such as checks for absence of software errors) this document does not consider functional safety (electro-mechanical failures), cybersecurity, operational or through-life issues. The exclusion of operational and through-life issues means that the approaches presented herein are largely invariant to whether the HAV is operated privately (by a User-in-Charge or 'UIC') or operated as part of a service by a licensed fleet operator (or 'LFO'). Therefore, where it is relevant, this document assumes that the HAV-under-test is to be operated by a 'UIC / LFO' (please refer to document Terminology for definition).

For the sake of achieving the project aims, but without prejudice towards other approaches to assurance, this document further assumes that assurance is undertaken within the context of existing vehicle and system approval frameworks and that parties such as Type Approval Authorities and technical services assume similar roles. However, the needs of other stakeholders involved in the HAV deployment (for example, service providers) are also noted insofar as these stakeholders may require information pertaining to the safety performance of the HAV.

In summary, considerations that are in scope, and therefore explored within this document, include:

- The execution of Minimal Risk Manoeuvres (MRM) and suitability of the resulting Minimal Risk Condition (MRC)
- Identification of ODD boundaries
- Validation of deployment service, including sufficient understanding of the deployment domain

In summary, out of scope considerations include:

- Functional safety (electro-mechanical failures), although checks for absence of some types of software errors may be in scope. Therefore, any requirements on monitoring for electro-mechanical failures are also outside the scope of this report
- Ethics, although some ethical issues are touched upon in relevant subject areas for example on use of Artificial Intelligence, and safety targets – the topic of ethics is not intended to be addressed in this report
- Cybersecurity
- Electromagnetic interference. Electromagnetic interference that is acceptable during nominal operation may be defined in specifications, however, considering whether the specifications are appropriate for operating conditions, the validating the risk exposure and test methods are not considered in scope for this document
- Operational procedures related to safety
- Through-life issues, such as:
 - Data recording
 - Accident investigation and reporting
 - Continuous improvement of the approval process
 - Over the air software updates
 - Periodic Technical Inspections ("MOT" tests), ongoing maintenance and testing
- Human factors operator/handover issues
- Non-ODD boundary related triggering of MRM

In support of this scope and the project aims above, the approach taken throughout this document is to:

- Be forward looking and consider developments required for longer term assurance mechanisms that are required to bring AVs to market
- Examine the different activities undertaken as part of assurance and bring some clarity on their specific contributions, both individually and collectively
- Consider practicality and proportionality to risks/controls in order not to overly restrict development and to facilitate smaller deployments and deployments involving smaller organisations
- Identify the needs of essential supporting stakeholders in the wider assurance ecosystem
- Provide a structured and justified framework to initiate constructive ongoing engagement with key stakeholders to further develop HAV assurance

1.2.1

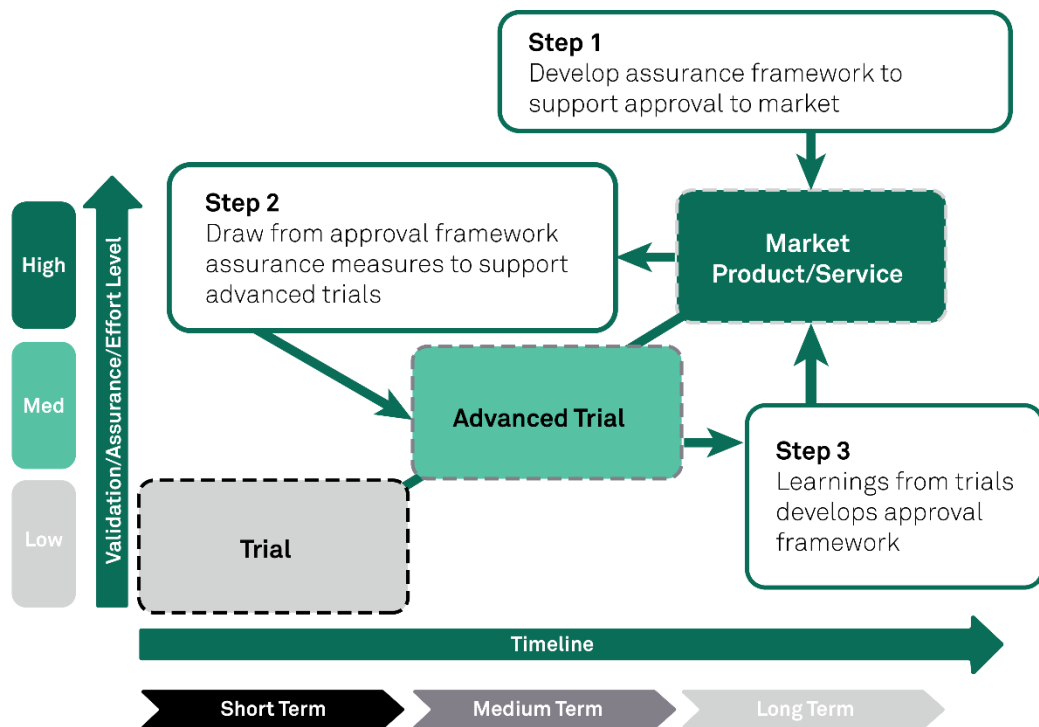
From trials to deployment - a process of learning approach

The intention of the CertiCAV project and this document is to take the first step towards developing a framework that may form the basis of an assurance scheme for approving the release of HAVs and their services to market.

Many of the considerations covered within this document could potentially be applied to advanced trials of HAVs, prior to market release. However, it would be necessary to clarify the definition of an advanced trial before the CertiCAV framework context to advanced trials could be defined (see Section 5.4.1).

The advantage of this approach is that applying elements of the longer-term system of approval to advanced trials will generate learnings that will further develop the assurance framework, effectively validating and continually improving the system of assurance. Figure 1 shows how a process of continual learning may be implemented.

Figure 1: CertiCAV assurance framework approach



1.3 Content

This report is organised into sections as summarised in **Table 2** below.

Table 2: Summary of report contents

Section	Title	Purpose
1	Introduction	Introduce the document.
2	Context and Assumptions	Define the common groundwork of context and assumptions made by the document. Includes setting out the use-case for this document, the scope & objectives that the assurance framework intends to fulfil and the specifications & definitions of the system it is applied to and assurance goals.
3	Common Assurance Practice	Set out the current state of HAV assurance practice. Identify key gaps, and present approaches for addressing them. Key topics include performance indicators and regulatory performance requirements.
4	System Development	Identify objectives and principles intended to provide guidance for HAV system development to cover assurance goals.
5	System Approval	Discuss a process of implementing practice for HAV system approval.
Appendix A	Workshop & Expert Consultations	Summarises information collected from industry engagement and experts.
Appendix B	State of the Art	Summarises key knowledge and findings from relevant state of the art.
Appendix C	Example HASS specification	Presents an outline example of a HASS specification for a parking valet deployment

2 Context and Assumptions

This section discusses the context in which the framework is applied and the definition of the system it is assuring.

The roles and activities of various stakeholders involved in HASS deployment are identified. This highlights the need for ADSEs to develop and provide a "HASS specification". The HASS specification defines and enables the activities of the other stakeholders associated with the HASS deployment.

2.1 CertiCAV's use-case

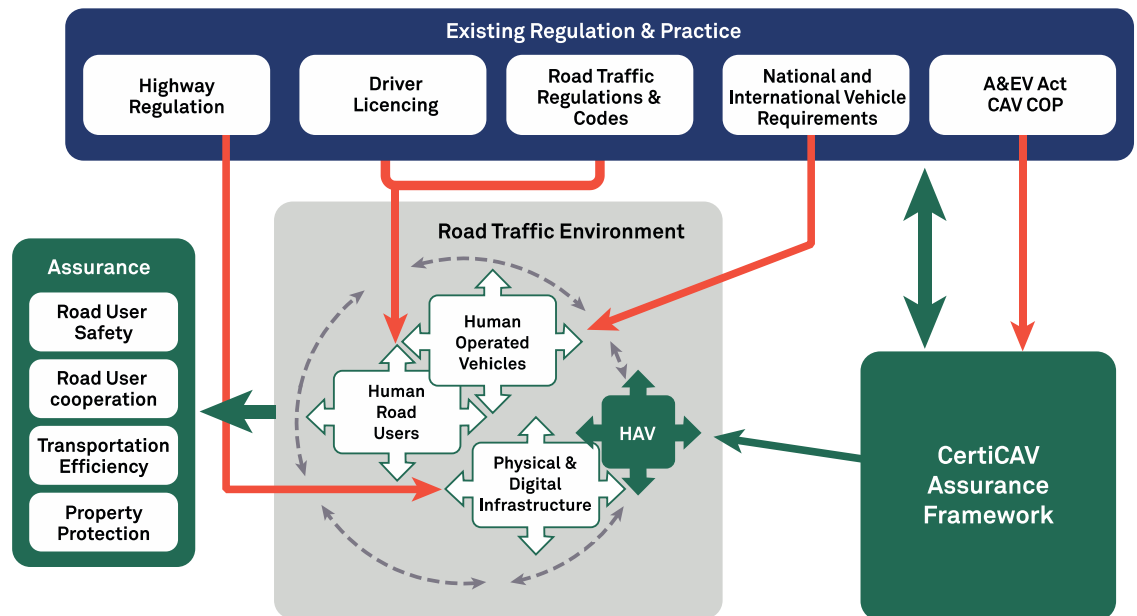
This document describes a framework intended to cover the assurance of perception and tactical decision making undertaken by an ADS integrated into a HASS, where the ADS and HASS are performing nominally. The assurance relates to the intended unmitigated on-road deployment of the HASS where it may encounter scenarios related to the intended on-road deployment. There are several organisations involved in supporting this deployment who are considered in our discussion.

2.1.1

Position of CertiCAV in overall system

Figure 2 below illustrates the positioning of CertiCAV within the wider operational environment for a HAV deployment.

Figure 2: Positioning of the CertiCAV Assurance Framework in a whole system



The road traffic environment is a complex multi-agent environment where currently a range of human traffic actors including pedestrians, cyclists, and motor vehicles interact with each other, and with physical & digital infrastructure.

In this context, 'assurance' relates to managing the operation within the road traffic environment toward achieving outcomes desirable for society. The primary assurance concern is safety, such as minimising the rate of road casualties. There are however other assurance concerns such as minimising conflicts between users, facilitating smooth and efficient traffic flow, and reducing property damage.

This is managed by a variety of regulations that define organisational and performance requirements, and practice that guides processes and responsibilities.

The place and role of CertiCAV is to define an assurance framework that enables the introduction of HAVs into the road traffic environment. This involves replacing the role performed by human licenced and insured motor vehicle drivers with automated systems which has wide-reaching consequences. This assurance framework for HAV has an input and output relationship with the existing regulatory and technical framework. It draws input from these such as the Automated & Electric Vehicle act, the Code of Practice for Automated Vehicle Trialling, the series of BSI publicly available specifications (PAS) for CAV, and existing regulations & codes managing the road traffic environment. It outputs feedback including identified gaps of needs to be addressed by developments to practice and regulations, and outlines approaches to address these. Questions that the framework needs to address include:

- How might the processes of driver licencing, and road traffic codes translate to best practices for HAV approvals?
- How might the practice of vehicle type approval account for approval of HAV?
- How might organisational and performance requirements be defined to provide assurance of the behaviour of HAVs introduced into the existing road traffic environment?

2.1.2 Context of assurance

2.1.2.1 Assurance vs Safety Assurance

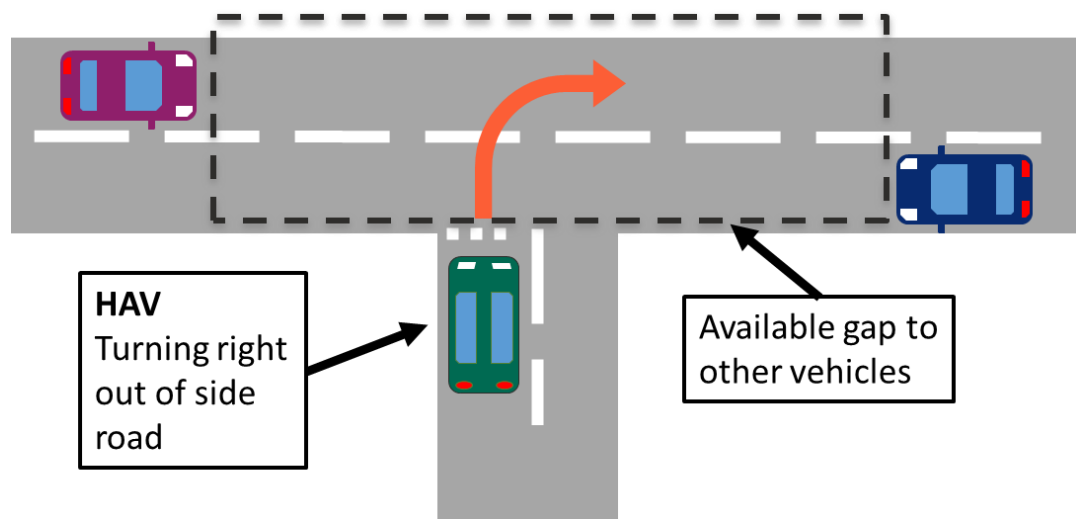
CertiCAV's primary assurance focus is safety, such as minimising the rate of road casualties. There are however other relevant assurance concerns such as minimising conflicts between users, facilitating smooth and efficient traffic flow, and reducing property damage.

Accordingly, CertiCAV refers to covering 'assurance', the term 'safety assurance' is used only when referring to the subset of assurance that is limited to safety concerns.

The distinction between assurance and safety assurance is particularly relevant as assuring acceptable HAV performance may involve trading off acceptance of some safety risk to achieve acceptable performance in other areas. For example, considering the scenario of a HAV turning right from a minor road into a major road at a priority T-junction shown in **Figure 3**:

- Considering safety assurance: With concerns limited to safety there are no non-safety cost factors under consideration to justify an increase in safety risk. Under this system the HAV, before initiating the manoeuvre, would always wait until there are sufficient gaps to ensure that oncoming traffic would not need to adjust their speed due to the HAV's manoeuvre and minimize safety risks.
- Considering assurance: With concerns expanded to include smooth and efficient traffic flow, a small safety risk appetite may be justified to reduce the cost of traffic disruption. Under this system, if traffic is heavy it may be justifiable to initiate the manoeuvre when gaps may require oncoming traffic to gently slow down, if waiting for longer gaps would significantly inconvenience traffic queued behind the HAV. However, challenges this justification must address includes the situation specific safety risk is not unacceptably high, and that the assumed risk does not outweigh the efficiency benefit.

Figure 3: HAV turning out of side road



This distinction also reflects the need to migrate from a traditional risk identification and control method of working, to a more holistic approach that is needed for complex systems such as HAVs. The NASA Aerospace Safety Advisory Panel (ASAP) advocated the adoption of such a holistic approach, stating that:

"... the ASAP applauds switching to a performance-based approach because it emphasizes early risk identification to guide designs, thus enabling creative design approaches that might be more efficient, safer, or both."¹¹

2.1.2.2

Areas of Assurance and safety

Existing best practice defines areas of assurance and safety in the context of automated vehicles; for example, PAS 1881¹² states that:

- Safety assurance for automated vehicles can be categorised into two interdependent areas: system safety and operational safety
- System safety is achieved through ensuring adequate functional safety (FuSa), Safety of the Intended Functionality (SOTIF) and cybersecurity. This forms an integral part of the vehicle development and includes the vehicle specification, design, implementation, and verification and validation of the automated vehicle's functions
- Operational safety assurance considers the interaction of an automated vehicle with the operating environment, including the route, safety driver or operator, passengers and other road users and road workers

However, in conducting the work and analysis of CertiCAV we have found conflicts with some of these terms. To avoid confusion, in the context of reading CertiCAV, we clarify that:

- The scope of current best practice for FuSa, SOTIF, and cybersecurity does not fully cover system safety. For instance, this report identifies that behavioural safety is an important consideration for HAV safety (see Section 3.4.5)
- System safety is not necessarily limited to the development and specification of the vehicle. The performance of a typical HAV involves a wider system architecture including GNSS, V2X, remote command/monitoring systems and over the air datasets such as mapping. Accordingly, this document refers to the wider system relevant for safe performance as the HASS (see Section 2.3.3)
- In addition to the interaction of an automated vehicle with the operating environment, wider operational processes could be considered part of operational safety. These may include for example monitoring and management of vehicle performance, operational environment, and vehicle maintenance
- It is challenging to create a clean divide between system and operational safety, some overlap may be inevitable. For example, SOTIF and ODD relate to the operational environment, and hence operational processes for monitoring and maintaining this environment.

¹¹ (NASA, 2011)

¹² (The British Standards Institution, 2020)

For the purpose of clarification, CertiCAV based on its scope (see Section 1.2) is intended to cover the assurance of:

- HASS performance, with respect to the nominal performance of an ADS's perception & tactical decision making. The validation of which includes determining that:
 - The HAV and its interaction with its environment performs acceptably in nominal conditions
 - The suitable detection and HAV response to the occurrence of non-nominal conditions

The CertiCAV framework suggests that the tolerance for nominal conditions is specified in a HASS specification (see Section 2.4). This could include, for example:

- ODD boundary specifications, such as a tolerance of sensory conditions that is nominal
- Specification or limitation on the behaviour of other actors, such as that they conform to certain road traffic rules or within defined bounds of reasonable/non-adversarial behaviour.

However, an aspect to be managed is that the specification is reasonable, for example that it does not exclude conditions that occur relatively frequently in operation and trigger a HAV response that is disruptive.

CertiCAV does not intend to directly address other areas, such as functional safety, operational procedures, or through-life issues (see Section 1.2). These may be discussed in part where they interact with areas being covered. However, some overlap gaps may prevail due to the incomplete coverage of CertiCAV. For example, considering the detection and response to the occurrence of non-nominal conditions, this may occur due to:

- Reaching the limits of the current ODD
 - This predominately relates to ADS perception and decision making which is in CertiCAV's scope and therefore, predominately concerns issues relating to the involvement of operational procedures or wider system architecture to monitor and communicate ODD factors
- Performance loss of the ADS systems or HAV actuators
 - This predominately relates to functional safety which is out of CertiCAV's scope. Although high-level readily identifiable matters may be discussed, issues involving a detailed consideration of functional safety are not covered. For example, CertiCAV discussion may identify that an ADS should detect performance loss and respond in an appropriate way including execution of MRM to achieve MRC. However, CertiCAV does not attempt to cover the identification and consideration of different types of performance loss and how to ensure MRM execution handles the perception and controllability impacts pertaining to each type of performance loss

2.1.3 Context of on-road HAV deployment

2.1.3.1 Definition of on-road HAV deployment

In the context of CertiCAV on-road HAV deployment refers to the use of a HAV, in its automated mode, on public roads. The deployment may be a trial or market use. The definition of HAV is provided in Section 2.3.1, for the sake of explicating the systems involved in deployment the HAV may be considered to be part of a HASS as described in Section 2.3.3.

2.1.3.2 Assumed regulatory process

The process of releasing an on-road HAV deployment operation for market use is assumed to be in three phases, as illustrated in **Figure 4**. To clarify, all three phases described in this section are a development of releasing an on-road HAV deployment operation for market use, none are for trials.

In phase 1 the ADSE develops the HASS. This is presented by the ADSE to the Regulator and/or Type Approval Authority (TAA) for system approval which if successful enables progression to the next phase. The system approval would have a focus on the system safety case.

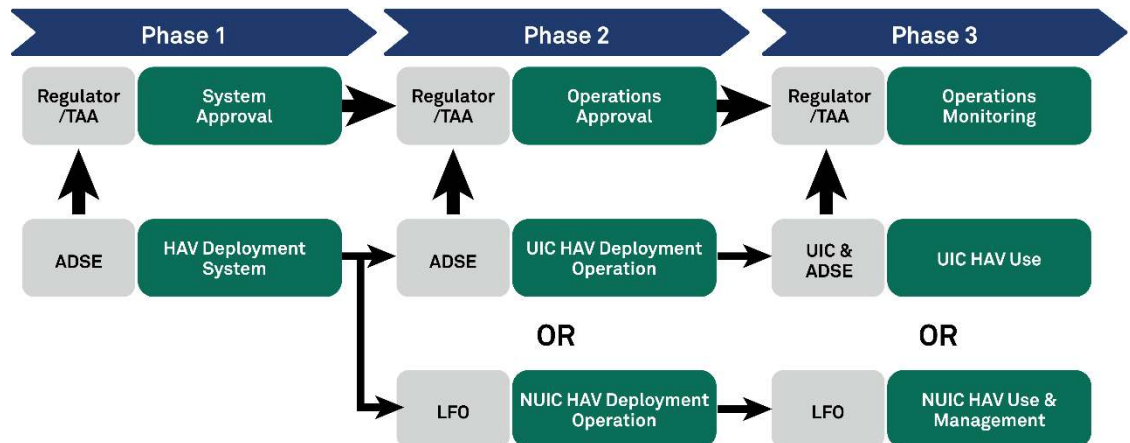
In phase 2 the deployment operation of the HASS would be developed. This is presented to the Regulator/TAA for operations approval which if successful enables progression to the next phase. If the deployment involves a HAV with a UIC then the deployment operation would be developed and presented by the ADSE, alternatively if it involves a NUIC HAV then this would be developed and presented by an LFO.

In phase 3 the HAV deployment operation would be released for market use. The operator would be responsible for the use of the HAV. If the use involves a HAV with a UIC then these responsibilities would be divided between the UIC (e.g. for appropriate use) and the ADSE (e.g. for through life software maintenance), otherwise if the use involves a NUIC HAV then an LFO would be responsible. During phase three the Regulator/TAA would monitor operations and the operators may have duties such as reporting incidents, providing in-use data and/or responding to advisory notices.

Although these phases are presented in a linear progression, various aspects of development may need to be progressed in a different sequence for practical reasons. For instance:

- An LFO may need to develop their deployment operation to derive their operational needs and provide these to the ADSE as requirements for the system development
- Accordingly, the LFO may need to check with the Regulator/TAA that the deployment operation is acceptable in principle in order to reduce the commercial risk that the system/operations that the ADSE/LFO are investing in the development of cannot be brought to market due to an issue in achieving 'Operations Approval'
- The approval/regulatory process in addition to shaping the technical needs of the system and operations, may also shape organisational needs. For instance, the needs placed on how an ADSE and LFO manage their relationship with each other and their supply chains depends on how regulatory requirements define organisational responsibilities, liabilities, and provision of access/evidence

Figure 4: Assumed regulatory process of HAV deployment to market



CertiCAV based on its scope is therefore focussed on covering phase 1 which involves the assurance of system safety aspects of the HASS. It is not intended for CertiCAV to cover phase 2 or phase 3. However due to the interdependent relationship between system safety and operational safety some areas of operational safety are discussed in this document at a high level.

2.1.3.3

Mitigated vs unmitigated deployment

During a HAV trial deployment it is normally expected for safety mitigations to be in place in order to comply with laws, and as part of risk management due to the incomplete safety validation of the HASS. In fact, the code of practice published by UK Government on 6 February 2019¹³ conveys the legal requirement is that 'A driver is present, in or out of the vehicle, who is ready, able, and willing to resume control of the vehicle'.

During such a trial the deployed vehicle system is not being operated according to the definition of a HAV (see Section 2.3.1) due to the involvement of a human in monitoring the dynamic driving task, and/or fallback operation and/or other driving task functionality.

We term this as a 'mitigated deployment'.

The aim of the system and operations approvals shown in **Figure 4** is to facilitate a HAV deployment that does not require human involvement in the driving task or being receptive to requests to intervene. We term this as an 'unmitigated deployment'.

¹³ (Centre for Connected and Autonomous Vehicles, 2019)

2.2 Stakeholder roles, responsibilities and needs

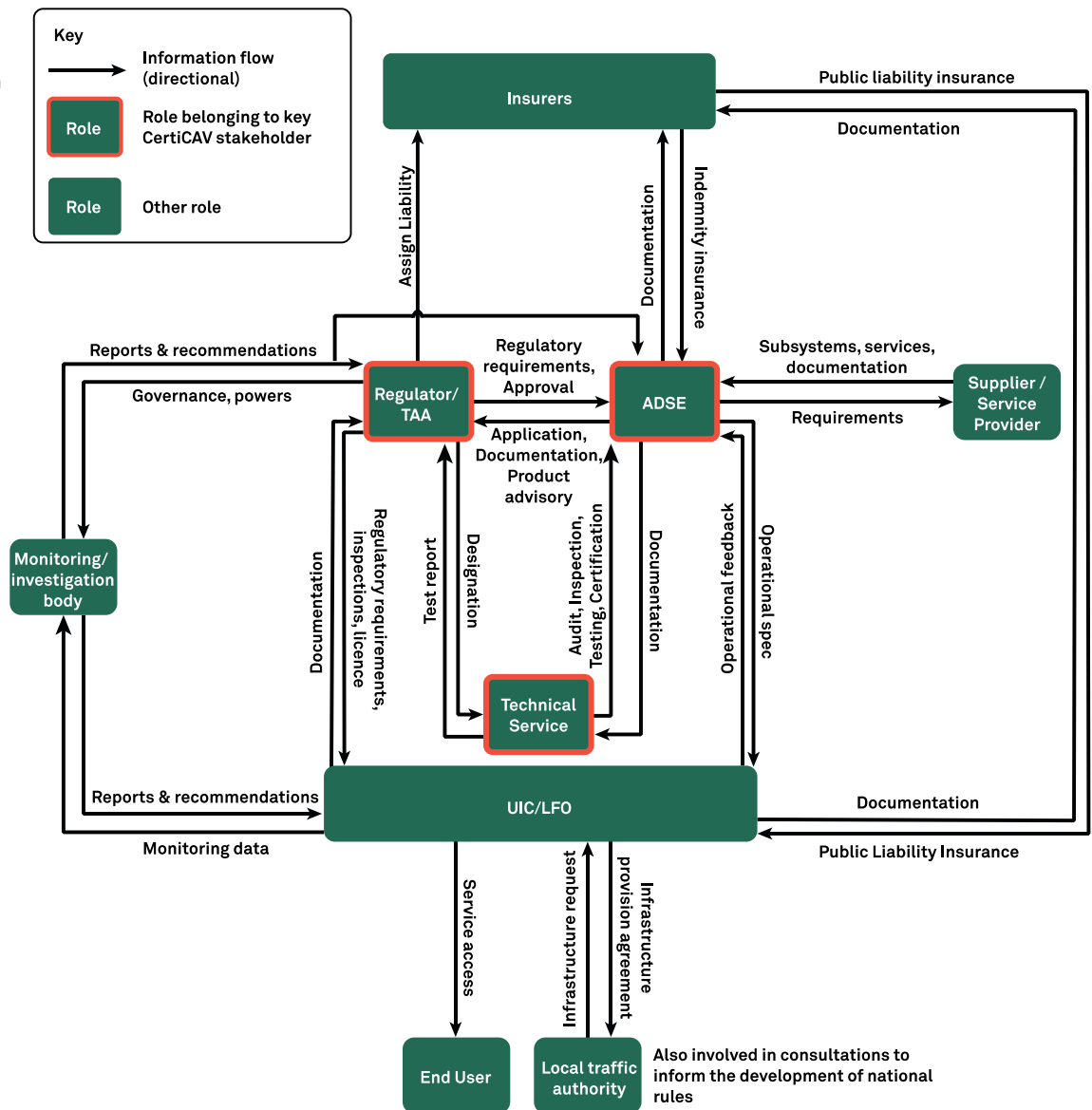
A HAV deployment would be supported by an ecosystem of stakeholders, several of whom have some involvement with assurance. For the purpose of this document we have forecasted what the HAV organisational ecosystem may become, based on existing defined practice and proposals, and pre-supposing deployment is for a HAV. Primary references are the BSI CAV vocabulary¹⁴ and the Law Commissions' automated vehicles consultations (CP1, CP2, and CP3) (see Section 8.5). This has resulted in identifying the Stakeholders described in **Table 3** with an organisational relationship as shown in **Figure 5**.

Table 3: Assumed Stakeholders in HAV deployment Ecosystem

Stakeholder	Description / Role
ADSE	The Automated Driving System Entity (ADSE) is the organisation or individual that presents a HASS for approval for use and is responsible for its system safety, relevant non-safety performance, and through life support, including co-ordination with the UIC / LFO and/or service providers involved in the deployment operation of the HASS. This is a development of the pre-existing definition of ADSE, as discussed in the Terminology . An ADSE may therefore be considered like the term 'manufacturer' referred to in existing vehicle type approval regulations. However, an ADSE has additional responsibilities for managing safety for the whole safety lifecycle, and co-ordination with vehicle/service operators to ensure the intended functionality is suitable for the intended use.
Supplier / service provider	Organisations involved in HAV deployment system through the supply of products and/or services to the ADSE, UIC / LFO or other entities.
Regulator / Type Approval Authority (TAA)	Organisations involved in regulating and/or technical approvals of the HAV deployment system.
Technical service	Organisations providing independent 3rd party testing, and inspection services that support the process of technical approval.
Insurers	Organisations providing Insurance products.
UIC / LFO	The entity assumed to be responsible for non-DDT related operation of the HAV.
Traffic authority	Local and/or regional organisations responsible for providing, operating, and maintaining road traffic infrastructure and regulating local transport services.
Monitoring / investigation body	The organisation(s) that undertake(s) governance monitoring of HAV deployments and/or investigating incidents. (For context only, monitoring and investigation is outside CertiCAV's scope)
End user	The end user of the HAV or HAV enabled service

¹⁴ (The British Standards Institution, 2020)

Figure 5: HAV
Deployment Ecosystem



We have undertaken a RASIC analysis to outline the roles and needs of the stakeholders in relation to HAV deployment. Firstly, in **Table 4** we identify and describe required tasks for supporting a HAV deployment. Then in **Table 5** each party's roles for undertaking tasks is allocated within the following categories:

- Responsible
- Approver
- Supporting
- Informed
- Consulted

In these tables, it should be noted that:

- Tasks that are out of scope for this report are shaded grey and not considered in detail.
- Currently market surveillance is undertaken by the TAA, and there is not a separate monitoring / investigation body organisation. Therefore, this function is considered part of the Regulator/TAA role
- The role of end user has been omitted for clarity

Table 4: Description of tasks relevant for HAV deployment management

Ref.	Task	Description
1	Produce HASS specification	Develop a HASS specification for the proposed HAV deployment.
2	Production quality and safety management	Develop and implement quality assurance system to ensure quality and safety control throughout system development. Including accreditation for relevant processes/items and a system safety case that argues the satisfaction of relevant requirements (including those defined in regulation/codes), supported by evidence that demonstrates the validity of the safety case.
3	System approval	Approving the HASS for use as defined by regulations and codes. This may include approval/ acceptance of relevant processes and items. Corresponds to phase 1 approval discussed in Section 2.1.3.2
4	Operations quality and safety management	Develop and implement quality assurance system to ensure quality and safety control throughout HASS operations and lifecycle. Including an operational safety case that argues the satisfaction of relevant requirements and is supported by justifying evidence.
5	Operations approval	Approving the LFO and/or deployment of HAV on roads as applicable. Corresponds to phase 2 approval discussed in Section 2.1.3.2
6	Operations safety manual	Produce guidance for operators on how to use the HAV safely.
7	Through life safety management	Provide safety-related updates to HAV throughout lifecycle.
8	Supply chain management	Specify requirements for subsystems/ services (e.g. map data).
9	Supply provision	Produce/ provide subsystems/ services in line with requirements.
10	Operations management	Run operations in line with safety case / ADSE guidance / operational safety standards.
11	Authorisation management	Decide who can use the HAV.
12	Deployment monitoring	Provide operational feedback/ data to ADSE and monitoring/ investigation body.

Table 5: RASIC identification for stakeholders involved in HAV deployment management.

Item	Regulator / TAA	ADSE	UIC / LFO	Technical Service	Supplier/ service provider	Insurer	Traffic Authority
Produce deployment specification	C	R	S		S	I	C
Production quality and safety management	I	R,A	C	I	S	I	
System approval	R,A	S	S	S	S	I	C
Operations quality and safety management	I	S(NUIC) R(UIC)	R(LFO)	I	S	I	
Operations approval	R,A,C	S	S(LFO)	S	I	I	A,C
Operations safety manual	A	R	C(LFO) I(UIC)			I	C
Through life safety management	A	S	R			I	S
Supply chain management		R,A	S		S	I	
Supply provision		A	A		R	S	
Operations management	I		R				
Authorisation management	A		R			C	
Deployment monitoring	A	S	R	S	S	I	C

Therefore assurance, for the context of the CertiCAV framework primarily involves the activities of three key stakeholders:

- Regulator / TAA: Undertaking due diligence that the HASS is sufficiently safe
- ADSE: Who is responsible for the HASS development
- Technical Service: Who supports technical approval of the system with impartial Testing, and Inspection services

However, assurance does involve the activities of other stakeholders who are involved in the HAV deployment by virtue of supporting the key stakeholders.

2.3 Common system definition

A common definition of a minimum number of sub-systems is required to facilitate a sufficiently explained and commonly understood system amongst stakeholders. The following common system definition is adopted for the CertiCAV assurance framework.

2.3.1 Highly Automated Vehicle (HAV)

A Highly Automated Vehicle (HAV) is defined in SAE J3016¹⁵ as a vehicle equipped with an automated driving system that operates within a specific operational design domain for some, or all, of the journey, without requiring human involvement in the driving task or being receptive to requests to intervene. SAE J3016 defines 6 levels of driving automation ranging from Level 0 to Level 5, on this scale a HAV is level 4. This document does not specifically consider level 5 automation; however, it may generally be assumed that the same principles apply to this highest level of automation with the exception that it would not be limited to operating within a defined ODD.

For the purpose of scope clarification, in this document a HAV may be:

- A passenger car
- Vehicles designed to operate on public roads that are not designed to carry people, e.g. delivery robot. Specifications/regulations are not established for this vehicle type, however for the purpose of this framework it is assumed they are designed and constructed to be reasonably conspicuous to other road users including their physical dimensions, lighting, indicator lamps, and horn
- Goods or passenger service vehicles with a maximum mass not exceeding 3.5 tonnes

We have not considered other types of vehicles in the development of this document. Although much of CertiCAV could be applied to vehicle types not in this list, the applicability and accuracy/completeness of content would need to be considered accordingly. For example, if the vehicle is designed to operate off road (e.g. footway/footpath) then CertiCAV content involving road traffic rules may be inaccurate or not applicable.

2.3.2 Automated Driving System (ADS)

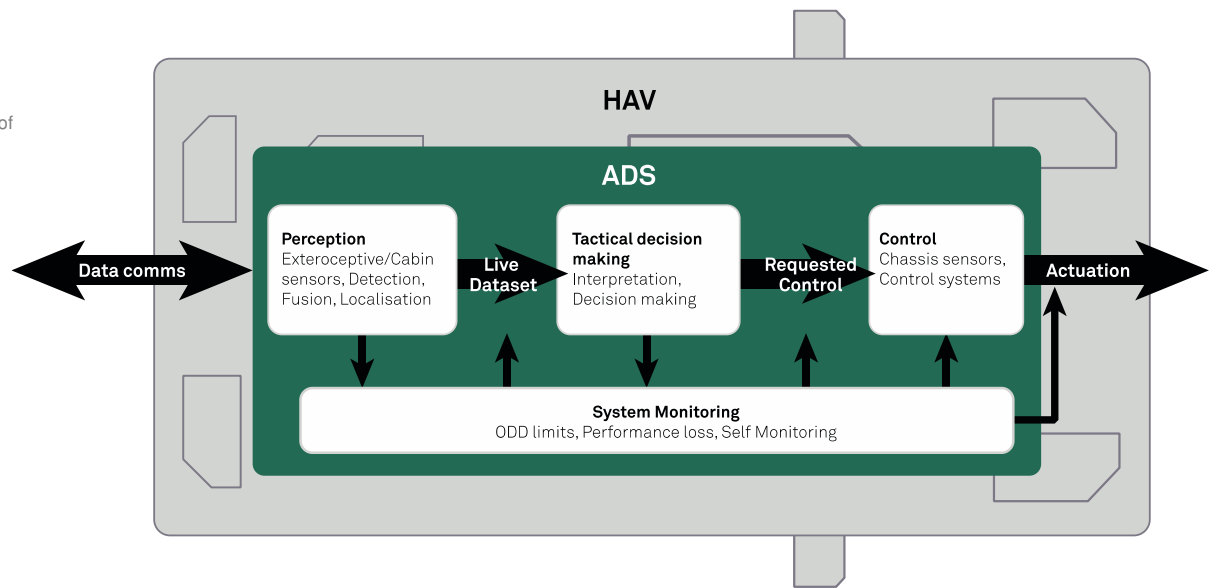
This document adopts the J3016 definition of an ADS which is "the collection of hardware and software that are collectively capable of performing the entire DDT on a sustained basis."

To support the development of the assurance framework we've defined a common sub-system architecture of an ADS as illustrated in **Figure 6** below. The purpose of this definition is to:

- Provide a common definition of ADS sub-systems for which there is a need to refer to in the definition of activities discussed in the assurance framework, whilst trying to keep the definition as minimal and simple as practicable
- Align where practicable with architectures defined in existing references, including PAS 1880

¹⁵ (SAE International, 2018)

Figure 6: CertiCAV common definition of ADS sub-systems



An overview of each of these defined common subsystems is given in the following subsections.

2.3.2.1 Perception

The collection of on-board systems that derives tactical situational awareness which is output to the live dataset. This includes sensing of the surrounding and cabin environment and undertaking object & event recognition and fusion processes. Perception includes the following subsystems:

- Exteroceptive sensors: Systems that directly collect data on the environment around the HAV
- Cabin sensors: Systems that monitor loads, passengers, operator/driver, and vehicles systems they interact with such as passenger doors or request buttons; as applicable
- Detection/fusion: Systems that process input data to detect objects/events and fuse data from multiple sources into a rationalised dataset. This may include the fusion of data received via data communications.
- Localisation: Systems that determines the position of the HAV within the environment

Note that *chassis sensors* are discussed in Section 2.3.2.5.

2.3.2.2 Live dataset

The set of live data that is input to the decision-making system. This includes a representation of objects around the HAV output from perception, and any other data input into tactical decision making including strategic routing, remote operator control, remote ODD monitoring data, occupant/cargo state, vehicle state such as vehicle dynamics and/or events/errors. The live dataset therefore encompasses any live (non-configuration) data output from systems that affects tactical decision making, including via HMI, data communications, and/or relevant system monitoring output are part of the live dataset.

2.3.2.3 Tactical decision making

The collection of systems that makes interpretations and decisions to determine requested control output based on two inputs:

- The live dataset, as described above
- Configuration dataset, any data configured onto the vehicle that influences tactical decisions including calibrations, and mapping

Tactical decision making may be considered to include all open loop control required to execute the DDT. This therefore includes all interpretation and decisions associated with real-time manoeuvre/path planning but does not include strategic routing decisions that are not part of the DDT, or closed loop feedback control that is part of control systems. Depending on the intended ADS functionality the interpretations may include predictive and/or probabilistic elements.

2.3.2.4 Requested control

The set of control requests that is input into the control systems from the Tactical Decision-Making systems.

This may include the four categories of control:

- a) Longitudinal motion control (acceleration, braking)
- b) Lateral motion control (steering)
- c) Statutory lighting and signal control (Head/taillights, Indicators, Horn)
- d) Control of other actuators (e.g. Windscreen wipers, doors, sensor servos)

2.3.2.5 Control

The collection of systems that outputs commands to the HAV's actuators based on the requested control input from tactical decision making and feedback from the chassis sensors.

This includes four categories of actuation:

- a) Longitudinal (acceleration, braking)
- b) Lateral (steering)
- c) Statutory lighting and signals (Head/taillights, Indicators, Horn)
- d) Other actuation (e.g. Windscreen wipers, doors, sensor servos)

Chassis sensors are systems that collect data on the physical status and performance of the HAV and its systems. For example measuring: acceleration, angle position, wheel tick, torque, or tyre pressure.

The implemented control commands may differ from the requested control input due to the influence of the control systems. The control systems may influence the implemented actuation based on feedback from chassis sensors to maintain HAV controllability or reflect physical limits of actuators. This is particularly true for implementation of motion controls (categories a and b) where it is typical to implement closed loop control systems to improve vehicle controllability and stability, for example anti-lock braking system (ABS) to maintain tyre traction during use of braking actuators.

Furthermore, the implemented actuation may differ from the control command due to the operation of actuators or the influence of physical dynamics.

2.3.2.6 Actuation

The physical systems that implement control actuations e.g. powertrain, brake servos, steering servos.

2.3.2.7 Data communications

Any wireless data communication with off board systems, typically through an On-board Unit (OBU) which provides a communication interface between external systems and ADS systems.

Functions may include:

- Strategic route planning
- GNSS
- V2X
- Remote safety operator / Fleet Management e.g. strategic route planning
- Mapping and Waypoint Datasets
- OTA updates

Outputs may be communications with nomadic on-board systems (for example a mobile phone) or off-board systems. For example, V2X co-operative awareness messages, status updates to FMS.

Data communications could potentially have inputs/output with any of the ADS sub-systems and the role of data communications has highly significant impacts on the definition of the ADS architecture.

For example:

- V2X data could provide an input into perception. However, how this data is handled would need to consider the safety and security implications associated with data received over the air from 3rd party sources
- Mapping datasets could be provided over the air rather than being part of ADS configuration data. In which case the potential impact on versioning and behaviour repeatability would need to be accounted for in validation and testing
- Data relevant for system monitoring, for example forecast weather
- Remote operator control/monitoring could directly input and/or output to the live dataset, tactical decision making, and/or requested control. This would involve a change in automation mode, and the implementation would need to consider safety/security implications including over the air communications and data transmission through 3rd party systems if applicable

The involvement of off-vehicle systems is particularly relevant if they form part of safety critical functions.

2.3.2.8 System monitoring

The collection of systems that monitors:

- The performance of the ADS systems (including self-monitoring) and HAV actuators
- That the HAV is operating within the limits of the current ODD

If system monitoring detects performance loss of the ADS systems or HAV actuators, or that the HAV has unexpectedly reached the limits of the current ODD, then it responds by:

- Switching to an alternative/degraded driving function/mode validated to operate appropriately for these conditions, which could be associated with a different ODD; or
- Outputting a request to the tactical decision-making system to implement an MRM. Or if this is not possible due to performance loss of tactical decision-making systems; then
- Directly implementing an MRM by inputting tactical directions to the control system. Or if this is not possible due to performance loss of control systems; then
- Directly actuating an emergency stop

This may be associated with event logging and/or communications to/alerting of: fallback capable operator, occupant, fleet management/remote systems, and/or monitoring authority.

2.3.2.9 HAV behaviour

HAV behaviour is not an item or subsystem of the ADS, rather is the general term given to any description of a HAV's actions.

For example, based on the actuation of applying the brakes the HAV behaviour would be that the HAV slows down.

A HAV manoeuvre is a semantic description given to a set HAV behaviour that may be triggered by a tactical driving decision e.g. a "lane change". This may be a useful structure for formally relating driving rules to HAV behaviour (see **Section 3.4.5.1**) and defining/generating scenarios (see **Section 3.4.9**)

2.3.2.10 Other considerations

This definition does not include some items that may be applicable to a wider context out of scope for this document such as:

- **Vehicle State System:** The system that performs event / state recognition of the HAV
- **HMI:** Systems providing an interface between humans and machines which in a broad sense may include interfaces with on-board operators/occupants, remote operators/users, and external road users
- **Supervisor:** The system interfacing data between the ADS and external systems including, routing instructions / driving requests from FMS/HMI, providing data to logging/monitoring systems. This may include, for example the logging of debug and performance data for development, validation, and/or monitoring purposes. Which may involve incidents detected by the monitoring system, or other detected events such as near miss, system performance events.

2.3.3 Highly Automated SuperSystem (HASS)

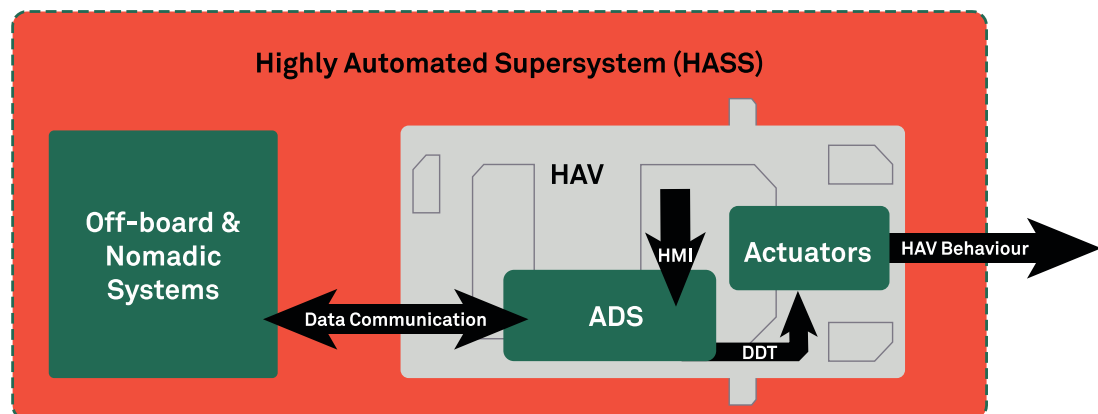
The Highly Automated SuperSystem (HASS) is defined as the collection of systems involved in the performance of HAV and its Behaviour. The HASS is illustrated in **Figure 7** and includes:

- The HAV (see Section 2.3.1) which includes:
 - Actuators that execute the DDT output of the ADS which results in performing HAV behaviour
 - HMI systems that provide an interface between on-board occupants/operators and the ADS
- The ADS integrated into the HAV (see Section 2.3.2) which includes:
 - Off-board and Nomadic Systems that may influence the ADS's performance of DDT through data communications

As the definition of ADS is "the collection of hardware and software that are collectively capable of performing the entire DDT on a sustained basis" it may be presumed that off-board/nomadic systems that influence the ADS's performance of the DDT are part of the ADS. However, we have specifically stated their involvement for the following reasons:

- To make it explicit that with the CertiCAV definition of ADS including data communication as an input (see Section 2.3.2.7) the collection of systems involved in performance of DDT are not necessarily limited to systems on-board the HAV. Off-board systems are potentially involved in the performance of DDT and therefore influence HAV behaviour
- The developer of a particular HASS could argue that their ADS is limited to on-board systems due to being capable of performing the entire DDT on a sustained basis with data communications being unavailable. However, the performance of DDT may still be influenced through data communications when they are available which may have impacts on safety or other assurance concerns

Figure 7: Illustration of a Highly Automated Supersystem (HASS)



Some of the functions of off-board & nomadic systems are listed in Section 2.3.2.7, including GNSS, remote safety operator, and V2X. For the context of CertiCAV it is assumed that principles are applied to limit the role of functions linked with the ADS wirelessly for the purpose of system/functional safety including:

- That the HAV is able to perform safely, including fallback operation, without reliance on, or availability of any wirelessly linked systems. For example:
 - Fallback operation is activated and performed by on-board systems and does not rely on active monitoring and/or intervention from a remote safety operator
- That safe operation of the HAV does not rely on correct function of any wirelessly linked systems. For example, that:
 - GNSS data being inaccurate or unavailable doesn't impact HAV behaviour in a way that causes unacceptable risk
 - Remote operated driving is limited to requesting driving paths, rather than overriding on-board collision avoidance functions
 - A 'Traffic Light Maneuver service'¹⁶ over V2X is not relied on as the primary means of detecting red traffic light signals. Although using as a secondary sensor to parity check detection from on-board vision sensors could enhance safety

Due to operational and functional safety being out of scope for CertiCAV (see Section 1.2) this document does not cover the management of HMI or actuators.

2.4 HASS specification

The HAV deployment use case set out in Section 2.1 is facilitated by several organisational stakeholders involved as described in Section 2.2. A specification of the HASS involved would need to be developed and provided to stakeholders so they can facilitate the HAV deployment. The following relevant stakeholder assurance activities would need to be supported by a HASS specification:

- ADSE: Argue and justify the satisfaction of regulatory requirements, provide assurance of quality and performance. Clarify the specification of nominal conditions.
- UIC/ LFO: Clarify the specification of the system is suitable for their proposed use, in the case of an LFO - put in place appropriate operational procedures
- Regulator/TAA: Clarify system approval activities
- Technical Service: Preparation and undertaking of independent testing, and inspection activities in support of technical approval process for the HAV deployment systems
- Insurer: Clarify risk factors, quality assurance and safety monitoring activities to facilitate provision of insurance services required for deployment
- Service Provider: Clarify their duties and responsibilities in supporting the HAV deployment system

¹⁶ (ETSI, 2018)

The subsections below outline identified deployment specification objectives and attributes that would need to be defined and shared with the stakeholders to allow them to undertake these stakeholder assurance activities. These are organised into the following five functional categories:

- Vehicle Operational characteristics
- Operational Design Domains (ODDs)
- ADS functionality
- HASS Architecture
- Deployment Management

Furthermore, the definition and provision of a HASS specification supports the overarching objective of cross referencing that the specifications for each functional category are appropriate and compatible with each other. For example, the ADS functionality should define a sufficiently high operational speed if the ODD includes high speed roads.

It should be noted that this content focusses on supporting objectives that fall within the CertiCAV scope of work as defined in Section 12, and does not attempt to cover wider topics such as functional safety, cybersecurity, and through whole lifecycle.

2.4.1 Vehicle operational characteristics

Objective: Define the outline physical and use specifications of the HAV to support:

- Classification of the deployment vehicle/service use of the HAV to facilitate system/operational regulatory processes
- Definition of vehicle dynamic & operational attributes of the HAV relevant for implementation of internal and/or independent assurance assessments.

We have identified that specification of attributes including the following would support the objectives:

- Outline physical attributes: length, width, weight
- Chassis dynamics attributes: Turning circle, steering rate, acceleration, braking
- Vehicle use: occupants/ cargo, Number/ weight, Type/ Restrictions (e.g. age, hazardous), Arrangement (e.g. Standing/ seated/ seat-belted, loose/ secured)
- Deployment domain: A geofence describing the area or route being operated. For some deployments, this may consist of all roads that conform to the ODD within a country.
- Vehicle configuration: Requirements and limitations for vehicle configuration (e.g. permitted tyres, towing attachments.)
- Exteroceptive/ chassis sensors: Type, location, field of view, range

2.4.2 Operational Design Domains (ODDs)

Objective: Define the operating conditions under which the HAV is designed to perform in automated driving mode to support:

- Clarification of boundary conditions under which the HAV will, and will not be operating in automated mode to facilitate vehicle/service operations, regulatory, and oversight processes (e.g. for an incident investigator to verify the environmental conditions were within boundaries of the ODD)
- Definition of nominal operational and boundary condition parameter specifications that the HAV is intended to be operated within, to support implementation of internal and/or independent assurance assessments (e.g. for generation of test scenarios)

ODD relates to the external conditions the HAV is intended to operate in; it excludes internal state/conditions of the HAV.

BSI PAS 1883 provides a definition of ODD terminology and ontology structure with the following attributes top-level concepts and attributes:

- Scenery: Zones, Drivable area, Junctions, Special Structures, Fixed road structures, Temporary road structures
- Environmental Conditions: Weather, Particulates, Illumination, Connectivity
- Dynamic elements: Traffic, Subject Vehicle

However, it is necessary to further develop the ODD taxonomy to provide a common schema format that can be applied to an ODD description shared between key stakeholders, and that satisfies the objectives stated above. It is critical that the ADSE provides a well-defined, and definitive ODD relating to the deployment which describes all environmental restrictions / boundaries relevant for the HAV operations.

ASAM is likely to produce a concept that contributes to these objectives as part of their 'OpenODD' project which started August 2020 and is due for completion July 2021. (see also Section 3.4.8.6)

A challenge for the definition of ODD is that differing levels of abstraction are required for different use cases:

- For ADS development and validation purposes it is ideal to describe ODD by quantified metrics that can be reproduced in test facilities. For example, quantifying sensory conditions with high spatial-resolution and continuously over the electromagnetic spectrum. For operations monitoring, governance, and insurance purposes it is ideal to describe ODD with relatively simple semantic descriptions that can be verified by human observation and measurements from relatively simple telemetry/measurement equipment.

A report produced by CPC and the National Physical Laboratory titled 'Performance Testing for Sensors in Connected and Autonomous Vehicles: Feasibility Study' discusses definition requirements to support sensor testing in detail, and relevant findings are summarised in Section 8.11.

A potential solution to this challenge would be to develop a common standard that satisfies the latter use case of defining the ODD with relatively high level semantic descriptions that can be verified by human observation and telemetry/measurement equipment. An ADSE would then have the flexibility to internally describe their ODD in a low level format that satisfies their development processes and would only need to collate these against the commonly defined higher level semantic descriptions to be able to readily provide the ODD specified to the common standard which would facilitate co-operation with assurance stakeholders.

2.4.3 ADS functionality

Objective: Define the intended functionality of the ADSs Perception, and Tactical Decision making to support:

- Clarification of the on-road behaviour of the HAV checking appropriateness for ODD and traffic performance
- Definition of the ADSs Perception capabilities and intended tactical driving functionality relevant for implementation of internal and/or independent assurance assessments.

We have identified that specification of attributes including the following would support the objectives:

- Perception
 - Object and Event detection specification, capabilities, and limitations. This includes the definition of detection zones internal and external to the HAV and for each zone: the location specification of related sensors and data communication, a flowchart of the fusion/processing applied to sensor data and data communications, the types of object/event detected, and the range/accuracy of detections
 - A list of known objects and events that may reasonably occur while the ADS is performing the DDT in the ODD that the ADS may not detect as objects/events or as a departure of ODD.
 - Localisation specification, capabilities, and limitations. This includes the method and accuracy of localisation and a flowchart of the related processing applied to sensor data and data communications
- Data Communication
 - Role of connected data in functionality of ADS
- Tactical Decision Making
 - A description of manoeuvres that may be implemented by tactical decision making and any relevant restrictive/permissive ODD factors for each manoeuvre. This may refer to a driving code and/or manoeuvre library: What rules are implemented or not, alternative policies for handling unimplemented rules, conditions/reasons where rules may be violated
 - What MRM and fallback operation modes may be executed and their triggering conditions
 - The specification should describe all limitations without gaps. For example, if it is defined that the HAV shall perform unprotected right turns then any environmental, situational, or functional factors that restrict performance of that manoeuvre, such as forward visibility limitations, should be described
- Control & Actuation
 - Lateral motion
 - Specification of turning radius and steering rates
 - Longitudinal motion
 - Desired acceleration rates/limits, modes
 - Non-motion
 - Controls actuated by ADS and their operational policy
 - Any vehicle chassis actuators that are NOT actuatable by the ADS or implemented in driving policies e.g. horn, lights, parking brake

The definition of intended functionality attributes may be cross referenced against the deployment ODD attributes, for example in more challenging environmental conditions the ADS may reduce operational speed and range of tactical driving manoeuvres.

2.4.4 HASS architecture

Objective: Define the architecture of the HASS, including off-board or nomadic electronic systems that are not integrated into the HAV but may influence the ADS's performance of DDT and therefore HAV behaviour through data communications. To support:

- Definition of which items of the system involve critical functions relevant for implementation internal and/or independent assurance assessments
- Public engagement and authority assessment/acceptance of safety/operational risks
- Clarification of the role and function of wider items in the deployment architecture to facilitate vehicle/service regulatory processes and scope/duty of service providers and insurance

We have identified that specification of attributes including the following would support the objectives:

- Equipment and communications supporting Vehicle/Fleet management systems; including vehicle routing, and monitoring. Certain routing restrictions may be in place to manage safety and capability limitations of ADS functionality
- Global Navigation Satellite Systems (GNSS) or other systems used to help vehicle localisation
- Communications equipment/services such as V2X
- Teleoperation systems
- Mapping / databases
- Local rules of the road

And would at a minimum include the following minimum information for each attribute:

- An identification/name for the system
- A description of how the system is used as part of the deployment
- Safety critical features

A complete description of the deployment architecture (e.g. a component diagram) would strongly support the objective by describing and illustrating to relevant stakeholders how the HASS systems relate and interact with each other.

2.4.5 Operational management

Objective: Define how aspects of the deployment relating to safety in-use will be managed.

To support:

- Definition of what changes may take place to HASS systems during its lifecycle and the relationship with change control processes
- Clarification of ongoing safety/quality monitoring role in safety management
- Definition of any required co-operation activities between stakeholders, for example provision/maintenance of physical/digital infrastructure
- Definition of safety critical human operator procedures, including vehicle operation, monitoring and maintenance. Definition of relevant competence and training procedures
- Establish a robust and sufficiently developed understanding operational and safety aspects of the deployment

We have identified that specification of attributes including the following would support the objectives:

- Entities involved in the deployment and their roles.
- How the vehicle is operated and managed.
- Insurance policy details (e.g. Memorandum of Understanding (MOU) of insurance provision of market product/service when approved)
- Change control management procedures: including HAV and deployment system software/hardware, ODD monitoring, Infrastructure monitoring/maintenance/co-operation, relevant attributes such as version numbers.
- Safety monitoring, incident detection.
- Through life management: E.g. Component inspection, maintenance, replacement schedule.
- Operational conditions assessment: Likelihood / Frequency of ODD boundaries being reached in-use and impact to safety and service performance. For example, for what proportion of time are site sensory conditions outside of ODD bounds, how frequently is fallback engagement expected

Several of these attributes are more relevant for HAV operations, rather than the HASS system, and would under the responsibility of the LFO, rather than the ADSE unless the deployment operation is a UIC HAV. Furthermore, with reference to the assumed regulatory process (see Section 2.13.2) they are relevant for the operations approval (phase 2) which is outside of CertiCAV's scope. However, we have presented this category as some attributes may be relevant for System development and authorisation practice, they represent an area of specification overlap between HAV system and HAV operations which is useful to cover with common specifications to assist co-operation and understanding between organisations.

Several topics relevant for this category are outside of scope of CertiCAV, e.g. incident management/response, occupant communication, emergency information manuals, authority engagement, operations procedures. The discussion in this document may accordingly not cover these factors, or only in a low level of detail.

2.5 Assurance goals

There are two key assurance processes; 'System Development', and 'System Approval'. This section sets out the assumed top level goals for each of these processes that are defined in lieu of context and assumptions defined in the preceding content of this section. These goals are assumed and further detailed in **Section 4**, and **Section 5** respectively.

2.5.1 System development

The ADSE undertaking system development should produce a system safety case proving the satisfaction that all relevant system requirements are satisfied.

This may include demonstrating that the ADS:

- Satisfies the driving performance criteria for all conditions that reasonably occur within the defined ODD
- Can detect when limitations of the ODD are reached and respond by executing an appropriate minimal risk manoeuvre to achieve a minimal risk condition

2.5.2 System approval

As part of supporting the approval of a HAV deployment an independent third party may determine via audit and inspection that:

- The HASS specification is adequately described and suitable
- Documents and processes are well developed, systematic, meet requirements and are traceable
- The safety case is sufficiently developed, provides full coverage of the deployment specification, has sufficient argumentation for coverage of inputs and reduction of unknowns, and satisfies safety requirements
- Evidence is sufficient for supporting safety arguments
- Resources and Tooling have appropriate quality controls
- Suitable practice and standards are applied
- There are suitable organisational safety culture and quality assurance

Furthermore, they may independently verify by evaluation of evidence and/or witnessing of tests that:

- There is sufficient quantity and quality of on-road trials monitoring to argue safety validation and results support
- Testbench testing:
 - Sub-system performance testing and characterisation
 - verification of sufficient behaviour in selected scenarios
 - Boundary / Edge case assessment
 - Volume, controllability/repeatability
- Physical Testing:
 - Verification of perception performance in selected conditions
 - Verification of whole system behaviour in selected scenarios
- Verifying the fidelity of the evaluation toolchain
 - Comparison of testbench and physical testing results
- On road test:
 - Verifying performance and assessing behaviour in real world deployment
 - Assessing deployment management, does it match deployment specification
 - Several tests covering specified sensory/traffic conditions

3 Common Assurance Practice

This section sets out the current state of relevant HAV assurance practice. It identifies key gaps and presents approaches for addressing them. Key topics include performance indicators and regulatory performance requirements.

3.1 Introduction

Assurance is supported by common practice that defines common requirements and processes that promote safe practices and facilitate interoperability between systems and stakeholders. This section outlines relevant existing practice, identifies gaps that should be considered further for automated driving and suggests approaches to fill those gaps. It proposes an approach for setting requirements, and how HAV performance may be measured.

This section discusses three areas:

- **Performance Indicators:** A common reference for performance indicators is beneficial in supporting the definition of requirements, assessment, and management of system safety/quality, and ongoing monitoring (see [Section 3.2](#))
- **Regulatory performance requirements for HAVs:** There is a need to develop regulatory requirements that define the performance standards that HASSs should conform to (see [Section 3.3](#))
- **HAV assurance practice:** There are needs for best practice to be developed for HAV assurance (see [Section 3.4](#))

These areas are complex and currently undergoing significant development internationally. Accordingly, in this document we do not provide a comprehensive system of practice, rather we focus on identifying key issues, gaps and suggesting potential approaches to address them.

3.2 Performance Indicators

Performance Indicators are metrics that quantify the performance of a system to evaluate the level of success, determine achievements, and argue the satisfaction of specifications and needs.

In the context of CertiCAV, performance indicators may measure either the performance of a whole HASS such as operating safely & effectively, or of sub-systems.

They play a significant role in assurance because they:

- Support regulatory requirements with an objective assessment of compliance and traceability to safety argumentation and evidence
- Are applied to Verification and Validation activities and Safety Case argumentation to define targets/specification, measure and evaluate their achievement. They may be applied within processes as part of assessments, testbench validation and process monitoring
- Are applied in independent assessments as part of compliance and/or approval activities

For example, UL4600 (see Section 8.15) sets mandatory conformance requirements for Safety Performance Indicators (SPIs) to be incorporated into the safety case. As the CertiCAV use case (see Section 2.1) includes a scope of assurance extending beyond direct safety concerns, we apply the wider term of 'Performance Indicators' rather than the more specific 'Safety Performance Indicators' term.

We've identified four types of Performance Indicator that are applicable for the CertiCAV use-case. These are:

- Event Occurrence
- Performance Models
- Formal Specifications
- Informal Specifications

The following sections provide a definition and discussion for each of these types of performance indicator.

3.2.1 Event occurrence

3.2.1.1 Definition

A measure of the rate of occurrence of describable events.

3.2.1.2 Examples

May relate to any type of describable and observable event, such as the rate of:

- Collision events
- Injury events
- Error of a computer vision system
- Occupants scoring their comfort below a threshold level
- A near collision or conflict event
- System failures and/or user intervention events

Events that are associated with a harmful consequence could be described as 'incidents'.

In the practice of functional safety, it is common to define safety levels that correlate to the consequential severity of a failure and define limits on the probability of failure for each safety level.

For example, in civil aviation the likelihood of failure of catastrophic consequence is restricted to being extremely improbable (10^{-9} or less per flight hour).

It is not possible to definitively collate the safety levels applied across various domains to failure rates due to differing quantitative and qualitative assessment methods and categorisation thresholds used. However, some typical Safety Integrity Levels (SIL) are shown in **Table 6**, and approximate mappings of SIL to failure rates in **Table 7**.¹⁷

Table 6: Typical safety integrity levels

Category	Safety Integrity Level (SIL)	Maximal Consequence of Failure
Catastrophic	4	Multiple fatalities
Critical	3	Single fatality
Marginal	2	Major injury or injuries
Negligible	1	Minor injuries and/or property damage
No Consequence	0	User inconvenience (no harm)

Table 7: Approximate mapping of Safety levels to failure rate across several domains

Range of Probability of Failure per Hour		General Programmable Electronics IEC 61508	Automotive ISO 26262	Civil Aviation DO-178/254	Railway CENELEC 50126 128/129
Max <	Min				
-	10^{-5}	(SIL 0)	QM	DAL E	(SIL 0)
10^{-5}	10^{-6}	SIL 1	ASIL A	DAL D	SIL 1
10^{-6}	10^{-7}	SIL 2	ASIL B	DAL C	SIL 2
10^{-7}	10^{-8}	SIL 3	ASIL B/C	DAL B	SIL 3
10^{-8}	10^{-9}	SIL 4	ASIL D	DAL A	SIL 4

3.2.13

Application

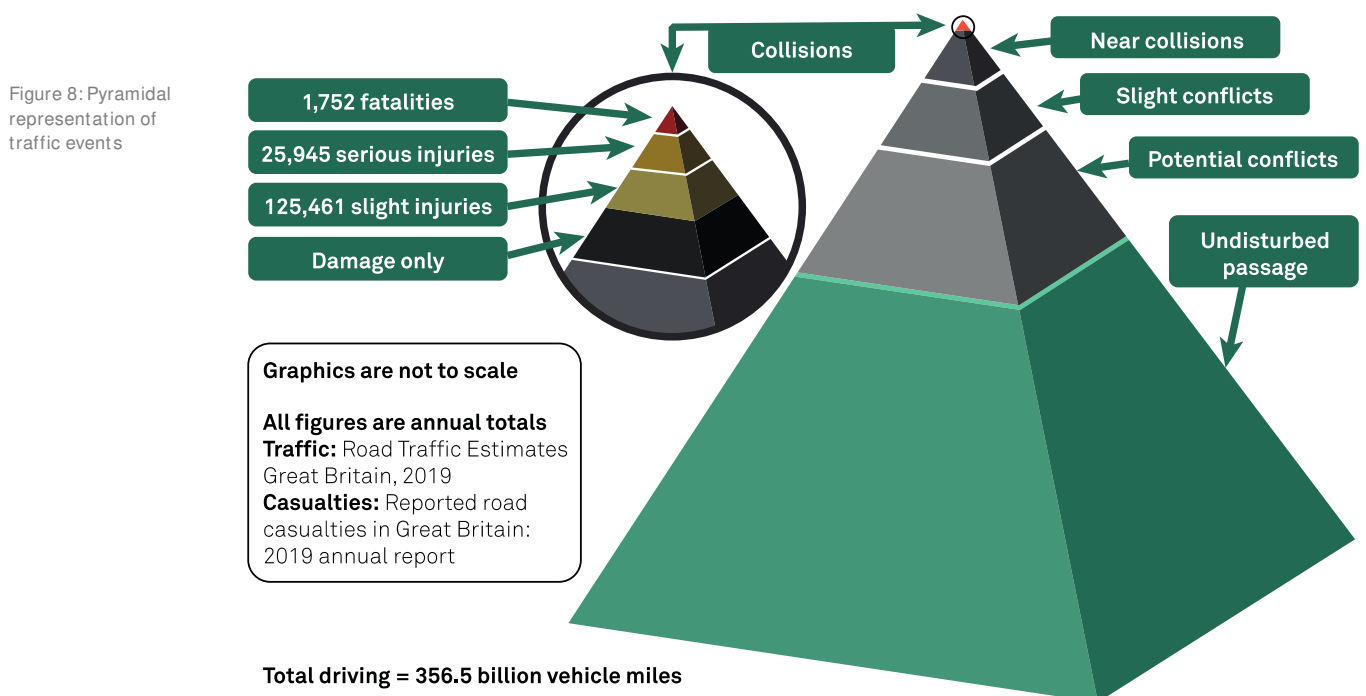
An advantage of using incidents as a metric is that they typically relate directly to an assurance need, such as absence of harm, and relate to definitive measures. They are therefore a metric that supports the validation of HAVs, and wider processes such as systems of regulation. However, a common challenge is that the acceptable rate of pertinent safety metrics including collisions and fatalities is very low, therefore a very large assessment coverage would be required to evidence achievement rates to a reasonable level of certainty.

As discussed in RAND's research (Appendix B 8.17), collecting robust evidence that a safety metric will be achieved is practically impossible pre-deployment. However, setting a measure of this kind could still have value: rather than evidencing statistical proof of incident rates it may be possible to assess evidence on the basis of whether it demonstrates capability of a deployment achieving incident rates. A process of ongoing incident monitoring (outside the scope of CertiCAV) could achieve statistically significant validation over time, or alert deployments with unacceptable rates: such a process could

¹⁷ (Reid, 2019)

provide a commercial incentive for developers to produce safe systems. However, care would need to be taken that data is accurate and representative.

A strategy for reducing the size of test-space to prove achievement of incident rates is to measure less severe incidents that occur more frequently. Road traffic incidents may be considered as a pyramid¹⁸ as shown in **Figure 8** whose volume represents distance travelled¹⁹ with the most severe and infrequent incident being fatalities²⁰ occupying a small volume at the top point followed by categories of decreasing severity and increasing frequency with a majority of driving at the base being undisturbed passage.



Furthermore, evidence generation of event occurrence is not limited to real life observation. Simulation testing of large sets of scenarios may provide wide coverage and detect a variety of incidents such as system failures. Collisions could be detected from simulation testing, and for example SAE J2980_201804 includes a reference table that indicates likely casualty severity based on the velocity delta of colliding vehicles, where the vehicles involved are passenger cars weighing up to 3.5 tons.

¹⁸ (Hyden, 1987)

¹⁹ (Department for Transport, 2020)

²⁰ (Department for Transport, 2020)

However, extrapolating outcomes and descriptions away from real life observation of definitive incidents such as collisions increases the potential for error and blurs the lines between measuring incidents, and measurement of models or specifications. For example:

- Detecting a collision in simulation is a definitive event that is relatively resilient to error, however determining casualty severity based on the velocity delta of the collision is prone to error, in particular if it does not account for the passive safety features of the particular vehicles involved
- Detecting near miss incidents in on-road use may be prone to false negative detections due to the method of monitoring or detection

This is particularly important if the indicator is used in a validation assessment of the determination of other performance indicators such as performance models. The potential margin-for-error of performance indicators used for validation needs to be accounted for in the determined confidence level of the validation.

3.2.2 Performance models

3.2.2.1 Definition

Performance Models are mathematical representations of a system performance factor such as risk or comfort or predicted event such as a collision.

3.2.2.2 Examples

Performance models are usually expressions that return a real-time value representing an abstracted system performance factor.

They may be based upon physical principles. For example, a time-to-collision model predicts the time until a collision between two actors assuming that both continue with their present velocity and speed.

They may be empirically derived, for example an event probability prediction model empirically derived from historical data of occurrence of those events, or epistemically derived from principles of physics.

3.2.2.3 Application

Two use-cases are identified for performance models.

Firstly, providing real-time metrics used as part of tactical decision making of HAV.

- A modified time-to-collision model could be implemented to trigger a collision avoidance braking feature
- A risk model could be implemented to plan the safest forward path

Secondly, assessing the HAV's performance based on driving data. An advantage is the ability to provide metrics with a relatively low level of abstraction which increases assessment coverage.

- Aggregating the values from a risk model over time would evaluate risk exposure

A challenge is that the model itself typically need to be validated to support a safety evaluation, and this may require significant quantity of data collection analysis. The validation may be specific to ODD factors.

3.2.3 Formal specifications

3.2.3.1 Definition

Formal specifications are statements expressed with formally defined vocabulary, syntax, and semantics. They should relate to explicit, quantified, and measurable criterion.

3.2.3.2 Examples

- Intel Mobileye's responsibility sensitive safety (RSS)²¹ concept which defines specifications for maintaining for certain situations. This includes safe longitudinal/lateral distances to maintain from other vehicles, a 'proper response' if those distances are encroached, right of way rules
- Nvidia safety force field (SFF)²² which similarly defines a specification following the principle that "Actors in traffic should apply a safety procedure or equivalent action before it is too late"
- Parts of road traffic law, which imposes specific and well-defined requirements on the operation of a vehicle (for example, do not exceed the posted road speed limit)
- Aggressive Driving is defined as repeated instances of manoeuvres (longitudinal and lateral accelerations) above specified thresholds completed by the ego vehicle that, cumulatively, are potentially less safe. The thresholds will vary by jurisdiction and/or culture
- A straightforward specification would be to maintain a clearance envelope from other objects. For example, to maintain 1.5 meter clearance when overtaking a cyclist. A development of this would be to increase the clearance distance as a function of the speed differential to the cyclist, to account for the aerodynamic loads placed on the cyclist²³

3.2.3.3 Application

Formal specifications can be built into a design process and are relatively straightforward to verify compliance with. They can be a useful medium for building consensus of sound principles and verifying their implementation in an ADS, for example, both RSS and SFF agree on various safety principles such as that 'right of way is given, not taken' which are applied in their specifications.

This means that it is possible to gain a high level of confidence in a system's performance in some respect before deployment. As such, well specified requirements form the basis of safety for many existing automated and high reliability systems. However, as the system becomes more complex, so does its specification. This increases the effort required in developing the specification. It also introduces a risk of 'specification bugs': cases where a system functions exactly as intended by its specifications, but not in a way that satisfies stakeholder needs. If a regulator takes on the responsibility of writing a specification, it also takes on responsibility for any errors of this type.

For example, if a requirement specifies that no traffic rules may be contravened this might prevent an ADS from committing a trivial infraction to prevent a crash. The RSS concept relies on assumptions on the behaviour of other road users, which is the subject of the "IEEE P2846: Assumptions for Models in Safety-Related Automated Vehicle Behavior"²⁴ standard.

Formal specification may be considered as being objective as they specify behaviour that is not subject to interpretation from an agent such as a driver. However, if two or more formal statements specify

²¹ (Mobileye, 2017)

²² (Nvidia, 2019)

²³ (Gromke, 2019)

²⁴ (IEEE, n.d.)

conflicting behaviour then subjective reasoning need to be applied to resolve the conflict if a formal system of assessing and resolving the conflicts is not established.

3.2.4 Informal specifications

3.2.4.1 Definition

Specifications that do not satisfy the definition of formal specifications are informal. These typically involve a degree of human judgement in their interpretation e.g. being "obvious", "reasonable", or "sufficient".

3.2.4.2 Examples

Currently defined road traffic rules and codes include several informal statements, for example from the highway code²⁵:

- Rule 114 You MUST NOT use any lights in a way which would dazzle or cause discomfort to other road users, including pedestrians, cyclists, and horse riders
- Rule 144 You MUST NOT
 - drive dangerously
 - drive without due care and attention
 - drive without reasonable consideration for other road users.

These rules do not provide compliance criteria related to objectifiable metrics. Other subjective statements that appear in rules include: 'plenty of time', 'you can see the road is clear', 'if it is safe', 'except in an emergency', and 'carefully and slowly'.

Legal tests are not analogous to specifications in this context, and a test considered to be legally objective may be an informal specification. The test that is applied in law for considering whether driving is dangerous is the objective test of the competent, careful driver. This is set out in the statute²⁶ as follows. A person is to be regarded as driving dangerously if:

- the way he drives falls far below what would be expected of a competent and careful driver, and
- it would be obvious to a competent and careful driver that driving in that way would be dangerous

Whilst such a test may be legally objective in the respect it supports a determination to be made from a 3rd person perspective, it is not a formal definition (see Section 3.2.3.1) and does not relate to explicit measurable criteria and is therefore an informal specification.

3.2.4.3 Application

Informal specifications may be considered as being subjective as they specify behaviour that is subject to interpretation from an agent such as a driver.

Requirements could define situations which a vehicle must be able to handle competently, without providing precise specification for the behaviour to be carried out. The ADS would then be judged qualitatively on whether its behaviour demonstrates this competency. This is analogous to the human

²⁵ (Department for Transport, 2019)

²⁶ (Legislation.gov.uk, n.d.)

driving test, which (in the UK) provides a structured way for a human examiner to make a judgement on behaviour.

However, the existing regulatory/practice framework applied to human drivers was not developed for regulating automated systems. A requirement such as 'Drive carefully and competently' has proven suitable for retrospective judgement of a driver's actions; however, this judgement only applies to a single specific scenario, and so cannot be used by ADSEs to write software that will always conform to the requirement. This is critical for ADSs because software typically lacks 'common sense', so cannot generalise the desired behaviour from one scenario to another superficially similar scenario. Additionally, from an assurance point-of-view, systems need to be tested in a very large number of scenarios and this makes relying on human judgement to assess the outcome of every scenario impractical.

The lack of explicitly objective requirements, such as the current system of requirements applied to humans, lends itself to self-assurance. Subjective requirements such as 'drive carefully and competently' are defined and it is left to the judgement of individual drivers to apply them. Similarly, when applied to HAV the ADSE may need to apply their own judgement based on their own evaluation of subjective measures.

Self-assurance leads to a reactive system of governance where investigation and control action is taken after a failure has occurred. This introduces increased risk of undesirable outcomes including harm, or reactively grounding a fleet of vehicles.

3.3 Developing performance requirements for HAVs

3.3.1 Introduction

Performance requirements can be defined in two areas

- **Specifications:** Defining how the system must perform. It may then be verified whether the system conforms to this requirement. Specifications in this context are deterministic: their conformance is not subject to randomness, and they are either satisfied or not satisfied.
- **Needs:** Defining the acceptable outcome of the system's performance. It may then be validated whether the system meets this requirement. In the context of assurance these are typically probabilistic in nature, relating to statistical rates with degrees of uncertainty.

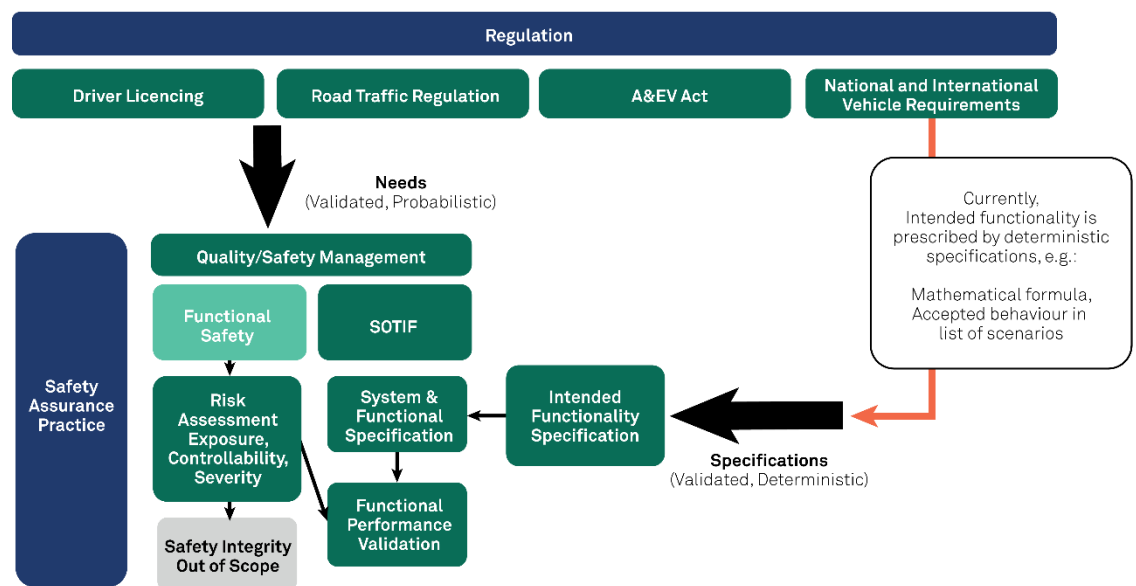
Currently, national and international vehicle requirements apply relatively prescriptive specifications for the 'intended functionality' of automated systems. In general compliance to needs based requirements is not directly referred to in requirements. For example, although for FuSa, ISO 26262 defines Automotive Safety Integrity Levels (ASIL) with broadly defined minimum failure likelihoods, compliance of these is only required indirectly through compliance to automotive QMS standards such as IATF 16949:2016.

The relationship between assurance practice and requirements is illustrated in **Figure 9**.

Considering an Advanced Emergency Braking System (AEBS) as an example:

- Intended functionality specification: is defined by the specifications and test procedures in UN-ECE regulation 131²⁷
- SOTIF: Addresses hazards caused by functional performance, e.g. false positive/ negative of object and event recognition systems, and validates the implementation of the intended functionality into the system and functional specification
- FuSa: Addresses hazards caused by the malfunctioning behaviour of electronic and electrical systems.

Figure 9: Current Interface between Assurance Practice and Regulation



When considering the application of this process to higher level automation and/or applying automation to wider ODD and/or covering more diverse functionality this raises two key questions.

Firstly, how will it be assured that an intended functionality will be sufficiently safe in use? Although SOTIF covers the identification and evaluation of hazards caused by the intended functionality, the methodology is essentially limited to defining that a systematic Hazard Identification and Risk evaluations is undertaken to determine that a validation target is met. SOTIF discussion on specification of the validation target is largely limited to stating they may relate to regulations, statistics, and functional requirements. There are further suggestions that a target could be an accident rate based on statistics, or follow rationales of: positive risk balance, 'globally at least as good' (GAME/GAMAB), or 'as low as reasonably practical' (ALARP). For an intended functionality of increased complexity, and in the absence of a human driver to cover false negatives or take responsibility, validating safety will be extremely challenging, needing to cover a very large quantity of variables and low tolerance for unknowns or error. A development of supporting practice for both defining suitable validation targets, and validating their achievement is required.

Secondly, how will the intended functionality itself be defined? SOTIF has been developed for and is applied to a system where the intended functionality of driving function behaviour is prescribed

²⁷ (United Nations, 2013)

elsewhere, for example ISO 15622:2010 for an adaptive cruise control function (ACC). The topic of developing and verifying an intended functionality for a driving behaviour function is therefore not addressed. It shall become increasingly challenging to manage the safety of HAV by defining prescriptive specification requirements because of the demands to:

- Cover requirements defined subjectively in road traffic regulations/ Highway code and driver licencing/ competency that are impractical to translate into an intended functionality specification (see Section 3.4.5.1)
- Implement tactical safety to control the exposure to external driving risks (see Section 3.4.5.2)
- Validate safe HAV behaviour when intended behaviour is not prescribed in standards (see Section 3.4.9.1)
- Ensuring the intended functionality is sufficient for taking full responsibility for the safety of the driving task. This includes the execution of fallback

It will therefore be impractical to define prescriptive requirements to manage assurance across a range of HAV; this motivates consideration of how requirements relating to performance needs may be defined.

3.3.2 Defining requirements for performance needs

Human drivers are regulated by requirements defined in Road Traffic Regulations, driving codes, and by driver licencing/ competency requirements.

The existing definition of requirements for human drivers including road traffic regulations and driver codes has evolved over time to influence suitable behaviour from human drivers and support the retrospective judgement of whether, for example, a driver's actions are 'what would be expected of a competent and careful driver'. However, it is assumed that monitoring the safety of HAV retrospectively would not be acceptable and that prospective validation of safety would be required.

We have identified three components that would need to be defined in performance requirement needs for HAV behaviour in order to support the validation of assurance:

- Conditions: statements determining where and how the requirement is applicable
- Limitations: statements determining performance thresholds
- Certainty: the level of validation required that the limitations are met during the conditions

For example, if a requirement is defined as:

- "A HAV must not cause injury collision incidents at a rate that exceeds once per billion miles of driving, validated as statistically significant with a p-value of 0.05"

This requirement:

- Has a condition of causation, if a HAV is involved in an injury collision, but was not at fault for causing the collision would not be subject to this requirement
- Has a limit on the rate of injury collision incidents of once per billion miles of driving, when the conditions of the requirement are satisfied
- The certainty is to be determined by validation is to satisfy statistical significance with a p-value of 0.05

We assume that the certainty component of requirements would be handled by guidelines applied to the validation practice in system development and approval which is discussed in Sections 4 and 5 of this document. However, further consideration is required on how conditions and

limits are defined that would cover a range of existing regulatory areas including road traffic regulations/driver codes, driver licencing, and the A&EV act in the context of HAV approval.

The above example is purely to illustrate the three components, and does not purport to be a complete solution. We suggest that ideally defined requirements would:

- When taken as a whole, define acceptable limits on risk
- Reference formal and measurable performance indicator metrics
- Be targeted, structured, and implementable in assessments
- Be free of conflicts
- Account for operating in a high risk multi agent environment. A HAV needs to be understood by other road users, and sufficiently mitigate the potential unreasonable actions or misuse of others
- Account for the need to operate effectively within an existing system of governance & infrastructure evolved over time for human use, co-operate with other road users and facilitate efficient transport
- Support ethics: For example, in addition to demonstrating the introduction of a HAV system is globally a positive risk balance compared to human driving, no particular group of people should be subject a negative risk balance
- Be proportional: For example, rather than having a single casualty incident rate limit, the limit could be proportional to the existing casualty rates for category of road, or other operational baseline risk factors

The needs to be addressed are therefore:

- The definition of measurable driving performance Indicators that may be referred to in requirements: This is discussed in Section 3.2 where types of performance indicators are presented
- Identification of what conditions and limits that would need to be in a set of requirements to address the above needs. This is discussed in subsequent Section 3.3.3

3.3.3 Driving performance criteria

As above, it is suggested to develop a set of conditions and limitations which when taken as a whole define acceptable limits on risk.

To explore this, we follow a process of:

- Defining an example set of criteria that describes fulfilment of safe and effective driving
- Consider the challenges of applying these criteria to requirements
- Considering requirement development options

This is followed with a case study example of how this may be applied, and discussion on how requirements relate to assurance process.

The purpose of driving performance criteria is to establish a set of statements which provide a complete qualitative description of safe and effective driving. The example criteria proposed in this document are intended to represent a plausible version of such a description. However, it remains to be seen whether all public concerns about automated vehicle driving standards will be covered by this list, so they may need to be updated as more evidence becomes available.

Note that:

- These are example criteria, that are limited to those applied to HAV behaviour. We do not attempt to consider criteria that would apply to the performance of sub-systems such as perception and decision making
- The goal in developing this list was to completely cover a definition of desirable driving, and not to try and rationalise into a concise list that minimises overlap between criteria

We assume that the priority is to cover safety (for both automated vehicle occupants and other road users) and to manage any potential negative effects on other road users. This means that many aspects of performance with an effect on user experience are excluded from our criteria. However, a limited amount of performance assessment is included, since a poorly performing vehicle could create a safety risk or seriously disrupt traffic flow.

In MUSICC, a previous CPC project, we published a proposed set of criteria for use as part of an automated evaluation process²⁸. These were taken as a starting point and reviewed as part of this project, with input from several key stakeholders. A revised and expanded set of criteria is presented below:

Table 8: Driving Performance Criteria

Ref	Criterion	Description
1	Do not cause harm	When conditions (e.g. other road user behaviour) are reasonable then collisions are not acceptable. The threshold of acceptance for HAV to cause collisions is very low.
2	Avoid harm, even when not the cause	Expected to allow for possibility of unreasonable behaviour by others, or hazardous conditions. The expectation is that a HAV should be able to detect and avoid/mitigate hazards with a performance at least equal to a reasonable human driver.
3	Provide occupant safety/comfort	For an omnibus driving smoothly is an important safety factor, for vehicles with seat-belted occupants this is limited to comfort unless accelerations are severe.
4	Follow Traffic Rules	HAV is expected to comply with road traffic regulations
5	Provide reasonable safety margins	HAV is expected to maintain safety margins that are safety resilient for both operation of HAV and appear reasonable to other road users.
6	Follow recommended driving practice	Expected to exhibit good driving behaviour, following guidance in highway code and other relevant practice. ²⁹
7	Facilitate established driving conventions	Expected to reasonably blend in with established driving practice, not to confuse or unreasonably hinder other road users.
8	Behave considerately to other road users	Expected to follow courteous driving practice to facilitate smooth and harmonious experience for all road users.
9	Do not unreasonably obstruct movement of traffic	Being overly hesitant/cautious and on a larger scale has environmental/ economic impact on traffic logistics.

²⁸ (Myers & Saigol, 2020)

²⁹ (The Police Foundation, 2020)

3.3.3.1

Conditions and limits in context of a human comparator

The description makes some suggestions on limits. The Law Commissions' CP3³⁰ paragraphs 5.60-5.65 discusses the principle of 'a positive risk balance compared to human driving performance' and observes that 'Given that the casualties that occur involve only a small minority of (often poor) drivers, a benchmark of "better than the mean" may be seen as too low.' We derived similar reasoning in the development of limit suggestions. Rather than applying the broad principle of achieving a positive risk balance, we referenced human performance benchmarks in a more context specific and conditional way that we suggest may be more ethically acceptable.

For example, for criterion 2 of avoiding collisions of which the HAV is not the cause, it is suggested that in this case a reasonable human is a good benchmark for hazard perception and response. Therefore, when avoiding hazards caused by faults of others, performing similar to a reasonable human driver is sufficient. This human benchmark comparison is conditional on:

- The comparison only applying only to events where the HAV is not the cause
- Using only reasonable humans as a comparator, for example based on studies where humans are known to be reasonably competent and attentive rather than holistic incident statistics that may include instances of unreasonable ego drivers

Furthermore, for criterion 1 it is suggested that tolerance for a HAV to be at fault of a collision is very low, and it may be expected for a HAV to perform significantly better than humans who may be inattentive or reckless.

3.3.3.2

Role of traffic rules

Much of the existing literature on HAV assurance either takes the view that legal requirements are not important for safety³¹, or that the assurance process must show absolute adherence to all rules of the road³² (see Appendix 8.17 for further discussion of the RAND report). In the safety criteria presented above, not only is adhering to traffic rules a key criterion, but additionally criterion 6 is included so that the HAV will follow practice that is recommended but not enshrined in law (which is the case for a significant amount of the UK Highway Code).

The rules of the road are critical for safety and co-operation because they set not only the ADSs expectations of what other road users will do, but also the expectations of other road users for the HAV. These reasonable expectations of other road user behaviour are key to facilitating the smooth flow of traffic, as road users can proceed in a way that reasonably ensures lack of conflict with each other without having to account for every possible action of other road users. As a simplistic example of this, imagine approaching a roundabout such that you can see the roundabout is completely clear. If you can see the road comprising the roundabout for an equal distance to both the left and right, some safety specifications will treat going around the roundabout either way as equally safe. This is clearly wrong, as one direction will result in the HAV driving directly into any hidden oncoming traffic.

An alternative way to ensure that behaviour is consistent with the expectations of other road users would be to directly collect data on road user behaviour within the system's deployment area. If enough data could be collected, this could be used to build a model of average and extreme actions for

³⁰ (The Law Commissions, 2020)

³¹ (Blumenthal, Fraade-Blanar, Best, & Irwin, 2020; Shalev-Shwartz, Shammah, & Shashua, 2017)

³² (UNECE, 2020)

that particular junction or road segment. Assuming the model was accurate enough, the ADS could use it to predict the behaviour of other road users, and ensure it conformed to the average behaviour itself, making the safety of the ADS independent of the actual rules of the road. This has a key advantage in that the ADS behaviour would reflect what really happens on the road, rather than what the rules say should happen. However, there are also several drawbacks:

- A large amount of data must be collected for even very infrequently used sections of road. Data will need to be kept up to date if road users' behaviour changes
- New or altered road sections cannot be navigated until data for them has been collected
- The overall safety level is at least partly determined the behaviour of the current population of road users, which may not be as safe as if more road users adhered strictly to the rules of the road
- It is likely to be extremely important to regulators that HAVs obey the law, regardless of the behaviour of human road users

The alternative viewpoint, that the HAV must always obey all traffic rules, would be the correct approach if the traffic rules were specified in such a way that they covered the full DDT, were precisely defined, did not conflict with each other, and allowed the free and fair flow of traffic in all situations where that is achieved by human drivers. If the traffic rules do not meet these criteria then slavishly obeying them may lead to a reduction in safety in some circumstances, as discussed in [Section 3.4.5.1](#).

3.3.4 Application challenges

There are a variety of challenges that need to be addressed when applying these criteria to an assurance framework as follows.

3.3.4.1 Overlap between criteria

The scope of various criteria overlaps with each other. For example, there is overlap between providing reasonable safety margins, recommended driving practice, and following traffic rules.

This means that the application of each criteria cannot be addressed in isolation, it must be considered in relation to how the other criteria are being treated.

3.3.4.2 Conflicts between criteria or within criteria

There are conflicts between criteria that need to be addressed.

Example: Following the code of not causing vehicles on the major road to slow down when pulling out from a side road may conflict with the need to avoid obstructing the movement of traffic.

Example: Following good practice of leaving recommended following distance may conflict with facilitating established driving conventions and increase global risk.

3.3.4.3 Metrics and traceability

To allow defined requirements to be traced against system validation outputs and system safety case argumentation it is necessary to reference common metrics. This role may be fulfilled by performance indicators as described in [Section 3.2](#). Ideally any referenced performance indicators would be formally defined and explicit – to allow direct mapping to formal verification and validation techniques. However, various performance indicators used currently for (human) driving such as compliance with rules/codes are not formally defined which introduces challenges referencing formal metrics to some criterion.

3.3.4.4 Complexity and completeness

The criteria are attempting to capture the correct behaviour for an actor involved in highly complex interactions in a highly complex high risk environment involving multi-agent systems and have not been formally verified or validated. The criteria we've set out in this document are not necessarily complete/and or optimal, amending them accordingly may highlight new challenges, issues, constraints and so forth.

3.3.4.5 Resolution of challenges

An approach for resolving challenges relating to overlap and conflicts between criteria would be to order requirement statements into a hierarchy. Singapore's technical reference for automated vehicles discusses the dilemmas of application of rules to automated driving as summarised in [Section 8.16](#). Whilst the driving performance criteria presented above are ordered approximately in a hierarchy, the resolution of this ultimately becomes a situation specific evaluation of safety and performance costs. Developing a system to resolve this on a wide scale is highly complex.

For example, considering the third criterion 'Provide occupant safety/comfort', if the HAV is a double deck bus that carries standing occupants, the potential severity outcome for scenarios involving heavy braking could include fatalities. However, if the vehicle is a passenger car where occupants are seat-belted, the potential severity outcome for the same scenarios may be limited to inconvenience/discomfort only, lowering this criterion on the hierarchy.

Referencing formal metrics to support formal verification and validation of requirements requires behavioural safety challenges to be overcome (see [Section 3.4.5](#)). This would include the formalisation of traffic rules/codes and modelling multi-agent risk, which are discussed in [Sections 3.4.5.1](#) and [Section 3.4.5.2](#) respectively.

There are multiple layers of complexity being covered by requirements, and if their coverage is not complete this leaves potential for unaccounted risks. It is likely that requirements themselves would need to be validated over a period of time of use, with monitoring and feedback identifying concerns to be resolved with iterative changes until the system of regulation is matured.

3.3.5 Requirement development options

The driving performance criteria in Section 3.2 attempt to provide a superset of statements which describe safe and effective driving. However, they are not detailed enough to be used as technical requirements. The descriptions of the driving criteria contain some hints at conditions, limitations, and certainty, but these are not fully defined. Also, they do not reference performance indicators, which makes it difficult to link them to metrics gathered in the validation process. This means that a further process to define requirements is required.

Defining these requirements could be technically challenging. Since the driving performance criteria are qualitative, it is impractical to formally show that any given set of requirements will meet them. This means that the set of regulatory requirements will have to be validated based on operational experience. An initial set of requirements should represent the best effort to set technical criteria for automated vehicle acceptance, but, given the complexity of the operating environment, it seems likely that they will need to be refined over time.

Section 3.3.4 identified that there are trade-offs to be made when setting requirements. Expecting perfect performance against all criteria is unrealistic, so the limitations aspect of the requirements must allow for some tolerable level of risk. Defining exact levels is a sensitive topic, but certain principles³³ are well established:

- There is a limit on how much risk any individual should be exposed to. This can sometimes be expressed numerically, but in practice depends on the level of societal concern
- Where a risk is significant, but not unacceptable to any individual, it should be mitigated as far as reasonably practicable (i.e. until the costs of further mitigation are greater than the benefits)
- Very low-level risks rarely justify regulatory intervention

As an example, these principles are applied to the first driving performance criterion, "Do not cause harm". A reasonable set of conditions and limitations for this criterion could look like those in the table below:

³³ (Health & Safety Executive, 2001)

Ref	Requirement	Rationale/ discussion
1a	Do not cause collisions at more than a defined rate per operating hour	Likely to be a relatively high level of societal concern about collisions caused by a HAV. This means that the tolerance for HAV caused collisions will be lower than other types of collisions. Setting a tolerable rate of zero is unlikely to be proportionate: beyond a certain point, the safety risk is negligible, so it may not be reasonably practicable to go further.
1b	Causality shall be determined with reference to a formal collision responsibility specification	The tolerance for HAV caused collisions is low. One way this can practically be verified is by reference to a formal specification. All HAVs must use compatible definitions of causality.
2	Where the HAV is designed to execute an action, which may foreseeably cause a collision, this must be justified by a validated risk assessment	System developers should be expected to reduce risk as far as is reasonably practicable, even if the overall rate of collisions caused is below the threshold.

Taken together, requirements 1a and 1b impose a limitation (maximum rate of incident occurrence) subject to a condition (a formal specification to define causality). Requirement 2 can also be read in this way; however, this is based on informal specifications (actions which may foreseeably cause a collision³⁴ which are not justified by a validated risk assessment) and a limit of zero. The requirements can also be related to performance indicators as discussed in Section 3.2. Note that this table does not cover the confidence level aspect of the requirement (this should be handled separately, as suggested in 3.3.2.)

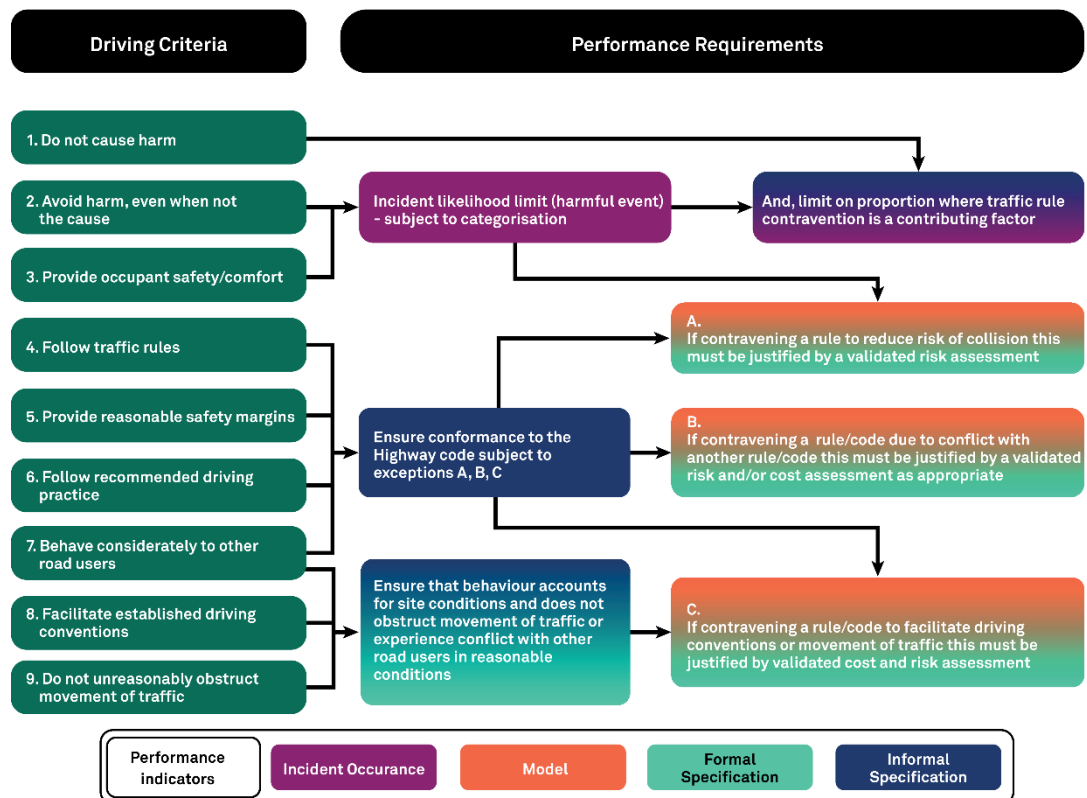
Some additional considerations, not considered in the example above, will need to be applied to the design of an overall set of requirements. These include:

- How to cover all aspects of the driving criteria with minimal overlap between requirements
- Practical application of the derived requirement to assurance practice. When applying to assessment there is a compromise between level of granularity of the evaluation, and assurance obtained from a certain number of tests

Figure 10 below shows a high-level example of requirements that attempt to satisfy assurance needs across all driving criteria:

³⁴ This part of the requirement could, in some circumstances, be easily formalised. For example, if a vehicle causes a collision in testing, it is likely to be foreseeable that the same could happen again after deployment.

Figure 10: Example performance requirement definition



Key features of this example include:

- A limit is proposed for both the total likelihood of incidents and the subset of these for which the system under test is determined to be responsible. There is deliberately no separate limit for not-at-fault incidents: otherwise, there are plausible situations where an ADS developer could be incentivised to increase the likelihood of their vehicle being considered at fault (in order to move a collision out of the not-at-fault category)
- Incident likelihood is subject to further categorisation. This could consider factors such as the level of challenge posed by the road environment, likely severity of harm, and the risk to different groups of road user (to ensure no individual is exposed to excessive risk)
- The Highway Code is included but treated as an informal specification. In practice, it may be desirable to formalise some aspects of the Highway Code, as discussed in 3.4.5.
- We have included road traffic rules in the category of Highway Code for clarity. In reality, the "must" rules in the Highway Code summarise legislation, which provides more detailed definitions. However, these definitions are usually also in the form of an informal specification, so the process presented here remains similar

Exceptions A, B and C are available to allow contraventions of the Highway Code to enable compliance with the driving criteria. Each of these fulfils a slightly different purpose:

- Exception A is required as it is currently not assured that the following the Highway Code will always permit the safest action in every situation. If this exemption did not exist, a much greater level of validation of road rules would be required to ensure that following them precisely in every situation cannot contribute to the causation of harm
- Exception B handles contradictions within the Highway Code. These are likely to be most important in edge cases. The need for this exemption could be reduced or removed if the Highway Code itself defined how such conflicts should be handled (e.g. by providing a hierarchy of rules)
- Exception C reflects an acknowledgement that sticking to the rules rigidly may not always be the best outcome for society in terms of convenience and traffic flow. It could, for example, lead to vehicles becoming stuck and blocking key parts of the road network. Again, it is possible that a better-validated Highway Code would not require this exemption

These exceptions refer to justification by a validated risk and/or cost assessment. These could be implemented in a number of ways, for example as real-time models, or in software design with an extensive set of situation specific rules, or a combination thereof.

The approach is designed to ensure that following rules is the default, with any departure requiring explicit thought and argument. It is recognised that developing such a validated risk and/or cost assessment would be challenging, however under this ruleset developers have the option to avoid this if their system specification and use allows. While the properties of the ruleset identify where exemptions may be needed, it is not necessarily the case that all will need to be used. The need to use exemptions may depend on the complexity of the deployment and the design of the ADS. Developers could choose to design a system or specify their deployment so that situations requiring exemptions are avoided.

3.4 HAV assurance practice

In this section we identify and discuss areas of HAV development processes. This is a highly complex area that is under active development internationally. Accordingly, this section does not provide a comprehensive system of practice, rather it focuses on identifying key issues, gaps and suggesting potential approaches to address them.

3.4.1 HAV assurance best practice

Best practice provides an important interface between System Developer and Independent assurance verification processes.

If best practice is sufficiently developed this facilitates independent testing, inspection, and conformance/certification activity to be integrated as part of the system development process and streamline assurance activities.

Existing discussion on automated vehicle system safety often cites the following three areas of validation/assessment practice: functional safety (ISO 26262), safety of the intended functionality (ISO 21448), and cybersecurity (ISO/SAE 21434). Which leads us to the following observations:

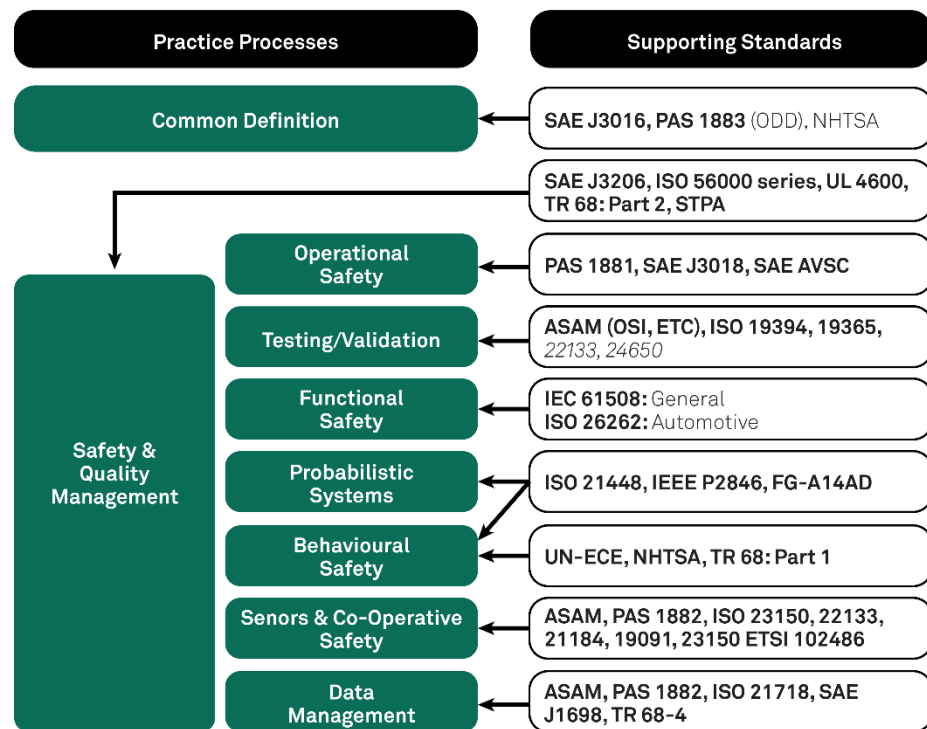
- All of these standards are intended to cover SAE level 0-2 driving automating systems, they therefore may be inappropriate for, or leave gaps when applied to, SAE level 3-5 ADS
- ISO 21448 SOTIF's normative content is limited to its objectives; the methodologies are informative
- ISO/SAE 21434 is weighted towards software validation methods and does not comprehensively cover cybersecurity design and assessment
- The scope of existing best practice applied to functional safety, SOTIF and cybersecurity does not completely cover HAV assurance

Figure 11 below illustrates some of the processes involved in HAV development and some examples of standards that inform good practice for these processes. Currently these areas are under very active development. Furthermore, due to the large scale and complexity of the systems and challenges involved, some gaps remain that are not covered by current practice development activities. Therefore, in several areas it remains uncertain what practice may be adopted by industry and regulators with there being competing practice proposals, or gaps in practice development.

In August 2020 BSI published a Connected and Automated Vehicles Standards Roadmap³⁵ which maps out relevant published standards, current/forecast standards activities, and identified standards gaps.

³⁵ (British Standards Institute, n.d.)

Figure 11: Example of HAV development assurance practice



An outline description of the identified processes:

- **Common Definition:** The definition of terminology and system architecture referred to in assurance practice (see Section 2.3)
- **Safety and Quality Management:** Ensuring that requirements are sufficiently defined, owned, and achieved
- **Operational Safety:** Discovering, specifying, and validating processes and attributes relating to safe use in operational conditions
- **Functional Safety:** Addressing hazards caused by the malfunctioning behaviour systems in vehicles
- **Probabilistic Systems:** Safe utilisation and operation of probabilistic systems
- **Behavioural Safety:** Identifying and Validating appropriate intended functionality
- **Sensors & Co-operative Safety:** Performance and safety of sensor systems and use of sensor data
- **Data Management:** The collection, use, and management of data
- **Testing & Validation:** Processes and tooling for Testing and Validation

This section continues to provide a more detailed discussion for each of these assurance processes. Further development of these processes would help to support system assurance needs.

3.4.2 Safety & quality management

This is the overarching process of ensuring that requirements are sufficiently defined, owned, and achieved. This may include satisfying the following objectives (non-exhaustive):

- Define an appropriate organisational risk culture. Assess and define organisational risk appetite that is demonstrably within regulatory requirements. Describe principles for vehicle testing culture
- Define and manage safety for the whole lifecycle of the HAV
- Identify and apply appropriate practice and processes
- Adopt a sufficiently well developed and systematic process for managing safety/quality and ensure safety requirements are satisfied
- Address risks across the entire supply chain with well-defined roles & responsibilities and ensure access to, and retention of, control documents and evidence
- Co-ordinate and sufficiently explain the deployment with necessary stakeholders and develop a complete system specification
- Document full specifications for nominal state of systems and integrate this into testing process
- Plan processes to account for the whole safety lifecycle of the product
- Organise safety for the whole system, ensure safety arguments are systematic structured, evidenced and documented
- Apply appropriate change management

Several of these objectives refer to areas out of scope for CertiCAV, including operational & organisational safety. However, the topics of system safety case and change management, which are considered of relevance to system safety, are discussed below.

A challenge to this process in the context of HAV is the current reliance on safety case which requires a high effort to develop, maintain, and review when compared to managing safety through conformance to specification standards. However, addressing safety through conformance requires the support of well-developed and validated safety standards. As HASSs are highly complex and diverse it will likely be impractical to standardise architectures and validation processes to a sufficient degree that replaces safety case. It would at the very least take a significant period of time for standards to develop sufficiently for assurance to be demonstrated by their conformance and it is likely that the ongoing development of standards will increase the scope of safety addressed through conformance, and thus reduce the scope of safety case – rather than entirely removing the role of safety case in addressing safety.

A system safety case is likely to be large, complex, and therefore challenging to scrutinise bottom up from its body of evidence. It may therefore be helpful to consider safety management as a top down process of ensuring that suitable risk appetite statements and safety principles are defined and owned at a top level in the organisation and embodied by the whole organisational structure and their activities. Furthermore, that safety management is co-ordinated across all system development processes, ideally by assessing risks in a common top level safety management process rather than addressing safety in separate potentially competing safety processes such as functional safety and cybersecurity without considering their respective inputs/outputs and impacts on each other.

3.4.2.1

System safety case

Existing practice often refer to safety assurance to be demonstrated by a safety case. For example PAS 1881³⁶ specifies requirements for safety cases for automated vehicles, with the intention of demonstrating that trialling and testing activities can be undertaken safely and securely. 'Safety case' is however a term used widely in a variety of contexts so can be interpreted to have a wide range of forms. For the context of demonstrating the achievement of system safety it is important that a safety case is a systematic and structured argument that justifies the satisfaction of the systems needs/specifications, which is in turn supported by evidence that demonstrates that the arguments are satisfied. It is therefore imperative to distinguish this definition of safety case from other definitions of safety case that are appropriate for a different context, for example the collation and provision of a prescribed set of documentation.

The application of safety case is already established practice for complex automotive systems, ISO 26262:2018 in its latest revision introduced the explicit requirement for safety case to be based on an argument. MISRA published guidelines for automotive safety arguments to support this³⁷

Singapore's technical reference for Autonomous Vehicles TR68: Part 2 'Safety': 2019 provides a scope chart in Goal Structure Notation (GSN) structured into Goals(claims), Strategies(arguments), and Solutions(evidence). It shows relationships to standards compliance, links to other references and applicable Assumptions/justifications

UL4600 describes practice for safety case relevant for system development and approval (discussed in Section 8.15). This includes requirements to structure safety case into goals, argumentation, and Evidence, and reference the argumentation to safety performance indicators.

Uber ATG have published their safety case framework³⁸ which follows a GSN.

The size and complexity of safety validation for HAV gives rise to the need to develop practice to co-ordinate the various validation processes and provide integrity of whole system assurance. Safety case development should ensure relational inputs and outputs of safety processes are identified and implemented.

For instance, in current practice a validation process (such as FuSa, SOTIF) will accept product release when residual risks are assessed to be acceptably within a specified validation target. However, for complex HAV deployments the impacts of residual risks are unlikely to be confined to the evaluation of a single validation process but have consequential impacts on other validation processes, therefore there may be a need to go beyond quantifying residual risks but to characterise them and ensure they're accounted for in all relevant validation processes. For example:

- The risk associated with being in MRC may need to be considered in relation to the likelihood/rate of MRC being triggered. This would be due to conditions assessed by a variety of processes including: functional safety, and cybersecurity assessment
- In Cybersecurity Validation a Threat and Risk Assessment may determine the security risk of HASSs being affected by security threats, however determining the safety assurance relevant safety risk would require these system impacts to be evaluated in other validation processes such as functional safety and/or SOTIF

³⁶ (The British Standards Institution, 2020)

³⁷ (MISRA consortium, 2019)

³⁸ (Uber ATG, n.d.)

- The residual risk of failures/hazards assessed in separate validation processes could be considered acceptable due to low severity consequence, however if their occurrence were considered together the severity consequence could be increased making the residual risk unacceptable. For instance
 - False negative detection of puddle depth by perception system; and failure of ABS sensor due to water submersion
 - Potential misuse of seatbelt monitoring system; and hazard of a hard braking event
- They may rely on the assumed availability and/or performance of systems not under the control of the developer
 - It would need to be assured that any assumed component specifications e.g. Exteroceptive sensors are robust
 - V2X or other through the air communications systems maintained/owned by others

3.4.2.2

Change management

Although out of scope for CertiCAV the need to develop practice for change management is recognised.

During system development it is impractical to repeat all validation/evidence generation processes every time a development change is made. Therefore, a change management process is required to robustly identify which verification and validation activities are affected.

It is expected that ongoing changes will be made to systems throughout the whole produce lifecycle, for example to reflect changes to road rules/regulations, changes to ODD, fix identified bugs or cybersecurity vulnerabilities. An appropriate process will be required of identifying and managing minor and major update changes. This would need to achieve a suitable compromise between ensuring that changes do not introduce safety or security issues, but the process does not detract from timely rollout of updates that may be beneficial to system safety, security, and effectiveness.

- Establish the stages of whole lifecycle for assets and implement an according change management process for software and hardware
- This needs to be associated with appropriate processes for product verification and validation testing

3.4.3 Operational safety

Although this document focuses on system safety, and not operational safety, these two areas are highly interdependent. Operational safety considerations provide an important input to system safety assurance. For instance, on defining how safety critical are the impacts of particular system performance features or system failures. The HASS specification in Section 2.3.3 discusses the specification of some operational characteristics that are identified as important to consider as part of system assurance.

Objectives that may be considered part of operational safety that we have identified as impactful for system safety include:

- Validating appropriateness of the ODD for the deployment operations
 - A complete and accurate definition of the ODD is a critical input for ensuring complete and accurate system validation e.g. in the process of defining test scenarios. ODD validation is therefore part of system validation
- Deployment domain conditions characterisation
 - Determining the frequency that ODD boundary conditions would occur during deployment operations would be required to calculate risks associated with fallback operation
- System/user handover
 - How the system reaches/leaves the ODD and the procedure for activating automated mode affects system design
- Monitoring / alerting
 - How the system performance will be and procedure of alerting and response to issues and incidents affects system design
- Organisational relationships
 - The responsibilities of organisations should be clear. For example, regarding liability, or performance validation
- Maintenance / inspection
 - For example, the regime of component inspection and component replacement may affect specification system component reliability and/or self-monitoring capabilities

3.4.4 Functional safety (FuSa)

This document, due to its scope does not cover FuSa which aims to address hazards associated with malfunctioning system behaviour. We have however identified two areas of interest in relation to HAV assurance that would require development of existing practice defined in ISO 26262:2018, which are addressing FuSa of:

- The ADS, which is a highly complex system with many malfunctioning behaviours that may have high impact consequences. For example, failure of an exteroceptive sensor; perception, decision making or control system. Furthermore, detection of malfunctioning behaviour may be challenging for items of the ADS whose function is probabilistic, such as exteroceptive sensors. Therefore, may require additional hazard classifications, and/or assessment techniques to address this
- Other HAV systems, such as actuators, whose safety case is impacted for higher automated mode use-case. For example: the FuSa case for an existing vehicle platform may typically rely on the physical input of the human driver to provide sufficient controllability to avoid harmful outcome in the event of power steering system failure, or braking servo failure – therefore, if such a platform is adapted for HAV deployment then alternative fallback would need to be implemented and the functional safety case re-considered

Furthermore, functional safety may provide as an output the exposure and system impact of system fault conditions. This is a relevant input to other safety assessment processes so that the relevant risks may be quantified.

For example, a FuSa hazard and risk assessment (HARA) may characterise the exposure and system impact relating to HAV scenario exposure and controllability of different malfunctioning systems such as:

- Partially deflated tyre: exposure probability high, controllability impact low.
- HAV internal power supply failure: exposure probability low, controllability impact high

Combining these attributes would be a relevant input to the generation/testing of scenarios for wider system validation as discussed in [Section 3.4.9](#).

3.4.5 Behavioural safety

As discussed in [section 8.14](#), SOTIF practice as embodied in ISO 21448 is best practice for validating that systems involving probabilistic sensor systems perform as intended i.e. that situational awareness is sufficient.

SOTIF is also aimed at addressing hazards arising from:

- Performance limitations
- Reasonably foreseeable misuse such as incorrect HMI (in conjunction with ISO 26262)
- Impact from car surroundings such as other users, physical and digital infrastructure, sensory conditions (in conjunction with ISO 26262)

There are however gaps which SOTIF does not cover in relation to system development of a HAV:

- It has been developed for and is applied to a system where the intended functionality of driving function behaviour is prescribed elsewhere, for example ISO 15622:2010 for an adaptive cruise control function (ACC)
- It is intended to be applied to systems automated to SAE level 0-2 where a human remains responsible for the driving task. Therefore, its objectives are to ensure the activation, termination, and performance of automated driving features is free from unacceptable risk. It does not cover ensuring that the combination of driving features provides sufficient safety in the operating environment when fully responsible for sustained DDT performance
- It does not explicitly cover the need for intended functionality to include suitable fallback behaviour response in the event of system performance limitations being reached
- In the context of SOTIF, scenarios are a sequence of scenes and the validation focuses on the performance of the perception of environment from these scenes. To effectively assess the performance of the HAVs driving function behaviour, a different process would be required. This would need to consider scenarios on a higher level of abstraction, including interaction between the HAV and its environment and surrounding actors.

To address these gaps we define behavioural safety as the:

- Field of safety concerned with addressing the safety and sufficiency of the intended functionality of driving features

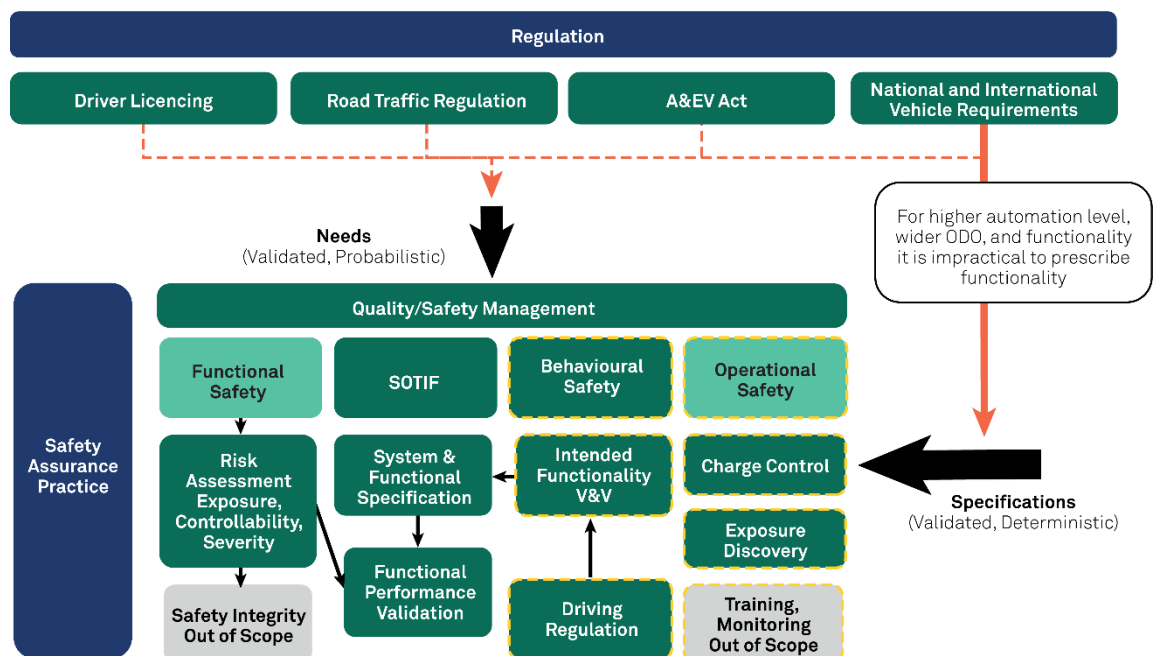
We identify the need for behavioural safety as a process to create, verify and validate the intended functionality of HAV driving features for a particular HAV deployment to implement an intended HAV behaviour that satisfies the following objectives:

- Suitably complying with road traffic rules and co-operating with other road users (see Section 3.4.5.1)
- Achieving acceptable safety of HAV behaviour by implementing suitable active and tactical safety controls (see Section 3.4.5.2)
- Implementing suitable behaviour to cover fallback operation including:
 - Ensuring MRM is suitable
 - Accounting for the risk of being in MRC after execution of MRM, including the likelihood of the MRM-triggering condition
 - Quantifying and managing potential for unknown unsafe scenarios arising from lack of human in loop as a fallback

The relationship between behavioural safety and SOTIF is illustrated in **Figure 12**. The development of common practice for this area of behavioural safety important for ensuring that the intended HAV behaviour is adequately defined and sufficiently verified and validated.

- The intended functionality of HAV driving features could be proprietary and developed by the system developer. In which case the common practice would support with references, harmonised process and achieving assurance through conformance demonstration.
- Or, it could be developed by regulators, in which case the common practice would similarly provide a reference and process for achieving this.
- Or, what could be considered most likely scenario, is a combination of both the above. In which case common practice would have additional value of providing a common system supporting co-operation of activities and communication between system developer and regulator.

Figure 12: Future Interface between Assurance Practice and Regulation



We have identified that the complexity associated with HAV deployment will likely have a significant impact on the complexity of the behavioural safety process, and therefore on the argumentation of the whole system safety case.

For example, if the example requirements discussed in Section 3.3.5 and shown in **Figure 10** are assumed then:

- If to demonstrate achievement of requirements for harm avoidance it is necessary to consider the contravention of rules then it would be necessary to develop the risk assessment to justify and implement this, for example by developing and validating real time risk assessment models as part of the safety argumentation.
- Such as: for a complex high speed deployment, scenarios may be identified where rules may need to be contravened to avoid harm due to inappropriate action of other road users.
- However, if it can be demonstrated that requirements for avoiding harm can be met without contravention of any rules, there would be no need for justification which simplifies safety argumentation
- Such as: for a low speed deployment in a controlled environment it may be identified that the suitable response to all inappropriate action from other road users is to stop or slow down which does not involve the contravention of any road rules

3.4.5.1

Road Traffic Regulations, Highway codes, Driver Codes, Driver Licencing, and competency

The behaviour of a HAV will be expected to satisfy requirements similar to those currently defined in road traffic regulations and driving codes and managed by the human driver licencing and competency system. This therefore needs to be reflected in intended functionality specifications.

- The Law Commissions have made various recommendations following two research report and consultation exercises that are summarised in Section 8.5, they published a third consultation in December 2020
- The UNECE WP 29 draft "Revised Framework document on automated/autonomous vehicles" states as a principle of "ensure compliance with road traffic regulations"
- Driving regulations and codes are not formal in the technical sense. These serve their current purpose of influencing the behaviour of humans and allowing retrospective judgement of whether a driver's actions were reasonable; they do not however support prospective formal verification and validation of conformance or safety performance
- Care needs to be taken that the way driving regulations are referred to in requirements does not restrict functionality from being implemented that would prevent harm

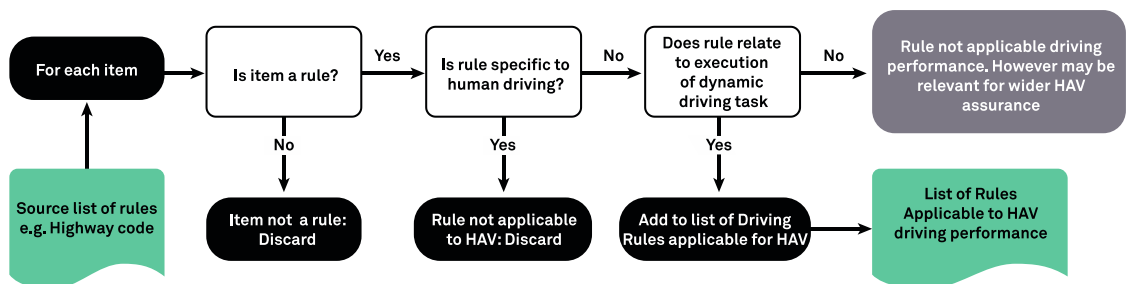
The current system has evolved over time with the purpose of being a suitable instrument for regulating and legislating the behaviour of humans; lagging measures of traffic injury statistics validate its success.

There are however various issues of applying this system to HAV including:

- The system is not designed or validated for regulating HAVs, the automated compliance to a set of existing rules and codes in itself does not guarantee satisfaction of assurance needs such as absence of collisions
- Many of the statements in driving rules and codes are defined subjectively:
 - They may not be explicitly defined, and/or do not relate to measurable metrics
 - Compliance may be subject to caveats
 - Compliance may conflict with compliance to other rules/codes and there is no clear hierarchy to resolve this
- Several rules/guidance are specific to human driving, for example relating to looking into mirrors before undertaking manoeuvres
- Many rules are not related to the driving task. Human drivers discharge various legal responsibilities outside performance of DDT, including:
 - Ensuring occupants under the age of 14 have appropriate crash restraints applied (seatbelt, child seat)
 - Ensuring vehicle is roadworthy: e.g. clear of snow, lights and indicators, exhaust attachment, engine not smoking and/or on-fire
 - Managing breakdowns (warning triangle) and incidents (helping casualties, calling emergency services, exchanging details),
 - Not knowingly assisting criminal activity e.g. drug dealing
- The highway code in particular intended to influence the behaviour of humans. It is not presented as rationalised and organised set of rules suitable for programming an automated driving system

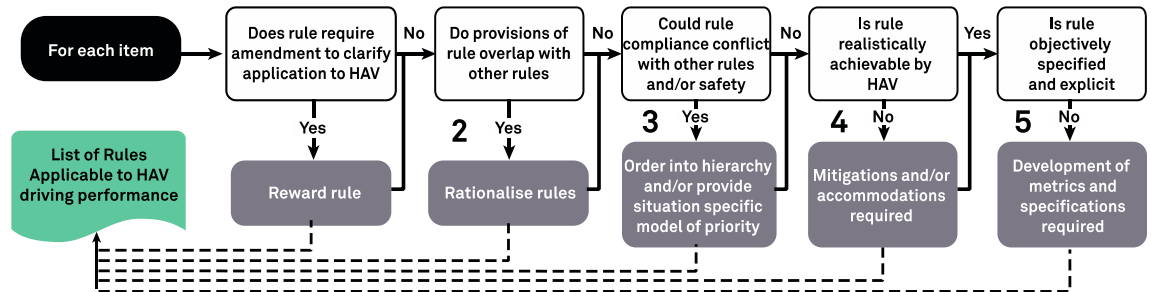
It'd be beneficial to address these issues through a codification process. As part of Singapore's technical reference for autonomous vehicles (TR 68 see Section 8.16) they filtered their existing theory of driving references to create a list of rules applicable for HAV. This process, adjusted for CertiCAV context and **Terminology**, is shown in **Figure 13** below.

Figure 13: Filtering an existing code to create a list applicable for HAV driving performance



This only filters an existing source of rules for relevance – to develop a codified and formalised explicit ruleset would require a further development process such as shown **Figure 14**.

Figure 14: Process of developing an explicit codified Highway code for HAV driving performance



This process involves five steps of evaluating each rule and making improvements when applicable (TR 68 partially implemented the first step). Completing all five steps for all rules would create a fully codified and explicit ruleset (assuming the source has full coverage), however this is envisaged to be an impractically challenging task to undertake over a short or medium term. Step 4 involves the challenging subject of ethics and whether it is acceptable to regulate existing users to enable HAVs, steps 3 and 5 involve the potentially even more challenging area of determining explicit specifications/models that determine correct actions involving a large number of complex variables in highly situationally specific scenarios.

However, providing even some partial common codification, for example of the highway code, would have positive impact on CAV industry as it would:

- Provide some quantification and categorisation of rules for analysis and understanding
- Be a reference that facilitates industry/expert discussion, feedback, knowledge sharing, and continual improvement
- Provide a technical reference for ADSEs, remove market uncertainties, provide harmonisation, streamline development work
- Provide a technical reference for authorities/independent assessors, to support assessments and approvals
- Even if mass market HAVs are realised and supported by strong international harmonisation, it would still be necessary to have country specific customisations/approvals to cater for country specific road traffic rules. Having codified country specific codes would help facilitate an effective development and control processes

Codification of The Highway Code will therefore produce rulesets that are more rationalised, organised, and specifically defined, and develop a reference of guidelines and codes that apply to automated vehicles. Any development in this area should be beneficial by better defining specifications and requirements for system developers and providing a reference standard for independent assessors and/or monitoring bodies.

Similar to formal specifications (see **Section 3.2.3**). It would help interfaces with other processes if the Codified Highway Code is formalised, in terms of using a formally defined vocabulary, syntax, and semantics. This could be informed by:

- Existing taxonomy definitions, e.g. PAS 1883 for ODD
- Legislative schedules, from the TSRGD³⁹ for traffic signs
- Relevant labels, such as DDT subtasks, or control actuations.

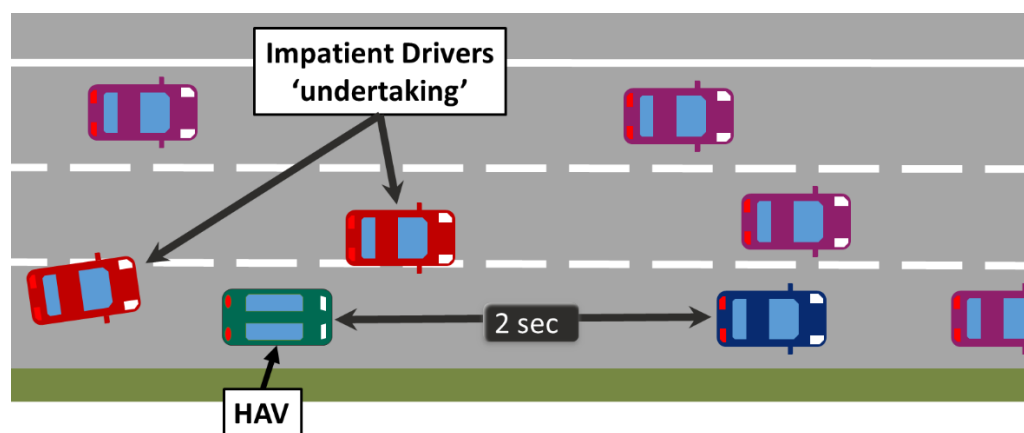
However, detailing a codified Highway Code would also involve tackling several ethical and technical questions, such as:

- If HAVs follow a set of safety principles in their behaviour would this leave them open to be abused by human drivers who know the HAV will give way to avoid a collision? To what extent should HAV behaviour account for the possible bad behaviour of other road users?
- To what extent should behaviour be dictated by rules, opposed to risk assessments?
- Should the development be limited to the direct impacts of the HAVs behaviour or should more global impacts be considered.?
- How should the compromise between risk and traffic/journey performance be addressed?
- Is a system of regulation that performs a little better than existing acceptable, or should more effort be put in to achieve a substantially better system?

As an example, consider a HAV driving on the outside lane of a 3 lane motorway:

- The HAV is following safe behaviour guideline of leaving a 2 second gap to the blue car it is following.
- However, the HAV is surrounded by many impatient aggressive drivers, who are driving with a shorter headway
- As a consequence of the HAV being an outlier by leaving a large following gap, it results in the red cars performing undertaking manoeuvres. This increases total risk and causes a 'shockwave' of braking to follow behind the HAV.
- The HAV could decrease total risk and improve traffic performance, by reducing its following gap to below the recommended 2 seconds. However, is it justified to contravene recommended driving practice and increase risk to the HAVs occupants?

Figure 15: Impatient motorway drivers



³⁹ (The Stationary Office, 2016)

It should be noted that the discussion on thus far is limited to the adaptation of an existing ruleset. However, this existing ruleset has evolved over time as suitable for regulating human drivers and is likely to have gaps when applied to HAVs. The identification of gaps to that need to be filled could include:

- Identifying principles/objectives that HAV performance needs to embody and representing these in new rules
- Creating rules for resolving ethics or logical resolution of rule conflicts. For example, to define collision responsibility
- Creating rules to cover identified gaps in rules. For example, to cover an automated driving concept that is adopted by industry, such as RSS which is not represented in existing rules. Developing rules to regulate an identified industry functional need. For example, if a new type of HMI is standardised for HAV's to communicate gestures to pedestrians, rules may be required to harmonise their use

3.4.5.2

Risk Modelling

It is typical for automotive risk assessments to adopt the model embodied in Functional Safety ISO 26262 (2018). This model decomposes the risk of traffic situations into Exposure, Controllability, and Severity factors where:

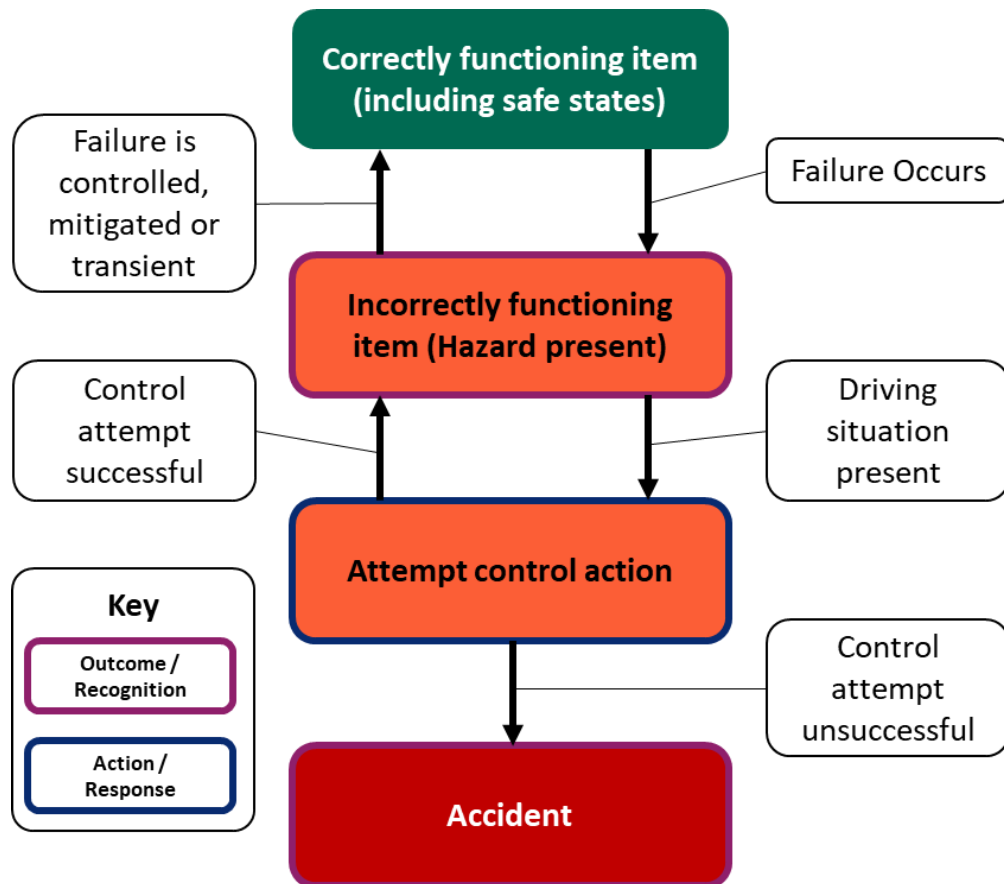
- Exposure (E): Is the likelihood of being exposed to a hazard situation
- Controllability (C): Is the likelihood of being able to act to avoid the harmful outcome of the hazardous situation
- Severity (S): is the level of harm that may occur in the event of a harmful outcome of the hazardous situation

Accordingly, the risk (R) for a hazardous scenario is calculated by:

$$R = E.(1-C).S$$

This is illustrated in the state machine model shown in **Figure 16** below, which shows the model as defined in section 10 of ISO 26262.

Figure 16: Functional safety, State machine model of automotive risk (ISO 26262, section 10)



However, it is important to consider the context of ISO 26262 and intended use of this state machine model:

- Functional safety assesses risks associated with malfunctioning systems; risks associated with driving behaviour or external hazards are not covered
- ISO 26262 is not intended to be applied to fully automated driving systems; it is assumed that a human maintains responsibility for the driving task. The four categories of 'Automotive Safety Integrity Level (ASIL)' applied by ISO 26262 to control the safety integrity of components of the vehicle system may not have sufficient granularity to control the safety integrity of higher complexity ADS systems

Broadly, automotive safety features can be classified as follows:

- Passive safety features are those that only become active during a collision, and work to provide protection and mitigate the severity of the collision
- Active safety features are those that are always active and work to avoid getting into a collision. Examples include anti-lock braking systems (ABS), automated emergency braking (AEB), and lane departure warning (LDW)

Passive safety features therefore target and reduce the severity component of risk scenarios. Examples include air bags, seat belt pretensioners, and vehicle crash protection.

The management of risks associated with route planning decisions could also be considered as an area of strategic safety. This could involve developing an understanding of hazards associated with routes and relevant attributes such as time or day and/or traffic conditions and implementing fleet management features to guide HAVs on less risky routes.

However, only the areas of tactical and active safety are directly influenced by the performance of the DDT by the ADS and therefore the focus of consideration in CertiCAV.

There are two generations of active safety features. The first generation actively monitors vehicle systems, examples include anti-lock braking systems and electronic stability control. The second generation actively monitors external conditions, examples include lane departure warning and autonomous emergency braking. In both cases these target the controllability component of risk scenarios by improving vehicle traction/stability, or warning/taking action when a hazard is present.⁴⁰

Therefore, the current application of passive and active safety features does not cover the exposure to hazard scenarios. Features of Level 2 Advanced Driver Assistance Systems (ADAS) and below are designed not to interfere with the human driver. Accordingly, when risks are assessed the exposure to hazard scenarios is input as an uncontrolled quantity caused by the human driver and external conditions which allows the functional safety risk to be applied. However, developers of higher level ADS that are taking full responsibility for the execution of driving task would need to implement features that adequately control exposure to hazardous scenarios. We term these as 'Tactical Safety' features to distinguish them from features targeted at controllability. Whilst active safety involves detecting the presence of hazards and taking control action to mitigate them, tactical safety involves forecasting hazardous conditions and taking pre-emptive action to prevent them from occurring.

In **Figure 17** the CertiCAV state machine risk model is presented which incorporates the control of hazard exposure.

Tactical safety systems are therefore:

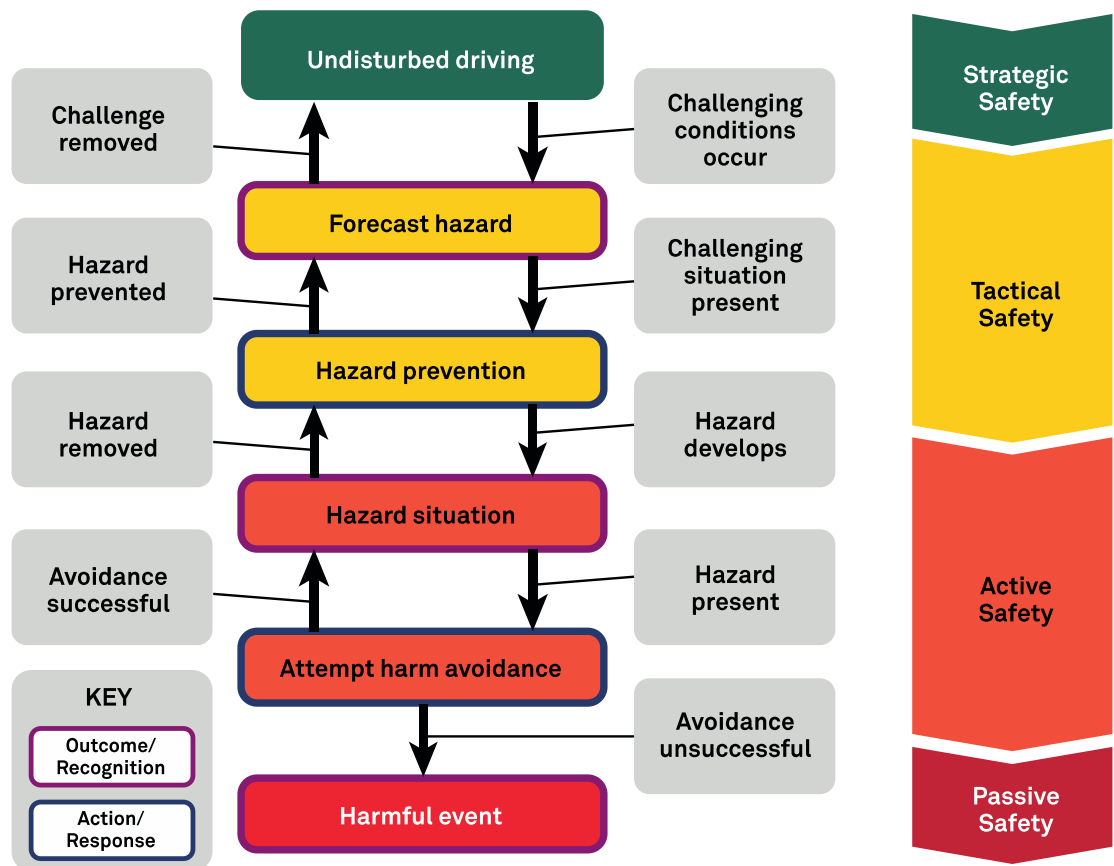
- Systems that act to reduce exposure to hazards

These contrast with active safety systems, that are systems activated in response to hazards.

Tactical safety may therefore be considered as the implementation of defensive driving to manage exposure to hazard scenarios. It is necessary to incorporate Tactical Safety into the risk assessment of automated driving features in order to facilitate the development of ADS that are intended to take responsibility for executing the driving task in place of a human driver.

⁴⁰ (Schöner, 2020) (Schöner, 2020)

Figure 17: CertiCAV state machine model of driver risk



A potential avenue of regulating/ assessing HAV safety would be to specify/ evaluate that performance of both Tactical Safety and Active safety is above a minimum limit:

- Does the HAV demonstrate sufficient Tactical Safety in on-road trials and simulation testing e.g. based on observed rate of potential/ slight conflict situations, Hazard perception and response
- Does the HAV demonstrate sufficient Active Safety e.g. based on quality of dynamic performance in simulation and proving ground hazard scenarios

This may be convenient in characterising risk in a similar way to insurance providers who have a mutual interest in validating safety performance and may have relevant datasets for evidencing appropriate domain specific minimum limits.

3.4.6 Probabilistic systems

With reference to the common definition of ADS sub-systems (see **Figure 6**), the live dataset which is crucial to tactical decision making is created by sub-systems that are probabilistic in nature and typically involve use of AI. This gives rise to the need for processes to cover objectives including the following:

- Safe specification and use of perception systems input to driving features (SOTIF)
- Certainty modelling, reporting, management
- Learnt system validation: explaining/ characterising expected performance and testing resilience
- Definition and appropriate use of AI systems including online and offline learning systems

Whilst it is established that there is a lack of testing infrastructure for sensors, the specification of such facilities is dependent on how industry specifies, validates, and manages the safety integrity of sensors. This indicates a priority to develop the technical framework, including characterising sensor performance, sensor models for simulation, validation testing and supply chain management before developing physical testing infrastructure.

3.4.6.1 Safety of The Intended Functionality (SOTIF)

The validation of probabilistic perception systems for road vehicles is commonly accepted to be covered by the application of the ISO/PAS21448 SOTIF standard. This covers objectives including:

- Validating the safe specification and use of perception systems input to driving feature
- Discovery and reduction of known and unknown unsafe scenarios
- Assessment and validation of false positive and negative system response

We have however identified the following gaps in SOTIF when applied to HAV context:

- Currently only the objectives of ISO 21448 are normative, the V&V processes and methodologies of SOTIF are informative only. Further development would be required to produce a normative conformance standard
- SOTIF discusses the need to identify and reduce known and unknown unsafe scenarios. However, the context is application to SAE level 0-2 driving automation systems where false negative system response is acceptable due to a human remaining responsible for the sustained performance of DDT. Further development would be required to provide approach for HAV context where acceptable thresholds for both false positive and false negative system response is low

A variety of top down or bottom up methods may be applied to validate and verify the intended functionality, and HAV behaviour; however, Systems Theoretic Process Analysis (STPA) is considered particularly suitable for the identification of software hazards and verifying congruence between intended functionality and HAV behaviour. This is relevant to the objective of SOTIF to certify the system behaves as expected – application of STPA to testing and validation is discussed in Section 3.4.9.3, an overview of STPA application to HAV is discussed in Section 8.6.

Furthermore, as the complexity and agency of automation increases to facilitate HAV this is likely to uncover new challenges that SOTIF does not cover.

3.4.6.2

Machine Learning and AI systems

HAVs typically employ probabilistic and/or AI systems as part of their perception and some also for their decision making/control systems. HAVs use artificial intelligence (AI) - machine learning (ML) algorithms at different stages of the vehicle's decision control loop. However, ML algorithms haven't yet evolved to address questions of safety verification, interpretability or *explainability*, and performance. Existing research and engineering efforts exist to answer these questions.

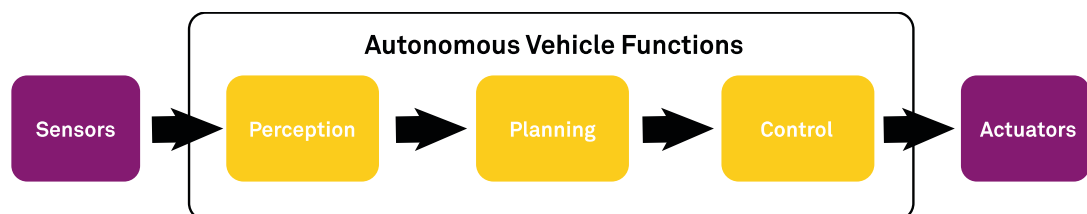
The probabilistic and complex nature of these systems present challenges that require development of new techniques and processes for managing and validating their performance.

To better understand AI's role in a HAV, it is important to recognise its use cases. Every autonomous vehicle function, as illustrated in **Figure 18**, has three basic functional blocks:

1. Perception: perceiving agents or features in the environment e.g. pedestrians, road signs, traffics etc
2. Planning: understanding the driving behaviour required to implement a driving goal given the information from the perception system
3. Control: route creation, driving manoeuvres identification and execution e.g. lane change, overtake, parking etc

AI may be implemented in any or all of these three functional blocks, each with its own challenges. ML is often used for vision (object detection and classification), sensor data fusion, driver monitoring, driver replacement, adapting operating parameters in the vehicle powertrain, anomaly detection and broadly for on-road vehicle motion control. There are a varied set of algorithms that can be used. Some classes of algorithms of important mention include deep neural networks (DNNs), reinforcement learning (RL), Bayesian networks, decision trees and clustering algorithms. To clarify the distinction between use of the terms ML and AI. AI is the broad area in computer science of general purpose algorithms that solve a task. While one may design specialized algorithms for specialized tasks, in AI, for all problems of a given type (such as planning, reasoning, language translation, route selection and discovery, constraint satisfaction), the same algorithm may be used to solve any problem of the same type. These algorithms are in some way intelligent because they can reason about a wide variety of problems, and more specifically scale well to large problem sizes. ML is a class (a subset) of AI algorithms used for learning tasks, including prediction, pattern discovery, correlation discovery and more broadly function learning. The idea is simply that ML algorithms may ingest large amounts of data, and, on their own, learn and take decisions on data they had not been exposed to. It is in this context that when ML algorithms are used in safety critical applications (such as in autonomous vehicles) their failure to predict a decision or outcome correctly could be catastrophic.

Figure 18: Generic functional process flow of autonomous vehicle functions



ML concerns

The aspect of ML algorithms that is of special concern is their transparency. The level of encoding of learned knowledge determines how easy it is to assess the safety and correctness of these algorithms.

Studies exist that describe the impacts ML has on the safety standard ISO 26262 and recommends how the standard may be adapted in the context of the ubiquitous use of ML in the hardware/ software architecture of HAVs.

One of the core safety issues is that these algorithms learn on a finite-size training dataset. Despite there being error metrics to assess the accuracy of learning, it is nonetheless only an estimate of the algorithms in-practice true error rate. This is because the error metrics are based on finite data available pre-deployment (such as a validation dataset, or even the training data itself), and therefore not fully representative of situations encountered when deployed. This then makes it clear that the quality of learning is strongly coupled with the quality of the dataset used for training. A dataset that is not sufficiently varied will lead to over-fitting and the algorithm not performing well in situations it hasn't seen before. In the context of HAVs, putting it simply, training over billions of driving miles is not sufficient. The quality of the miles used for training is important; achieving sufficient coverage of the different scenarios a vehicle is likely to encounter in the real world is important when choosing a training dataset.

An upcoming field in the area of Safe AI is the Explainable AI (XAI) concept, which is being championed by DARPA, USA. As the XAI concept, itself is very new and developing, its use in the HAV domain is still being established and defined. However, as AI-Based applications in HAVs will have safety critical implications, there is a strong expectation to have a level of interpretability and explanation of the results. While the XAI concept is gaining traction in the research community (especially for perception functions), another approach for safety of AI-based systems is a relatively unexplored concept (especially in an HAV domain) of "safe learning". Learning is a key part of an AI-development process. However, depending on the intended functionality, the type of learning needed for the system may differ. In a HAV, AI systems may be part of perception, planning or control (or a combination of them) functions. While AI systems in perception will have a classification functionality, AI-systems in planning or control will have a behavioural functionality. Thus, we divide the research problem of Safe AI for:

1. Perception
2. Planning and control.

The case for Formal Verification

A trained model must be assessed for its adherence to safety and correctness rules, violation of which could be catastrophic, and is therefore considered a failure. The question is whether there exist scenarios that can lead to failure of the system. While simulating the learned model with a list of test scenarios is possible, to discover failure scenarios, this is dependent on human expertise for concrete scenario generation. Furthermore, it is important to understand that an inability to discover a failure scenario via simulation does not guarantee the non-existence of a failure. This is because, the number of scenarios a vehicle may encounter are theoretically infinite, and simulation would explore only a finite number among these.

Formal Verification and AI/ML

Formal verification aims to provide safety and correctness guarantees for mathematical models against a specification. The advantage of formal approaches over simulation-based approaches is that the former can statically examine all possible behaviours in the model under observation, while the latter is restricted to exercising only explicit test-cases. Existing formal verification methods suffer from

scalability challenges, and hence complex analyses must be hierarchically decomposed into simpler ones, and an assume-guarantee framework could be used to hierarchically verify complex systems.

The ease of formal analysis is tightly coupled with the explainability of the learned model. For instance, it is easier to assess the correctness of decision trees in comparison to neural networks. This is because decision trees, by their nature, are explainable. They clearly specify data attribute boundaries used to make decisions. On the other hand, in a neural network the decision function is usually a complex polynomial or non-polynomial equation in terms of the neuron functions, bias values and network edge weights. Due to the ability of neural networks to learn (approximate) highly non-linear functions, they have been widely used in AV applications. Formal analysis is presently restricted to problems that can in some way be represented symbolically. This includes applications of AI, and more specifically ML techniques, in perception, planning and control. The reason we combine AI and ML at this stage is that symbolic reasoning engines used to reason and verify decisions at applicable stages in the AV architecture are pre-designed, defined, and not learned from data.

The existing work applying formal analysis as an avenue for safe learning follows one of three directions, given safety specifications (the richness of the specification can vary across studies), (1) given a learned model prove that it is safe, (2) given a learned model provide a safe wrapper that prevents it from doing anything bad (with respect to specification), and (3) learn the model safely. A selection of the current state of the art in formal verification for neural network, and reinforcement learning models is given below:

- **Neural Networks:** Methods and tools have been developed to verify feed-forward neural networks and some classes of DNNs. Studies also exist to provide robustness certificates for DNNs quantifying how well they perform against input perturbations. Efforts have also been made to scale formal verification of neural networks using pruning techniques, where parts of the DNN's structure are pruned to achieve quicker outputs, while maintaining the model's accuracy. Neural networks are famously used in perception, where vision techniques decipher objects in images, and control, where a control function (such as engine or braking control) is performed in a feedback loop. Tools such as Sherlock and ERAN exist for verification of feed-forward neural networks
- **Reinforcement Learning (RL):** Methods for the probabilistic verification of Markov Decision Processes (MDP) exist. Therein, given a formal requirement, expressed in temporal logic, the probability of violating the property is determined. Later in the document we expand on how requirements expressed as formal constraints can be used to learn MDPs that adhere to requirements. RL techniques are used in the phases of planning and control in AVs, where the sequence of actions taken by the AV must be safe

It is important to take deliberate steps in ensuring safety of machine learned components. While formal verification techniques are being developed for safe learning at the component level, in terms of scalability, at the system level they have a long way to go. A conservative approach would perhaps use formal verification techniques wherever possible to identify hazard and failure scenarios that are input to simulation and real-world testing. This would be beneficial in systematically uncovering risks, and thereby providing increased confidence in the safe use of AI.

Explainable and Safe AI

In the past, studies in ML (and AI) were focused on developing algorithms and models for systems, not necessarily in explaining how these models behaved. This is mainly attributed to access to cheap

computing power and the advances in machine learning. Explainable AI deals with AI systems (ML and beyond) that learn from data and take actions, explaining their reasoning in a manner that is easy to understand and can thus be validated, thus improving the trust and transparency in complex AI systems. Causal reasoning is central to explainable AI. For instance, a model that identifies cars on the road from images must be able to list causes for the conclusion it makes; where the car is located in the image, and what attributes of the image it uses to reason that the image contains a car. Ontologies describing the operating environment of an interpretable AI model, together with a causal reasoning framework would make for a model that is easily explained.

Broadly there are two types of AI algorithms/models. White box (intrinsically interpretable) models and black box models (knowledge is encoded in an obscure form). Black box approaches for explainable AI algorithms and models may be further classified into two categories: (1) Model Agnostic Methods, and (2) Example-Based Explanations.

White box AI approaches

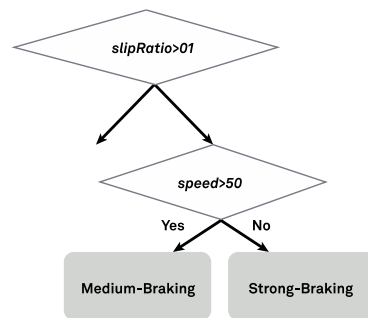
Linear regression and logistic regression may be considered white box approaches because the form in which associations are made provides transparency in the direction of association between attributes and the outcome conclusion. These methods, however, can't be applied when the relationship between attributes and outcome is non-linear. In such cases decision trees succeed.

Decision trees are some of the most widely used learning techniques in ML. With decision trees the dataset is split multiple times, each time on an attribute and a cut-off value. Splitting ends at the terminal nodes of the tree where data is homogenous with respect to the outcome and an explanation for the outcome can be derived from the path, from the terminal node to the tree's root node.

The Decision rule list and Bayesian rule list structure are a pair of similar models, the former being non-probabilistic and the latter probabilistic, that are explainable through a list of "if then else" rules. Decision trees and rule lists are inefficient for describing linear relationships. Additionally, non-linear relationships are effectively represented using step functions dictated by the conditions in the nodes of the decision tree (or the if-else of decision rules). An example is shown in the figure. As a result, a small perturbation in the input could result in a value-jump in the prediction (such as medium braking over strong braking). Nonetheless, decision trees are popular due to their intrinsic explainability, which is why they are often chosen for use for planning and control in AVs. Even though their expressiveness is limited, provided that learned decision trees produce an explanation (which is explicit in its structure) and can hence prove to be safe, it makes for a good candidate for safe learning in safety-critical tasks.

While white box approaches have significant advantages with respect to transparency and interpretability, they are not always employed. The core reason is that they are limited in their expressibility and accuracy. Linear regression for instance is only applicable when the relationship between attributes and the outcome is linear, while decision trees and rules are best suited for situations where the relationship can be expressed using a set of conditional rules. The result is that for more complex non-linear relationships, for a decision tree/rule-based model, a slight change in the input can drastically change the outcome, and this can be undesirable. Black box approaches are usually more powerful and can represent arbitrary functions and are more robust to input noise.

Figure 19: example decision tree



Black box explainability

Well known black box approaches include approaches such as neural networks (NN), deep neural networks (DNN) and Markov decision processes (MDP). NNs and DNNs are specifically useful in perception and studies are now being undertaken to use DNNs in control applications. MDPs are traditionally used in planning and control.

There are model agnostic methods for explainability that aim to explain outcomes of models using various methods of approximation. The advantage being that a single technique is usually applicable to different learning models. One of the oldest of these is the Partial Dependence Plot (PDP) which plots a chosen feature (or two) against the target outcome. For instance, one would plot speed of the vehicle with the likelihood of an accident. Individual Conditional Expectation (ICE) is like the PDP but plots the feature in every instance of the dataset against the predicted outcome. The PDP would be the average of all the instance plots. The issue with PDP and ICE is that they don't consider the fact that features in a dataset may be internally correlated, artificially creating unrealistic relationships that do not occur in the data. Accumulation Local Effects (ALE) plot attempts to solve this but is nonetheless restricted to considering at most two features. Another method, that of feature importance determines how relevant a feature is in determining an outcome. This is achieved by examining what happens when the features values are changed by permuting or shuffling values around in the dataset. The prediction error is then used to determine the importance of the feature. If shuffling values increases the error, the feature is important, otherwise it isn't.

Surrogate based approaches use inherently explainable models (like decision trees) to approximate the black box model by using the dataset along with the prediction made by the black box model. The benefit of doing this is that we achieve explainability in an otherwise obscure model, while the limitation is that, because black box models are usually more accurate, the prediction by the twin white box model is usually less accurate. Such an approach is promising since it provides explainability where explainability otherwise doesn't exist, however to what extent an equivalence can be established between the black and white box twins remains to be determined.

Providing an example-based explanation is an alternative option. Such an approach uses instances from a dataset to explain the models reasoning. Some such techniques include the k-nearest neighbour model (explains using examples in the neighbourhood cluster), counterfactuals (explain by showing examples in which changes in attribute values flip the outcome), using prototypes (representative examples) and criticisms (non-representative examples). Counterfactuals seem to be the more reliable among the approaches. A counterfactual better represents how a person thinks. For instance, the counterfactual for the explanation "the car braked strongly because a pedestrian was detected less than 10 metres away from the path of the vehicle", would ask what happens if a pedestrian is either not detected or more than 10 meters away along the vehicle's path. If the answer is

that the vehicle doesn't brake, then this strengthens confidence in the explanation. In implementation, the inputs in the constraints representing the explanations would be inverted, and examples from the dataset would be chosen to illustrate the counterfactual. These instances would represent circumstances where some aspects of the input attributes are changed, and thereby how the outcome prediction changes. Symbolically for a conditional "A $\rightarrow$ B", such approaches would look for situations where A is false (A doesn't happen – via some perturbation in attribute values forming constraints in A), and provide them as an explanation, with the expectation that in those instances the event B is false (B doesn't happen).

Learning safely

For some black-box approaches, it is also possible to approach the learning problem by keeping safety at the centre of the learning process, we call this safe learning. What this means is that in lieu of learning a model and then proving the learned model to be safe, one would ensure that the process of learning is itself safe, guaranteeing that the learned model is safe by construction. Present studies in safe learning are on MDPs and are therefore limited to planning, with studies only beginning on applications to control.

In the paradigm safe learning, requirements are formally encoded into constraints expressed in a logic language (common examples include linear temporal logic (LTL) and signal temporal logic (STL)). For a logic language encoding, a quantitative measure is determined to express how well the model meets or fails to meet the requirements. For instance, for the language STL, a measure can be defined indicating a degree of robustness with which the requirements are satisfied. Here zero or negative values indicating failures, with the more negative indicating more catastrophic failures, while positive values indicate that the model meets specification, with the large values indicating that requirements are satisfied with high levels of confidence and robustness to noise. The latter being desirable, safe learning aims to learn the model while optimizing the satisfaction or robustness measure, thus assuring the models adherence to requirements.

A decision strategy can be learned in a way that does not violate a set of properties specified in STL. System dynamics are learned using DNNs. The DNN model is used to predict system states. A Markov Decision Process (a learning model for learning best actions) is then learned, using the DNN to predict future states, in such a manner that maximizes the robustness of the specification in STL.

Alternate approaches exist for pruning and shaping MDPs to satisfy safety criteria. These approaches use requirements to first generate a model policy that describes only permitted behaviours. At each system state a subset of all actions (only those that are permitted by the requirements) can be taken. Although a variety of action sequences (the control policy) may be permitted, many of them would be undesirable from a performance or quality of service perspective. This is achieved by iteratively constraining the model, and pruning sets of action sequences to maximize the models performance. Here performance is defined using a reward/penalty policy. The MDP consists of a set of states (representing the situation) and transitions (representing actions), with rewards/penalties associated with being in a state and/or taking an action. Rewards/penalties may also be associated with a sequence of state-action pairs. A policy determines how control moves through the states in the MDP. Constraint-based learning uses constraints that associate a cost with movement in the MDP. The constraints impose an upper bound on the cost incurred which implicitly constrains the actions that can be taken. The policy that is learned will then be constrained to have state-action sequences that adhere to requirements (expressed using the cost function) and thus is implicitly safe.

Summary and Conclusion

Approaches for learning in safety-critical application of AI algorithms, and more specifically ML, are broadly of two types; those where learning is done safely (such as constraint-based learning of MDP control policies) and those where the model can be validated through explanations or formal verification. For now, the prior (safe learning of MDPs) is applicable in the areas of planning and control of AVs while the latter (explainability of learned models) can be applied to vision and perception. Furthermore, explainable AI approaches are also useful in situations where neural networks (a black-box approach) and decision trees (a white-box approach) are used for decision making in planning and control. For the latter, an AI reasoning engine taken together with rules explaining how the learned model behaves would be used to validate the models' reasoning.

3.4.7 Sensors & Co-operative safety

With reference to the common definition of ADS sub-systems (see **Figure 6**) this includes environment and cockpit sensors as part of the perception sub-system and a data communications input.

Practice is required to ensure the suitable specification and use of sensors and data systems including the following objectives:

- Specification of sensor systems (e.g. exteroceptive and cabin) including suitability for intended functionality for example range of view, refresh rate, precision (related to **ADS functionality** see Section 2.4.3)
- Characterisation of sensor performance, for example based on environmental conditions
- Safe concept of operations: In particular the involvement of any safety critical systems that are external to the HAV should be described and controlled (related to **HASS architecture** Section 2.4.4)
- Assessment and management of sensor faults (related to functional safety see Section 3.4.4)
- Performance monitoring/management (related to operational safety see Section 3.4.3)

The modelling and testing of sensor systems is a complex and rapidly developing area. A feasibility study was published by NPL and CPC in July 2020⁴¹ includes appendices with examples of attributes that characterise various sensor types. The findings are summarised in Section 8.11 of this report. Of key relevance is the conclusion that a priority area is developing the technical framework for how industry specifies, validates, and manages the safety integrity of sensors including characterising sensor performance, sensor models for simulation, validation testing and supply chain management. This technical framework would then inform the specification of testing infrastructure for which there is also an industry need.

The role of exteroceptive sensors is particularly important as these provide situational awareness (e.g. scenery for localisation, actor position/pose for tactical decision making). Cabin sensors influence tactical decision making as the state of internal occupants/cargo affects interpretation of rules and risk assessment. For example, a HAV bus could drive at a slower speed and or with smoother acceleration when it has standing occupants - to reduce the risk of them falling. Similarly, the HAV bus could stop with a smaller safety clearance from a pedestrian who has stepped out in front of it, to balance the risk

⁴¹ (Connected Places Catapult and National Physical Laboratory, 2020)

of its standing occupants falling. A HAV passenger car could delay pulling away from parking space until all occupants have fastened their seatbelts.

Another area of key importance is that of co-operative safety which has the potential to provide valuable data for promoting safe and efficient HAV behaviour. However, care must be taken that the way co-operative data is utilised does not create functional or security related safety hazards (see the following Section 3.4.7.1).

3.4.7.1

Co-operative safety

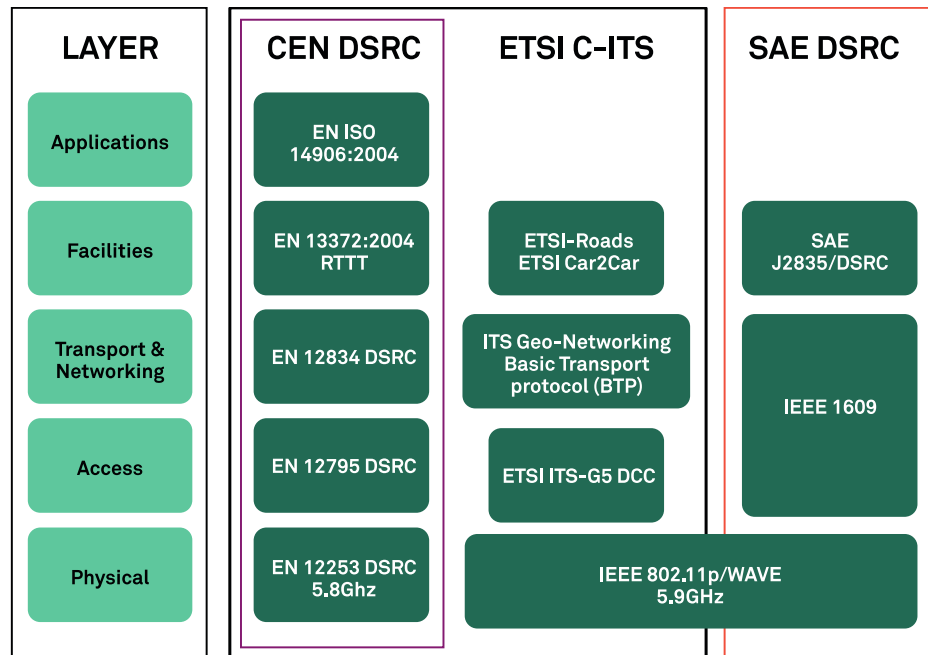
Although out of scope for CertiCAV the need to develop practice for co-operative safety is recognised.

There is a significant European technical framework for Co-operative Intelligent Transport Systems and services (C-ITS) for both Wi-Fi (ITS-G5) and cellular (C-V2X) physical layers. ETSI C-Roads and Car2Car frameworks define services utilising vehicle to infrastructure and vehicle to vehicle communications, these define so-called day 1, and day 1.5 services cover and driver information/warning services and co-operative awareness/data exchange to support partially automated features. Further development is required on how to apply existing facilities safely and effectively to services for HAVs, and what new facilities/services are required to support HAVs. Traffic authorities and system developers need more clarity on what responsibilities and requirements must be satisfied to implement C-ITS safely. Traffic authorities need an understanding of what infrastructure and services they'll need to provide and maintain to support HAVs.

For instance, the C-Roads technical framework defines a Signal Phase and Timing Extended Message (SPATEM) and topology Map Extended Message (MAPEM) facilities and a service for providing Green Light Optimal Speed Advisory (GLOSA). The existing framework requires further development to consider whether the SPATEM and MAPEM facilities are sufficient for messaging traffic signal and topology data to support HAV use-case and define practice/specification on how these facilities may be implemented and used safely. As an example, consider the case of a traffic authority providing a roadside I2V unit that communicates traffic signal data to HAV.

- If the HAV architecture uses the I2V data to detect red signals the data is safety critical, therefore the service provider would need to ensure both the availability and safety of the data e.g. free from false negative and cybersecurity threats
- If the HAV architecture uses the I2V data to parity check red signal detection from on-board camera sensor the data is not safety critical, therefore the service provider may need to ensure a level of availability but would not need to ensure safety of the data

Figure 20: Illustration of Wi-Fi based C-ITS technical framework layers



3.4.8 Data management

There is need to develop data formats standards to support industry harmonisation and facilitate assurance processes including independent assessments, monitoring activities. This continues to be a rapidly developing space, for instance involving standards activity regarding monitoring data.

Objectives to be satisfied include:

- Data filtering and evidence retention to justify safety arguments
- Collection of monitoring data from physical use of HAV to support system validation, and monitoring activities
- Transmission of data to defined standards between organisations.

The initial set of OpenX standards that are being developed by ASAM (including those under development) derive some content on formats for data to support assurance processes including OpenDrive, OpenCRG and OpenSCENARIO, later additions to this series are OpenXOntology, OpenLabel, OpenODD. These are particularly relevant for virtual testing, however, are also important for defining common standards to support scenario definition, and physical testing telemetry.

3.4.8.1 OpenDRIVE

The OpenDRIVE standard defines a file format for the precise description of the road network, its usage is mainly for simulation applications. The (at time of writing current) OpenDRIVE 1.6 schema includes road geometry description, surface properties, road markings, signposting, and logical properties such as lane type and driving direction. OpenDRIVE file format has been widely implemented by several simulator tools including VTD, CARLA, OpenPass, and Esmini. It can be used to generate the road topology within the simulators; however, the current OpenDRIVE schema does not cover many potentially relevant fixed road structures such as buildings, lamp posts, etc. The ASAM community are investigating the possibility of including more scenery elements in future OpenDRIVE versions.

3.4.8.2 OpenCRG

OpenCRG is a file format standard for the description of the road surface, it can be used to describe the precise road surface elevation data as well as other surface properties such as friction coefficient or grey values. These attributes are relevant for vehicle dynamics, and potentially relevant attributes of sensory conditions for exteroceptive sensors as vibrations may affect their performance. OpenCRG is targeted to be used for endurance simulation, driving simulation and traffic simulation. OpenCRG can be referenced within the OpenDRIVE file for additional road information.

3.4.8.3 OpenSCENARIO

OpenSCENARIO is a file format standard for the description of dynamic content for driving and traffic simulators. The current release of OpenSCENARIO, version 1.0, is in .xml format, whereas the future version 2.0 is targeted as a domain specific language with its own syntax and semantics. Since OpenSCENARIO 2.0 is still under development, we focus on OpenSCENARIO 1.0. OpenSCENARIO's main use case is to describe complex manoeuvres of the traffic participants such as vehicles, pedestrians, it is capable of describing manoeuvres using atomic input values as well as using a pre-defined trajectory. An OpenSCENARIO file structure consists of parameter declaration, index to catalogue, index to external files, definition of scenario object, manoeuvre story board. The story board can be further divided into initialisation, acts, sequences. Parameter declaration enables user to pre-define parameters and assign names to them for later reference/usage. The catalogue will host the properties of: vehicle, pedestrian, and MiscObjects; and serves as the model for instantiate the corresponding objects. For example, properties for vehicles could include drag coefficient, maximum acceleration/deceleration, number of gears, dimensions etc. The OpenSCENARIO file will reference to the catalogue when defining a dynamic object. Indexing to external file could include files such as OpenDRIVE or OpenCRG. The definition of scenario object will then instantiate traffic participants using the catalogue and assign them with unique name/ID, the traffic participants names will be used during the story board description. A story consists of acts and sequences. Acts are triggered at a specific condition such as when speed or distance equal to certain values, acts contain manoeuvre groups. A sequence consists of the manoeuvre groups of one or more traffic participants, manoeuvre groups consists of a list of manoeuvres which the traffic participant will perform.

3.4.8.4 OpenXOntology

OpenXOntology is a work-in-progress standardisation project with the aim of providing common ontologies (definitions, properties, and relations) to be used across ASAM standards. Within the OpenXOntology, work has been carried out at different levels: core, domain, and application. Core ontology provides a list of high level concepts (such as physical objects, activities, events etc) which are not unique to the road traffic domain, it enables reasoning in a four dimensional space (x,y,z,time) by including the temporal properties. Domain ontology consists of all the objects, activities and abstract concepts that are within the road traffic domain, examples including vehicle, traffic lights, manoeuvre etc. Application level ontologies are unique to specific applications, taking OpenSCENARIO as an example, the corresponding application ontology will be constructed according to the OpenSCENARIO file format. Examples of the elements of an OpenSCENARIO application ontology include triggers, storyboard, private actions, and bounding boxes. The idea of initiating the OpenXOntology standard is to allow common core and domain ontologies to be used by various applications, for example an OpenSCENARIO designer would be able to import the base domain and core ontologies and use them in conjunction with the OpenSCENARIO application elements for reasoning. The OpenXOntology

project team is currently investigating OpenSCENARIO 1.0, OpenDRIVE 1.6, the OpenLABEL concept project, ODD standard (BSI PAS 1883) to separate the domain concepts from the application level concepts and consolidate them to form a unified domain ontology.

3.4.8.5 OpenLABEL

OpenLABEL is another work-in-progress standardisation project with the aim of setting up a common labelling standard within the industry. Currently the OpenLABEL concept project has finished, whereas the formal standardisation project has not yet started. During the concept project, the project group concluded the labelling methodology, labelling structure, file format and scenario labelling method. The labelling methodology handles how objects within a scene (a point in time) can be labelled, examples are semantic segmentation or 2D bounding boxes. The labelling structure deal with how objects and their relations are labelled, the current method is RDF triple format, subject-predicate-object (e.g. Pedestrian follow truck). The labelling file format is in .json format. The scenario labelling defines the meta-model within which all the labelling classes are structure. Currently, the labelling uses a three-buckets model, the first bucket includes custom labels which contains organisational information, versions etc. The second bucket contains all the information for defining the operational design domain (ODD), which at a high level consists of scenery, environment and dynamic (macroscopic traffic). The third bucket consists of behaviour labels such as manoeuvres as well as scenario event triggers. For the standard project, labelling will be further divided into object labelling, scene labelling and scenario labelling.

3.4.8.6 OpenODD

OpenODD is a concept project with the aim to define a language to describe the operational design domain (ODD) of ADSs. ODD refers to the operating environment (road type, weather conditions, traffic conditions) in which a vehicle can drive safely. The OpenODD format would enable users to draft, compare, exchange and re-use ODD specifications, it will also have the ability to describe conditional ODD statement, such as motorway is only suitable when there is no rainfall. The OpenODD will enable users to interpret and perform analysis, therefore its attributes are measurable and verifiable. This may be highly relevant for the definition of ODD as part of HASS specification (see Section 2.4.3).

3.4.8.7 In-use data recording

PAS 1882, currently in drafting phase, is intended to provide a reference practice for data recording of HAV in-use. This may have a focus on post incident event investigation.

During system validation, performance metrics providing evidence for the system safety case (see Section 4.2) would be collected. As part of System approval, it may be appropriate for the regulator to access/monitor this evidence – or specify in-use data recording conforming to some other specification. As part of the demonstration part of CertiCAV we have developed a definition of a 'CertiTrace' data format as an example of a data format that would support independent assessments.

3.4.9 Validation and testing

The purpose of this area is to provide the process and tooling to undertake the testing and validation activities for the HASSs. Whilst validation activities are included in other identified practices such as Functional Safety and SOTIF, this process reflects the need to co-ordinate the generation and coverage of test scenarios, and tooling/methodologies to undertake testing. This includes satisfying the following objectives:

- Safety process release judgements, documentation and management of safety anomalies, Hazard Analysis and Risk Assessment schedules, Risk mitigation methodologies
- Apply validation classification levels to all system items and apply corresponding safety controls
- Ensure sequential progression of verification and validation methods from test bench and off road testing before progressing to public road environment or reduced monitoring
- Accredited and Validate testing process and testbench
- Monitor all testing to identify and reduce previously unknown scenarios, and detect issues/ events
- Evidence from V&V activities must be recorded and argue an integral progression of maturity
- Verifying that system performance satisfies system specifications
- Co-ordination of inputs and outputs between various system development processes

3.4.9.1 Validation as a requirement

As discussed in Section 3.3 it may be impractical to fully define regulatory requirements for HAV based on specifications which gives rise to the need to consider how regulatory requirements relating to the definition of needs could be defined and implemented.

This in turn places demands on validation practice to define and manage validation that needs are satisfied. Other industries such as commercial civil aviation do involve statistical targets in technical regulations however their challenges are different. A particular challenge to overcome for HAV is accounting for operating in a high risk multi agent environment, it is not just a case of making sure the behaviour of the HAV by itself is safe, but that is understood by other road users, and sufficiently mitigates the potential unreasonable actions or misuse of others in the road environment. Some issues and approaches to cover this area are discussed in this document for example as part of safety & quality management (see Section 3.4.2) and behavioural safety (see Section 3.4.5).

3.4.9.2 Virtual Testing Management

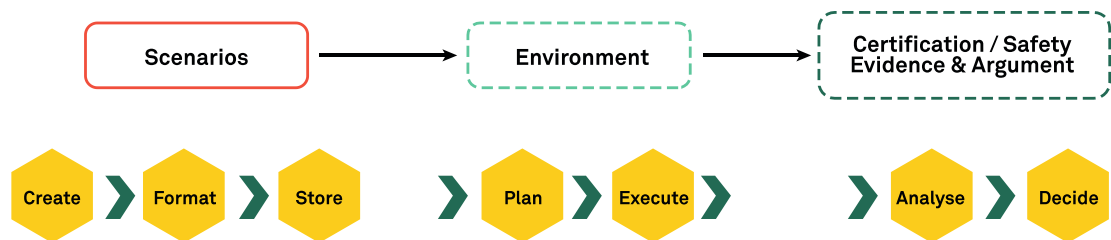
Virtual testing is an important part of system safety management and V&V processes which may be considered to comprise of three key elements:

1. Scenario generation
2. Testbench
3. Evidencing & safety argumentation

Be it at a system level or at a component level, the V&V of each aspect of in the V-cycle will comprise of these three elements. The first step of V&V process will involve creation or identification of scenarios which will need to be formatted in a common/standard language and subsequently stored in a repository to enable scenario sharing and exchange. The second step will involve planning and executing the test scenario in a testing platform (simulation or real-world). A test medium could be a

digital simulation, simulation, test track testing or on-road testing. While each test medium has its distinct role in the V&V process, there exists an evaluation continuum which involves a progression from simulation to real-world testing. As the complexity of scenarios increase along with their associated risk, the aim of a V&V process should be to front load the scenario execution to simulation based testing. This allows us to learn about the system as early as possible while also making it cheaper and less resource intensive. The third step of the V&V process involves analysing the results of the execution of the scenario in a test medium and subsequently deciding if the ADS is safe or not. This involves using the generated safety evidence, making a safety argument and documented in the safety case.

Figure 21: Virtual testing management process



There are therefore three key aspects to consider relating to scenarios:

1. Scenario generation
2. Scenario format
3. Scenario storage

These are discussed subsequently.

3.4.9.3

Scenario Generation

It is widely accepted that there is a need to shift from using "number of miles driven" as a metric to a scenario based testing approach. However, it is essential that the scenarios used as part of the test programme have a potential to reveal shortcomings in the system. A recent semi-structured interview study involving Verification & Validation experts around the world suggested the move to Hazard Based Testing. It found that experts are interested in identifying "how a system fails" as compared to "how a system works". Hazard Based Testing and corresponding test scenario generation could be classified into two main categories⁴²:

- **Data based scenario generation:** This involves using real-world data from accidents or near misses or insurance claim records to identify vehicle or environment attributes that contribute to crashes (minor, major or fatal). For UK based data, one could use the Stats19 data (police records of accidents) or the RAIDS database. It is important to mention that using a data-based approach, implicitly assumes that HAVs will fail similarly to human-driven vehicles unless we have data from CAV trials and deployment. While this assumption will likely not be true for all failure modes, it does provide us with the situations that human drivers find difficult

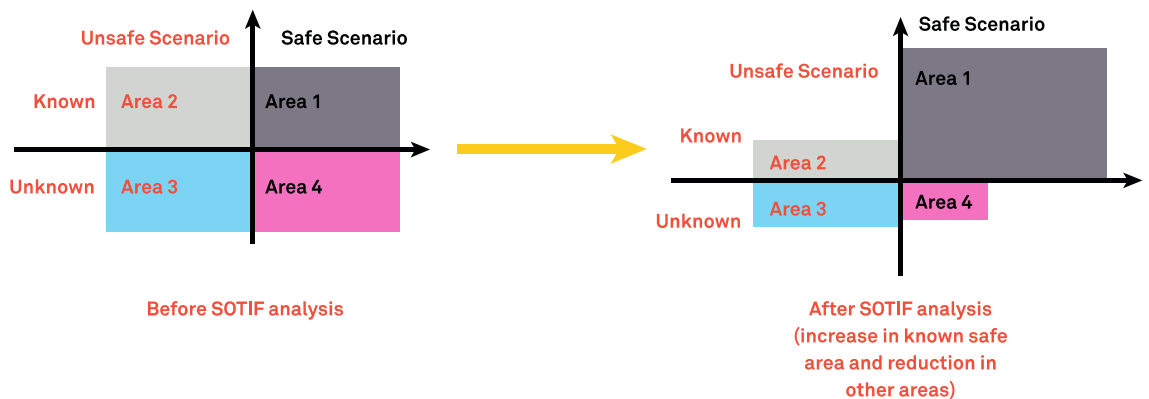
⁴² Brackstone et. al., 2020. OmniCAV: A Simulation and Modelling System that enables "CAVs for All". Proc. of the 2020 IEEE Intelligent Transportation Systems Conference.

- **Knowledge based scenario generation:** In order to overcome the possible limitation of data-based scenario generation (due to the assumption stated above), a second complementary approach is needed. Knowledge-based scenario generation identifies scenarios which will reveal potential failures and hazards based on the analysis of the system architecture. As the aim is to identify failures and hazards, various hazard identification methods may be used, e.g. Systems Theoretic Process Analysis (STPA), Fault Tree Analysis (FTA), FMEA, HAZOP etc. A recent report by Waymo⁴³, it is suggested that STPA enabled them to identify failures and hazards that other methods didn't capture. This may be due to the nature of the STPA analysis which is based on systems engineering and focusses on system interactions

As mentioned earlier, it is essential to focus on the quality of scenario generation as compared to just the number of scenarios generated. From a safety perspective, this aligns with the quadrant approach as per SOTIF 21448 which suggests that the scenario generation process should be to reduce the Area 2 (known scenarios) and Area 3 (unknown scenarios) while increasing Area 1 (known safe scenarios). Methods like STPA help uncover the missing and incorrect requirements along with their corresponding scenarios.

It is important to highlight that the hybrid approach of data based test scenario generation and knowledge based test scenario generation is key to ensuring alignment with the SOTIF 21448 standard. Some of the other standards that may be considered include ISO 34502⁴⁴ (draft standard).

Figure 22: SOTIF analysis example



3.4.9.4

Scenario Format

In order to enable scenario exchange between various stakeholders, it is important to create a common language for scenarios and scenario related V&V artefacts. It is important to highlight the diversity in the stakeholders involved in the HAV assurance process which will impact the type and kind of common language adopted in this area. From a simulation based testing approach, there is a growing consensus in the industry to use ASAM OpenX standards which include OpenDrive, OpenCRG and OpenSCENARIO, later additions to this series are OpenXOntology, OpenLabel, OpenODD. See Section 3.4.8 for an overview of these standards. While these standards cater to simulation execution,

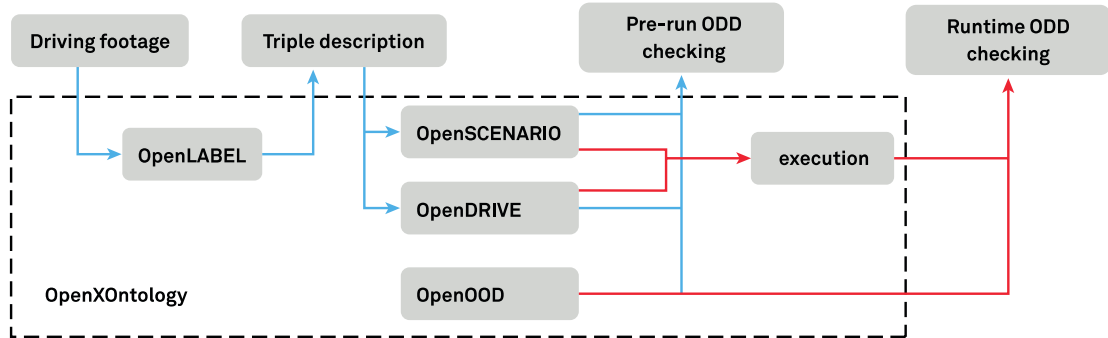
⁴³ <https://storage.googleapis.com/sdc-prod/v1/safety-report/2020-09-waymo-safety-report.pdf>

⁴⁴ <https://www.iso.org/standard/78951.html>

there is a need to overlay these standards with a higher layer abstraction format which would be more useful for stakeholders like regulators, policy makers, system engineers etc.

Figure 23 below is an illustration of how the OpenX standards relate to each other, and furthermore an example of how they may be applied to a scenario based testing activity workflow.

Figure 23: Relationship between ASAM OpenX standards



This workflow illustrates the process from receiving sensor data, to converting this data into scenarios for execution, and compare the scenario with an existing ODD specification as part of the analysis, OpenXOntology is used for all the steps. The input data can be driving footage recording, it can then be converted into OpenLABEL data format by using OpenXOntology, the output is a list of triple description equivalent to the driving footage recording. This set of triples forms the abstract/functional scenarios and can be converted into OpenSCENARIO/ OpenDRIVE file format by using ontology-based scenario detailing method. Assuming there is an ODD specification described using OpenOOD format, the OpenOOD, OpenSCENARIO and OpenDRIVE can be used together to carry out a pre-run ODD coverage checking. This is to indicate whether the intended scenario is within or outside the ODD specification. The OpenDRIVE and OpenSCENARIO files can also be executed in a simulation. During the runtime simulation, simulator can output OpenLABEL ground truth message and user can also perform runtime ODD checking when compared with the OpenOOD file format.

This workflow contains a data driven scenario generation approach, similar workflow can also be constructed for a knowledge driven scenario generation, execution by using the ASAM OpenX standards series.

3.4.9.5

Scenario Storage

The third aspect of the scenario stage is the scenario storage. As the type of scenarios and the type of ODDs for which we need scenarios would vary a great deal, it is essential to create a common repository for scenario collection and exchange. This will also allow to prevent re-invention of the wheel by multiple organisations creating scenarios for same ODD or same HAV. The UK has played a leading role in creating scenario database with an initial proof-of-concept developed as part of the MUSICC project. The learnings of the project have helped developed the National Scenario Database as part of the Midlands Future Mobility project. Other international initiatives on scenario databases are also being developed, e.g. Safety Pool, SAKURA database etc. There is a need for international collaboration on the subject of scenario databases to make the process efficient and scalable.

3.4.9.6

Relationship between ODD, desired behaviour and scenarios

As part of the HASS specification (see Section 2.4), a manufacturer will define the ODD and the ADS functionality defines the desired behaviour of the HAV which may or may not include requirements laid down in regulation/codes. This enables the creation of the argument around the evidence for

behaviour safety. One could consider the desired behaviour of the HAV as its behavioural competency. It is critical to ensure that the scenarios tested as part of the assurance process consider both the ODD and the desired behaviour competency.

An ODD definition can be defined from the perspective of an end user (e.g. local authority, insurance provider etc.) or a system specifier/ ADS manufacturer. Similarly, behavioural competency of an HAV may have various levels of abstractions. For example, a manufacturer may claim that it can perform an "unprotected right turn" on "sunny days" only, while another manufacturer may claim that an "unprotected right turn" is not part of its behaviour competency. Behaviour competency describes also identifies the assumptions on the behaviour of other road users.

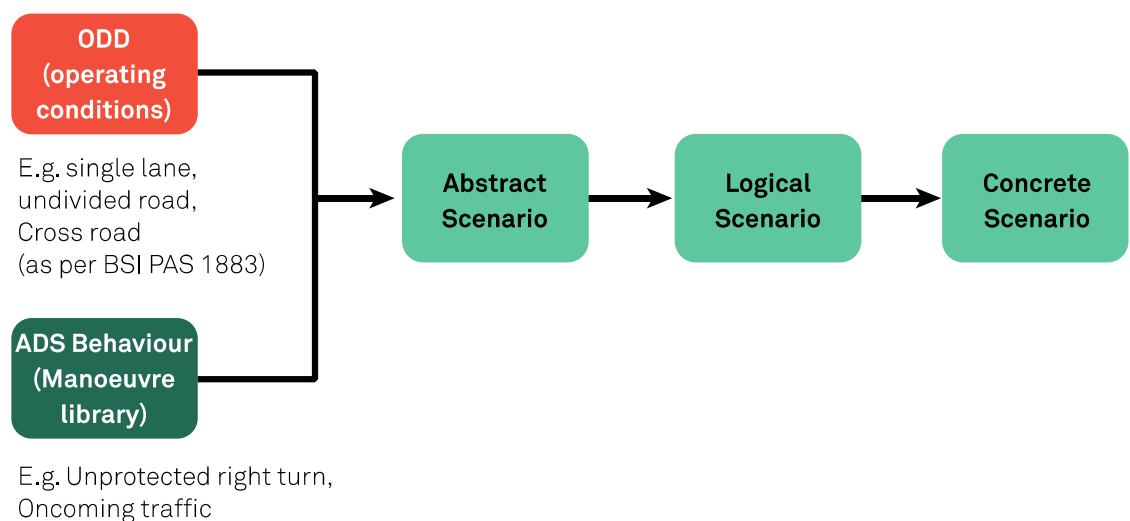
Therefore, the scenario generation process needs to take into consideration both the ODD and the behaviour competency defined for the HAV. One can identify abstract scenarios by overlaying expected behaviour from an ADS and other road users on the defined ODD of an ADS. For example, performing an unprotected right turn (behaviour) on a single lane undivided cross-road (ODD) with oncoming traffic (behaviour) would be an abstract scenario.

Furthermore, certain aspects of behaviour (both of the ADS and other actors around the ADS) may be restricted by the defined ODD. For example, the manoeuvre "lane change" may be prohibited on a single lane undivided road. It is important to highlight the relationship between ODD, behaviour and scenarios. While the three constructs are related to each other, they are distinct in nature. As mentioned earlier, the ODD defines the operating conditions for an ADS.

The definition of a HAV manoeuvre library would be useful for providing a formalised common vocabulary for the specification of ADS functionality (Section 2.4.3) and scenario generation. This may illustrate standard sets of HAV manoeuvres for a HAV. Driving rules/codes or developments thereof for HAV such as a codified Highway Code (see Section 3.4.5.1) that provide rules and guidance applying to HAV behaviour could be used to derive HAV manoeuvres, or refer to a HAV manoeuvre library.

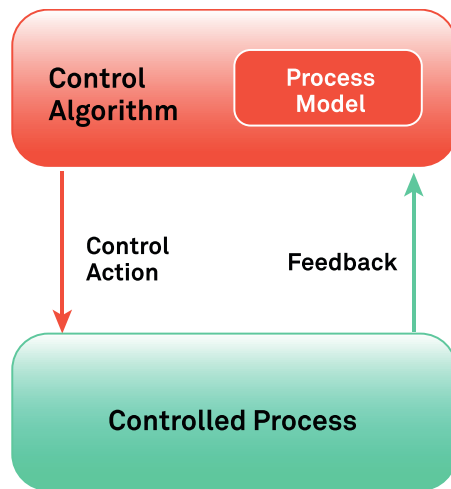
As discussed earlier, the behaviour of a HAV may be restricted due to many factors e.g. its ODD or system design. The creation of a manoeuvre library allows for HAV behaviours to be described in either a prescriptive format or a restrictive format.

Figure 24: ODD based Scenario generation process



In the development of safety argumentation, it is essential to decompose each criterion further into more tangible and measurable performance indicators. We suggest an ODD-based and behaviour-based scenario generation process. The foundations of the proposal are inspired by control theory. Consider a control loop as illustrated in **Figure 25**. In order for the controller to perform an action (i.e. give a control command), its process model will have a certain belief about the environment, and it will receive some feedback from the sensors which will provide context to the situation. Based on the process model beliefs (i.e. assumptions) and the feedback provided by the sensors (i.e. context), the controller will issue a control command.

Figure 25: Typical control loop structure



Therefore, in order for a set of goals to hold true, the controller will have a set of assumptions and context. Only under those assumptions and context would the driving criteria hold true. For example: Let's consider a driving goal (goal 1): Do not cause collisions.

For goal 1 to be true for a particular intended functionality, the following assumptions could be assumed to hold true:

- Ego vehicle shall follow the rules of the road
- All other road actors shall follow the rules of the road

It is important to highlight here that the set rules of relevant road rules themselves will be a function of the ODD. For example, if the ODD involves a motorway junction, the rules of road are different to those compared to a city centre roundabout junction. Therefore, we can further refine the above conditions as:

For goal 1 to be true, in a given ODD, with the following assumptions and contexts need to be true:

- A1: Ego vehicle shall follow the rules of the road for the given ODD
- A2: All other road actors shall follow the rules of the road for the given ODD

Given a codified Highway Code, one can now use the Highway Code rules for the given ODD as a pass criterion for the high-level goal ("do not cause collisions"). Additionally, the assumptions for the goal to be true serve as a criterion for understanding if the intended scenario actually got executed in the test setup.

3.4.9.8

Selection of Test Medium

As part of the evaluation continuum mentioned earlier, there is a growing consensus that, as an ADS progresses through its V&V cycle, the primary test medium will gradually shift from digital world to real-world testing. However, the aim is to front load most of the testing process to digital world or simulation-based testing. Within this goal, the aim is to use the evidence generated from the simulation-based testing to:

- Inform the downstream testing process (test-tracks and real-world testing)
- Form part of the safety evidence and corresponding safety case

Both these aspects have an implicit assumption that the results from simulation-based testing are valid and are a good representation of the real-world. This is a big assumption and a non-trivial task to validate. In order for simulation based testing evidence to be used as a part of the safety evidence or to inform reduced real-world testing, it is imperative that the simulation tool is validated itself. While there is a lack of consensus on the definition of a "validated simulation tool", activities have been initiated by the Validation Method for Automated Driving (VMAD) – SG2 on developing consensus on this aspect. It must be noted that in order to validate a simulation tool, the validation process can be focussed on two levels: 1) the tool itself, or 2) output of the tool. While the former focusses on the tool development process, the latter would compare the results from a simulation-based test to real-world test.

Furthermore, the fidelity and validation of the simulation tool will need to be a function of the scenarios it is being used for testing and the safety evidence it is expected to generate. Stakeholders using simulation to generate evidence for their perception system will need to focus on the fidelity of the environment model, world model, sensor models etc. in order to have validated set of results. Stakeholders using simulation-based testing for generating evidence for planning function, would need to focus on map fidelity instead of the environment model.

3.5 Summary Findings

This section discusses the current state of HAV assurance practice, identifies key gaps, and presents approaches for addressing them. These include:

- Performance Indicators (see Section 3.2)
 - There is a need to develop these as a common reference to supporting the definition of requirements, assessments, management of system safety/quality, and ongoing monitoring that the system of assurance addresses needs
 - It is through common and formalised performance indicators that traceability can be provided between regulatory requirements, and the system developers quality & safety case argumentation along with the formal V&V outputs and body of evidence that justifies the argumentation
 - We have identified four categories of performance indicator, each of which may perform different roles in an assurance framework
- Regulatory performance requirements for HAV (see Section 3.3)
 - There is a need to develop regulatory requirements that define the performance standards that HASSs should conform to in particular for safety
 - We have identified that performance requirements may be considered in two categories. There are performance specifications that are verified against, and performance needs that are validated. To date automotive technical regulations typically rely on relatively prescriptive specifications. However, we've identified that a HAV's relative complexity and responsibilities in performing the driving creates the need for an increased role of needs based requirements.
 - We have explored how the development of performance requirements could be approached and identified specification with components of conditions, limitations, and certainty. This significant work area would address the "how safe is safe enough" question and is interdependent on other work areas such as the development of a codified Highway Code and performance indicators.
- HAV assurance practice (see Section 3.4). In this section we consider how current best practice may be applied, however the more significant focus is discussing the development of practice. There are needs for best practice to be developed for HAV assurance both to support the development of regulatory performance requirements, and to facilitate the assurance of their achievement
 - There is a significant need for practice to develop, verify and validate the intended functionality of HAV driving features to ensure it is complete and sufficient for taking full responsibility for driving in a potentially wide range of vehicle and domain attributes. This area we term behavioural safety, and includes
 - The handling of road traffic rules and co-operation. Developing a codified Highway Code would develop an understanding of desired behaviour of HAVs (Section 3.4.5.1)
 - The risk control strategy for HAV is not limited to response/mitigation of hazards, as the HAV's behaviour influences the exposure to hazards in the multi-agent environment it is operating within. Risk modelling should therefore account for this area of tactical safety, in addition to the more established area of active safety (Section 3.4.5.2)
 - There is a need to develop guidance on Minimal Risk Manoeuvres and Minimal Risk Conditions.

- Probabilistic systems are typically involved in ADS perception, and may be involved in decision making. Some of this area is covered by current SOTIF best practice. However, practice for V&V of systems trained using machine learning is insufficiently mature, and additional research and development on this topic is needed
- Practice is required to ensure the suitable specification and use of sensors and data systems, to support modelling and definition of physics-based sensors
- Although outside of CertiCAV scope, the need to develop harmonised standards and principles for appropriate use of co-operative sensors is recognised.
- Validation and testing (see Section 3.4.9)
 - On-road testing is important for validation; however, risks need to be managed, in particular to the public.
 - Practice is required for the management of complexities of HASS development and management of system safety case.
 - System development processes for HAV have various interrelated inputs and outputs with each other that need to be co-ordinated.
 - Developments are required to manage the significant role of virtual testing in HAV assurance. Tooling needs to be validated to have appropriate fidelity for the attributes it is modelling and the role of the evidence it creates in safety argumentation.
 - The development of standardised a HAV manoeuvre library (3.4.9.3) would help specify ODDs and generate scenarios. Furthermore, it could formalise part of the vocabulary for a codified Highway Code (3.4.5.1).
 - Scenario based testing is important for assessment and practice is required for generation of scenarios and mapping these to coverage of safety and performance. Both the ODD and Behaviour of a HAV are important for determining the set of scenarios that are relevant for HAV assessment, testing and validation.
 - In order to use safety case evidence based on simulation-based testing, there is a need to create a standard set of metrics and process for validation simulation testbench depending on its intended use.
- While BSI (The British Standards Institution, 2018) and the UK government (HM Government, 2017) have created excellent principles for HAV cybersecurity, it may be beneficial to establish a detailed process for assuring the cybersecurity of HAVs (noting this is out of scope for CertiCAV)
- A common challenge is that the performance indicators, requirements and codes currently applied to human drivers do not quantify performance or support traceability to formal verification or validation of performance achievement. There is therefore a need to formalise these to support the assurance of HAV

4 System Development

This section considers relevant areas of assurance in the context of HASS development.

4.1 Introduction

It is assumed that an ADSE would develop a HASS, intended for on-road deployment, and would need to ensure that requirements defined by a Regulator/TAA are satisfied. They may additionally need to ensure that system specification (e.g. required by LFO) and/or performance needs (e.g. required by traffic authority and/or LFO) are satisfied.

The key objective of System Development, therefore, is to ensure that the HASS satisfies all relevant performance requirements. It is assumed that this shall be demonstrated by a system safety case. For example, key top level items to be demonstrated are: that appropriate HAV behaviour is identified and verified and validated as sufficiently safe for all conditions that reasonably occur within the deployment domain; that the ADS is aware (via on-board and/or off-board sensing) when the limits of the ODD are reached, and that it responds by executing an appropriate fallback to achieve a minimal risk condition.

4.2 System Safety Case

The system safety case should be a systematic and structured argument that clarifies all relevant performance requirements for the HASS and justifies their sufficiency.

A complete safety case involves a variety of design, specification, verification, and validation processes. In Section 3.4 of this document we have identified and discussed areas of practice for HAV assurance. Following in this section we discuss principles and objectives, relating to this practice that are considered relevant for the assurance of system development.

We suggest that, in general consensus with existing practice, a safety case should be structured into Goals, Argumentation and Evidence, for which we have identified the following objectives.

- Goals
 - Define safety goals that ensure full coverage of requirements
 - Relate to well defined and quantified performance indicators
 - Consider system safety concept and safety by design
 - Assess and manage relationship between processes that develop/validate system specifications (e.g. behavioural safety/intended functionality), with other processes those impact, and the safety argumentation (e.g. selection and structuring of performance indicators)
 - Assess and manage the potential interaction of residual risks between validation/development processes, ensure that characterisation of residual risks is robust.
 - Cover safety culture and validation strategy, such as relationship of whole system validation with individual system validation processes and test pillars, reduction of uncertainties, and evidencing
- Argumentation
 - Provides organised structured arguments according to a defined format.
 - Ensure relational inputs and outputs of safety processes are identified and implemented
 - Ensure system wide risks and performance aspects are identified and assessed
- Evidence
 - Use evidence with auditable, defined formats.

Consideration could be given to adopting and demonstrating compliance to UL 4600 as practice for formatting, consistency, and completeness of system safety case. UL4600 is discussed in [Section 8.15](#).

The need to relate to well defined and quantified performance indicators is considered to be of particular importance. [Section 3.2](#) of this report discusses performance indicators, furthermore UL4600⁴⁵ categorises types of performance indicators and recommends their role in the argumentation of safety case.

[Section 3.3](#) of this report discusses approaches for developing performance requirements for HAV which includes:

- The concept of driving performance criteria as a methodology for identifying a superset of HASSdriving performance requirements
- Application challenges and an example of performance requirement definition (see [Sections 3.3.4](#) and [3.3.5](#))

Over time, it is expected that regulatory bodies will develop and clarify regulatory performance needs for HAV on-road deployments. However, in any case a key quality assurance activity that a System Developer would need to undertake is to assess and clarify what performance needs relate to the HASS they are developing and describe top level goals, to be applied to system development, that provide full coverage of these.

The safety case should ensure that all system development processes are co-ordinated and ensure integrity of overall safety management and assessment.

⁴⁵ Undertakers Laboratories UL 4600

The following objectives are identified as relevant for safety and quality management of the HASS:

- Adopt a sufficiently well developed and systematic process for managing safety/quality and ensure relevant requirements are satisfied including regulatory obligations. This should include key safety and QA documentation e.g. safety release judgements, safety anomalies management, Hazard Analysis and Risk Assessment schedules, and risk mitigation methodologies
- Identify and apply appropriate practice and processes
- Address risks across the entire supply chain with well-defined roles & responsibilities and ensure access to, and retention of, control documents and evidence
- Co-ordinate and sufficiently explain the deployment with necessary stakeholders and develop a complete system deployment specification
- Document full specifications for nominal state of systems and integrate this into testing processes
- Define and manage safety for the whole lifecycle of the HAV. For example, setting out inspection and maintenance processes for software and hardware
- Define an appropriate organisational risk culture. Assess and define risk appetite, ensuring this satisfies regulatory requirements and relevant guidelines. Describe principles for vehicle testing culture
- Organise safety for the whole system, ensure safety arguments are systematic structured, evidenced and documented
- Apply appropriate change management

Of key importance is ensuring that the HASS specification is well structured, accurate and sufficiently detailed for the following reasons:

- This is a vital input for supporting System Development Verification and Validation activities
- This information would be required to support collaboration with key stakeholders including suppliers, customers, and approval authorities

Section **2.3.3** of this document identifies a structure and key attributes for HASS specification and it is suggested that a sufficiently developed specification would cover satisfy these. The system developer should ensure that the deployment specification is sufficiently co-ordinated with relevant stakeholders.

Another aspect to consider is that of change management. The ADSE should establish stages of whole lifecycle for all items of the HASS and implement an appropriate change management process for software and hardware. This needs to be associated with appropriate processes for product verification and validation testing.

4.2.1 Operational safety

Although operational safety is outside the scope of CertiCAV, several operational safety attributes are relevant inputs for HASS development. Objectives for operational safety that we have identified as relevant for system safety include:

- ODD appropriateness for the deployment domain
 - A complete and accurate definition of the ODD is a critical input for ensuring complete and accurate system validation e.g. in the process of defining test scenarios. ODD validation is therefore part of system validation
 - Deployment domain conditions characterisation
- Determining the frequency that ODD boundary conditions occur in operation would be required to calculate risks associated with fallback operation
- System/user handover
 - How the system reaches/leaves the ODD and the procedure for activating automated mode affects system design
- Monitoring / alerting
 - How the system performance will be monitored and the procedure of alerting and responding to issues and events. These needs to be facilitated in system design
- Organisational relationships, the responsibilities and activities of organisations should be clear
 - For example, the maintenance & inspection regime assumed to be undertaken by a LFO including component inspection / replacement may affect specification system component reliability and/or self monitoring capabilities

The HASS specification presented in this document (see Section 2.3.3) discusses key attributes relevant to operational safety. This includes definition of the ODD (Section 2.4.2) and operational management (see Section 2.4.5).

4.2.2 Functional safety

This document, due to its scope does not cover functional safety. However functional safety is recognised as an important process. Firstly, as part of safety in design and risk assessment in its own right. However, of greater relevance to CertiCAV the functional safety process provides a relevant input to other safety assessment processes so that risks relating to a combination of other safety factors, such as SOTIF or behavioural safety may be assessed and quantified.

In this document we set out some relevant considerations for the practice of functional safety in Section 3.4.4.

4.2.3 Behavioural safety

Creating, verifying, and validating an intended functionality for driving features that satisfies relevant needs is considered an important part of HASS development. Relevant objectives that we have identified include:

- Suitably complying with road traffic rules and co-operating with other road users (see Section 3.4.5.1)
- Achieving acceptable safety of HAV behaviour by implementing suitable active and tactical safety controls (see Section 3.4.5.2)
- Implementing suitable behaviour to cover fallback operation including:
 - Ensuring MRM is suitable
 - Accounting for the risk of being in MRC after execution of MRM, accounting for the likelihood of the MRM-triggering condition
 - Quantifying and managing potential for unknown unsafe scenarios arising from lack of human in loop as a fallback

A system developer would need to adopt or create suitable processes to achieve this. The challenges and complexity of achieving behavioural safety are likely to vary significantly depending on the complexity and challenges associated with the HAV deployment.

The system safety case argumentation should incorporate performance indicators that quantify the achievement of behavioural safety, including:

- Active Safety controls
- Tactical Safety controls
- Co-operation with other road users and road traffic rules
- Fallback operation, including coverage/suitability of MRM and residual risk of MRC

4.2.4 Probabilistic systems

The live dataset which is crucial to tactical decision making is created by sub-systems that are probabilistic in nature and typically involve use of AI. This gives rise to the need for processes to cover objectives including the following:

- Safe specification and use of perception systems input to driving features (SOTIF)
- Certainty modelling, reporting, management
- Learnt system validation: explaining/characterising expected performance and testing resilience
- Definition and appropriate use of Artificial Intelligence(AI) systems including online and offline learning systems

There is general consensus across industry to apply SOTIF practice for the validation of perception systems. Accordingly, it'd be expected for a system developer to demonstrate conformance by ensuring the objectives of ISO 21448 (sub-clauses 5.1, 6.1, 7.1, 8.1, 9.1, 10.1, 11.1 and 12.1) are covered by system safety case goals, providing arguments for the satisfaction of these objectives and evidence that justifies their achievement. However, in the development of a HASS gaps that are not covered by SOTIF would need to be addressed as discussed in Section 3.4.6.1.

Furthermore, the use of AI should be sufficiently defined, explained, and managed (see Section 3.4.6.2). This should include ensuring there are no intrinsic biases as a result of the training dataset.

The system safety case argumentation should incorporate performance indicators that quantify the suitable performance of probabilistic systems, this might include:

- False detection rate and severity
- Certainties of object data attributes

4.2.5 Sensors & co-operative safety

Identified relevant objectives relating to the suitable specification and use of sensors and data systems include:

- Specification of sensor systems including suitability for intended functionality for example range of view, refresh rate, precision (related to ADS functionality see Section 2.4.3)
- Characterisation of sensor performance, for example based on environmental conditions
- Safe concept of operations: In particular the involvement of any safety critical systems that are external to the HAV should be described and controlled (related to HASS architecture Section 2.4.4)
- Assessment and management of sensor faults (related to functional safety see Section 3.4.4)
- Performance monitoring/management (related to operational safety see Section 3.4.3)

The HASS specification should include vehicle operational characteristics (see Section 2.4.1) and sensor performance aspects of perception within ADS functionality (see Section 2.4.3) and relevant data communications that influence HAV behaviour (see Section 2.4.4); of particular relevance the involvement of any safety critical functions over data communications should be identified and controlled.

The specification and characterisation of sensor performance is an important input for developing sensor models for simulation, validation testing and supply chain management.

4.2.6 Data management

Identified relevant objectives to be satisfied by data management include:

- Data filtering and evidence retention to justify safety arguments
- Collection of monitoring data from physical use of HAV to support system validation, and monitoring activities
- Transmission of data to defined standards between organisations

There is growing consensus of use of the OpenX data standards (see Section 3.4.8)

The system safety case should reference all relevant evidence to justify the satisfaction of safety arguments in described and traceable formats.

Relevant data from physical testing of vehicles should be identified, monitored, and analysed to reduce unknown scenarios and further performance validation against known scenarios.

PAS 1882, currently in development, provides guidance on the collection of monitoring data as part of on-road use of vehicles. When published this may provide directions on monitoring data requirements and/or provide a reference for standards that may be adopted for a wider use case of data collection and formatting.

4.2.7 Validation & testing

This area includes the co-ordination of various system development processes to generate and test a set of scenarios that provide sufficient validation coverage. It also includes the tooling/methodologies to undertake testing, and should satisfy the following objectives:

- Support safety and quality management, in particular managing risks associated with physical use of system and use in public on-road environments
- Monitor all testing to identify and reduce previously unknown scenarios, and detect issues/events
- Evidence from V&V activities must be recorded and argue an integral progression of maturity
- Verifying that system performance satisfies system specifications
- Co-ordination of inputs and outputs between various system development processes
- Identification and mitigation of any potential failure cases that arise from subsystem interactions, for the cases where subsystems are tested independently

We have identified the following principles that are considered an important part of validation practice due to its integral role in testing culture:

- Manage the progression of validation of all system items by tracking validation levels, and apply appropriate safety controls. This could include Safety process release judgements, documentation and management of safety anomalies, Hazard Analysis and Risk Assessment schedules, risk mitigation methodologies
- Ensure sequential progression of verification and validation methods from test bench and off-road testing before progressing to public road environment or reduced monitoring
- Validate testing processes and testbench and apply accreditation where applicable
- Apply performance indicators to monitor, track and record organisational and system safety maturity during V&V activities, apply these in coherence with the organisations risk appetite, testing culture and regulatory requirements to argue the release of systems to progressive testing environments

Validation and testing activities should clearly demonstrate sufficient coverage of the range of scenarios that may occur within the ODD. This coverage is likely to require:

- Use of multiple comprehensive scenario identification processes, using data and/or structured methodologies to ensure there are no gaps in ODD coverage (see also Section 3.4.9.3)
- Automated generation of randomised variants of the scenarios identified
- Methods to identify and focus testing on the critical scenarios for the particular ADS, accounting for the architecture and technologies it uses, and any failures or near failures found from previous testing
- For more complex ODDs, automated methods to assess performance in scenario tests. These are needed to enable a sufficient number of scenarios to be tested, and to ensure complete and correct application of the performance criteria to ADS behaviour

Section 3.4.9 discusses various approaches for testing and validation.

4.3 Summary

This section discussed objectives and principles to be followed by an ADSE to guide HASS development and sufficiently cover assurance needs.

There are advantages in gaining assurance through demonstration of conformance to good practice. However, there is, for at least for the foreseeable future, substantive lack of established best practice and harmonisation for HAV systems that would be required to facilitate this.

Therefore, for the time being the development of a safety case is considered to be important for the management safety and quality during HASS system development. This section focuses on the identification of a set of principles, objectives and requirements applied to the practice of system safety case. This includes various development processes to be covered in the system safety case, these in turn reference the discussion of HAV assurance practice in [Section 3](#).

In summary assurance should be achieved by the development of a system safety case that:

- Is formalised, structured and consistent
- Relates to performance indicators to provide traceability to requirements and to evidence
- Covers and co-ordinates system development processes
- Demonstrates achievement of safety requirements and is consistent with defined organisational risk appetite

5 System Approval

This section builds on the content of previous sections to highlight key considerations for the approval of a HASS for on-road deployment.

5.1 Introduction

The context of this section assumes the regulatory process defined in Section 2.13.2.

It is assumed that a Regulator/TAA, supported by an independent technical service shall undertake approval inspection and testing of a HASS intended for on-road deployment as part of a system approval. This section discusses this first phase of system approval.

It is assumed that upon completion of this, there would be a second phase of operations approval, subsequently a third ongoing phase of operations monitoring may continue to be an important part of assurance

It is not practical for independent 3rd parties to have full access to all data and repeat all evaluation activities to the same level of detail as undertaken during development of the system. Practically, independent testing needs to be performed within resource and time limitations.

Current automotive good practice often involves the conformance of systems and/or processes to established best practice standards such as ISO 26262 for functional safety. This facilitates quality/safety management to be better integrated into the system development process, which improves efficiency. Furthermore, demonstrating conformance to best practice provides strong evidence for assurance purposes. This, however, requires best practice to be defined, whereas currently there are substantial gaps in standards/practice coverage for HASSs which presents a barrier to integrating conformance processes in system development and approval activities.

In a similar fashion, there would be benefits of engaging independent approval bodies early. This could involve integrating the independent technical service activities in parallel with development processes rather than implementing development and approval activities sequentially. This however requires the support of standardised and well-developed approval practices, and careful management of change control challenges as discussed in Section 5.13.

There is a wide diversity in the potential specification of a HAV, where and how it is deployed, and the system supporting it. This in turn creates a wide variability in the nature, complexity, and relevant risks of a HASS. The process of testing and inspection therefore needs to be flexible, suitable, and proportional for the subject HASS, however some consistency is also required.

5.1.1 Context and Objective

The purpose is to define a framework process for the system approval of a HASS intended for on-road deployment. The context of the on-road deployment and system approval is defined in Section 2.1, the definition of a HASS is defined in Section 2.3.3.

It is assumed that an ADSE would present a HASS for approval with a HASS specification (see Section 2.3.3) and a system safety case. It is assumed that regulatory requirements would be defined (see Section 3.3)

The objectives of the system approval would be to determine via audit and inspection that:

- The HASS specification is adequately described, suitable and co-ordinated
- The ADSE has applied sufficient Quality and Safety Management
- The system safety case is sufficiently developed, provides full coverage of the HASS specification, has sufficient argumentation, evidence for coverage, and demonstrates achievement of regulatory requirements

Furthermore, to verify by evaluation of evidence and/or witnessing of tests that requirements are met through:

- Evaluating sufficient quantity and quality of on-road trials monitoring to argue system maturity and safety validation
- On-road testing: Deployment management and HAV behaviour competence
- Testbench testing: Toolchain capabilities, sub-system performance
- Physical Testing: perception performance and HAV behaviour in selected scenarios
- Physical and testbench comparison: fidelity of the evaluation toolchain

5.1.2 Deployment risk classification

To implement proportionality of independent assurance a Deployment Risk Classification (DRC) assessment could be applied to scale the coverage of independent assurance activities.

For example, this could be calculated based on the matrix shown in below which determines a deployment risk class between A-G based on:

- The maximal severity consequence of a failure. As an example of how this could be derived, **Figure 27** is a table for determining potential severity consequence from the collision delta velocities between two passenger vehicles. This could be expanded upon to include additional factors such as: HAV occupant configuration (are they seated/restrained or standing), payload, collision with vulnerable road users.
- Scale of the deployment, which affects risk exposure. For instance, number of kilometres driven per annum: D1<1000km; to D4>1million km
- Deployment complexity: The complexities involved in the deployment specification such as the ODD and ADS functionality. C1=low speed, highly controlled environments, restricted

manoeuvres; C4 = Dynamic deployment in a complex and/or informal environment e.g. in close interaction with vulnerable road users

It is suggested that the DRC would scale the coverage of independent assessment, e.g. number of concrete scenarios; the performance requirements and acceptance criteria should remain the same.

Figure 26 DRC Matrix

Severity Class	Deployment Scale	Deployment Complexity			
		C1	C2	C3	C4
S1	D1	A	A	B	B
S1	D2	A	A	B	B
S1	D3	A	B	B	C
S1	D4	B	B	C	C
S2	D1	B	C	C	D
S2	D2	B	C	D	D
S2	D3	C	C	D	E
S2	D4	C	D	E	E
S3	D1	D	D	E	E
S3	D2	D	E	F	F
S3	D3	E	F	F	G
S3	D4	E	F	G	G

Figure 27: Velocity delta based hazard severity determination (SAE J2980)

Collision Type	Range for	S0	S1	S2	S3
Front	Min. km/h		>4-10	>20-50	>40-65
	Max. km/h	<4-10	<20-50	<=40-65	
Back	Min. km/h		<4-10	>20-50	>40-60
	Max. km/h	<4-10	<20-50	<=40-60	
Side	Min. km/h		>2-10	>=8-30	>16-40
	Max. km/h	<2-3	<8-30	<16-40	

5.1.3

Change control

This document does not address change control but acknowledges that this will be an important factor for HAV.

Currently UN Regulation 79 has provisions that⁴⁶ (para 9.1) 'Every modification of the vehicle type shall be notified to the Type Approval Authority which granted the approval. The Type Approval Authority may then either':

- (para 9.1.1) 'Consider that the modifications made are unlikely to have an appreciable adverse effect and that in any case the vehicle still complies with the requirements';

⁴⁶ (United Nations, 2017)

- (para 9.1.2) 'Require a further test report from the Technical Service responsible for conducting the tests'

It is envisaged that a HASS would have relatively frequent incremental software updates throughout its lifecycle, and this would make it impractical to notify the TAA for a judgement on every modification. This creates the need for practice to be developed for change control and relate this to independent assurance activities. This is particularly relevant if some of the approval activities are integrated into the development process, rather than sequentially undertaking approvals after completion of development.

For highly mature systems, it is envisaged that issues would be discovered over time that should be addressed promptly. This would need to involve processes of managing change, ensuring that the potential impacts of implementing change are known and appropriate verification and validation activities undertaken prospectively or retrospectively. Independent assurance activities could involve, for example, a process of selecting independent tests that have some focus on scrutinising the change, but also some randomisation to spot check change before it is implemented, or a system of retrospective testing of systems at regular intervals.

Furthermore, change needs to be managed during the product development and initial approval cycle. It is not practical to recommence verification, validation, and approval from the beginning every time a change is initiated, however checks and balances do need to be in place to identify which items are affected by change and ensure that changes are validated.

5.2 Approval process

This section sets out a devised approval process with sequential assessment steps. **Table 9** describes these steps and the objectives they satisfy.

For this work we presume that the steps would progress sequentially with the outcome of each step either being to pass and proceed to the next step, minor non-conformance/corrective actions and proceed to next step, or major non-conformance and to require corrective action and repeat assessment prior to proceeding to next assessment step. Further developments could finesse this process, for example including some parallel working of activities for improved efficiency.

It is intended that successful completion of this approval process would approve the HASS for the use described in its specification and the safety concept applied by the system developer. However, it should be noted that

- This approval relates to the first phase of system approval of the assumed regulatory process (see Section 2.13.2), and there would be a second phase of operations approval
- The assumed regulatory process includes a third phase of ongoing monitoring after approval
- The nature of the approval release depends on how phase 2 and 3 approvals are developed (outside of CertiCAV scope). For example, there could be an initial probationary period of enhanced monitoring and ad-hoc inspections to build further trust and confidence in the system and operations before fully releasing the HASS operation

Table 9: Assessment steps and objectives

Step	Activity	Objectives
1	HASS Specification & Safety Check	Assess that the HASS specification is well described/ developed, suitable and co-ordinated.
2	Road Test ⁴⁷	Inspect that appropriate deployment management processes are in place. Assess that Driving performance satisfies appropriate competency level.
3	Audit and Inspection	Assess that the organisation has applied sufficient and appropriate process, policies, tooling, and resources.
4	System Safety Case / Evidence Evaluation	Assess that the system safety case meets requirements and has sufficient arguments and evidence.
5	Scenario Generation	Create effective Scenarios for testing.
6.1	Testbench Testing	Verify testbench capabilities. Assess and verify HAV behaviour in scenario tests.
6.2	Physical Testing	Validate perception performance. Assess and verify HAV behaviour in real world controlled scenarios.
6.3	Fidelity Check	Validate toolchain fidelity.
7	Road Test	Refer to step 2 and discussion in 5.2.2

Following in this section we discuss each assessment step in further detail.

⁴⁷ There are advantages/disadvantages to conducting towards the beginning or the end of the approval process, or there could be several road tests. Please refer to discussion in Section 5.2.2

5.2.1 HASS Specification & Safety Check

The objective of this step is to assess that the:

- HASS specification is well described/developed, suitable, and co-ordinated
- Adopted safety principles and system safety case development are adequate
- Clarify that the information provided sufficiently defines the HASS and organisational structure is sufficient for planning and undertaking of the approval

The intention of this step is to facilitate an initial check that the ADSE and HASS are suitable for committing to the approval process and to plan and agree the approval process to be suitable for the details and specification of the HASS and associated deployment based on information provided.

To undertake this, it is assumed that the ADSE would provide a HASS Specification and supporting documentation that would be assessed by the Technical Service.

The following could be considered for the assessment of the HASS specification:

- That it is well described and complete
 - Section **2.3.3** of this report discusses HASS specification, it is suggested this may be used as a reference to define structure and content requirements
 - There should be evidence that attributes are specified according to defined, structured, traceable, and appropriate formats. This could include conformance to defined standards and/or best practice, e.g. for ODD BSI PAS 1883, or in the future OpenODD, or ISO 34503
 - Attribute specifications should describe bounds and limitations for operational/nominal conditions and configurations. The wider approval could presume that the HASS is expected to handle under nominal operation any attribute conditions that are not explicitly defined as boundaries/limitations in the specification
 - The whole HASS architecture should be described including off vehicle systems, with a particular focus to those influencing the performance of safety critical functions
- That it is suitable and co-ordinated
 - The various parts of the specification should be suitable and coherent with each other. For example, attributes for HAV operational speed and manoeuvres (ADS functionality) should be appropriate for the road conditions it is intended to be deployed in (ODD).
 - Common guidelines, such as codified Highway Code (see Section **3.4.5.1**) could be developed to support this
 - The specification should cover the full HASS with particular relevance to systems involved in safety critical operation of the HAV, including items that may not be fully under the control of the ADSE organisation. Where appropriate there should be evidence of co-ordination and agreements with relevant stakeholders to demonstrate that access and information that would reasonably be required as part of the approval process will be available. For example, telecommunications service providers, roadside traffic infrastructure providers
 - It may be appropriate to ensure that consultation is undertaken with such as sponsor/client/customer, local/regional traffic authorities, or emergency services. The purpose of this would be to ensure the nature of the HASS and future deployment is sufficiently explained to relevant supporting stakeholders and agreed to be feasible

- That specifications are reasonable, for example, the specifications of tolerance of nominal conditions may include limitations on the behaviour of other actors, such as that they conform to certain traffic rules (see Section 2.12.2). However, these may be considered to be unreasonable if these are restrictive to the extent that the HAV behaviour would be very uncooperative to other road users, or demonstrate incompetent execution of the driving task.

Documentation would be required to undertake the assessment that the adopted safety principles and system safety case development are adequate. It may be considered inappropriate for the ADSE to share a full system safety case at this stage; however, an outline of the system safety case and safety concept including safety/risk strategy could be provided. The assessment of this could consider the following:

- That safety principles are adequate:
 - Provided statements of organisational safety culture including risk appetite and safety/quality assurance are well developed, define suitable thresholds, and are owned at a sufficient level within the organisation
 - There is sufficient quantity and quality of on-road trials monitoring to argue safety validation
 - Performance monitoring data from trials is coherent with the validation management of HASS items: for example that the frequency and nature of incidents suits current validation levels/release statements
- That safety case development is sufficient:
 - The safety case is structured, provides full coverage of the deployment specification, has sufficient argumentation, and satisfies safety requirements
 - There is sufficient development of evidence, in particular quantity and quality of on-road trials monitoring to argue validation of system safety

We have identified the need to clarify that the information provided sufficiently defines the HASS and that the organisational structure is sufficient for planning and undertaking of the remainder of the approval process. It may be appropriate to verify co-ordination and engagement with relevant stakeholders. e.g. police and emergency services. This may include checking there are agreements/ co-ordination in place with 3rd parties who have responsibilities in supporting safety critical system architecture to support the approval process: for example, roadside V2X infrastructure provided by local transport authorities, telecommunication services, and HD map data services.

Due to the diverse range of configuration and complexity of a HASS and associated deployment it is suggested that as part of this approval step planning of the remaining approval steps would be undertaken based on the review of the HASS specification and supporting documents. This may include, for example, deciding which assessment activities are applicable and to what detail/ scale which would determine the programme and financials for the approval process. The deployment risk classification (see Section 5.1.2) could be used as part of this process.

5.2.2 Road test

Objectives of the road test are:

- Verifying performance and assessing behaviour in real world deployment
- Assessing deployment management, does it match deployment specification

It is assumed that for the road test the HASS would be operated in the same way as on-road trials/advanced trials for which the HASS was previously deployed. This would include the same safety mitigations, which generally include an on-board safety driver actively supervising the driving task and ready to intervene.

The test could be undertaken in a similar manner to a human driving licence test with an examiner in the passenger seat observing and assessing the performance of the driving task. They could record major and minor faults observed on a check-list associated with demerit points to determine a pass/fail decision. However, it is suggested that the type of faults and how they relate to demerit points differs from a human driving test.

- Minor vehicle control faults, such as touching the kerb which may be acceptable for a provisional human driver, are unlikely to be acceptable for a HAV
- All traffic rules should be conformed to, unless conditions are satisfied to justify rule violation that were communicated before the test
- The assessment calibration may aim to determine whether the HAV's execution of the driving task demonstrates sufficient competency for its deployment domain
- To establish suitable driving criteria, it may be beneficial to reference guidelines of driving behaviour, a CHC would be a highly relevant reference (see Section 3.4.5.1). Other references could include: the highway code, DVSA instructor guidelines or advanced driving guidelines
- The engagement of system fallback should be acceptable if this is demonstrated to be due to system specification limitations, such as ODD, being reached. However:
 - The frequent or inappropriate occurrence of system specification limits being reached could lead to an assessment conclusion that the HASS specification is not appropriate for the operational conditions
 - If system fallback is engaged where it cannot be demonstrated that specification limits were reached, then this could lead to an assessment conclusion that the HASS specification is incorrect or performance of the ADS monitoring system is inadequate
- The involvement of any human fallback, such as an intervention from a safety operator demonstrates that the HASS is unsuitable for unmitigated deployment and therefore should be automatically failed
- Additionally, during the road test it may be witnessed whether the deployment management implemented is sufficient and conforming with the HASS specification

The assessment could be supplemented by data recordings from onboard telematics unit, video cameras, and/or monitoring data. This would provide more objective/quantified records and post drive assessments. For example: telematics measurements would provide quantified measurements of speeds and jerks, forward facing camera would provide a record of surrounding conditions, a camera facing at the human driver controls would provide a record of if and when a safety operator intervenes in the execution of the driving task.

To prepare for the test it is suggested that:

- The technical service determines suitable test specification based on the HASS specification. The definition of the tests should suit the attributes of the HASS specification such as ODD and ADS behaviour. It may be appropriate to undertake several tests to cover a range of ODD conditions, such as time of day, or weather conditions; however, it is expected that the road test covers typical conditions that occur naturally within the operating environment during HAV deployment, rather than attempting to identify and re-create corner conditions as part of the road test
- The technical service should share test protocols; however, if it is planned to follow a fixed route but the HASS is intended to deploy with dynamic routing, it may be appropriate to withhold sharing of the route until shortly before the test to reflect the dynamic routing specification and guard against the HASS being trained specifically for the test route
- The ADSE should provide information on relevant boundaries, limitations and conditions relating to HAV operations so these may be accounted for in the testing protocol. Including:
 - Relevant operations management and procedures. There should be an understanding on protocol for fallback operation, guidelines followed by safety operator and when they would be expected to intervene
 - ODD (see Section 2.4.2)
 - There should be an understanding on protocol for fallback operation, guidelines followed by safety operator and when they would be expected to intervene
 - Relevant limitation and conditions relating to ADS behaviour. Of particular relevance the conditions under which any contravention of traffic rules is considered justified should be explained
- There should be a briefing and communication between all parties involved in the road test
- Any person external to the ADSE who occupies the HAV, e.g. the examiner should be inducted on any safety critical procedures or equipment such as emergency stop button

The following could be considered regarding placing the road test step early in the approval process.

- An advantage is the activity requires relatively little time/effort, which minimises the impact of repeating assessments due to discovery of non-conformance/corrective issues. A road test provides exposure to sufficient scenarios to gain a good understanding of the maturity of the HAV system and appropriateness/competency of HAV behaviour
- A drawback is operating a HAV on-road presents some risk. There would be better assurance of safety of the road test if it is undertaken after other assessment steps including audit and assessment and physical testing have taken place. However, for the context of system approval it is assumed that the HAV would have already been deployed on-road as part of trials/advanced trials and therefore have established safety maturity of on-road deployments

5.2.3 Audit and inspection

The objectives are to:

- Assess the organisation has applied sufficient and appropriate process and policies
- Has used sufficient and appropriate tooling, and resources

This may include assessments that:

- The HASS specification is adequately described and suitable
- Documents and processes are well developed, systematic, meet requirement and are traceable
- The safety case is sufficiently developed, provides full coverage of the deployment specification, has sufficient argumentation for coverage of inputs and reduction of unknowns, and satisfies safety requirements
- Resources and Tooling have appropriate quality controls
- Evidence is sufficient for supporting safety arguments
- Suitable practice and standards are applied
- There is a suitable organisational safety culture and quality assurance process
- Facilities, resources, and tooling provide the coverage and capabilities described in specifications/documents or are sufficient for safety argumentation
- There is sufficient quantity and quality of on-road trials monitoring to argue safety validation and results support

Sections 4.2.1 through to 4.2.7 of this report discuss practice for assurance processes which could be referred to for establishing objectives to be demonstrated and conformance criteria. The former and latter areas of safety & quality management, validation and testing are particularly relevant for the management and assessment of safe HASS performance. For example, demonstrating conformance to the following validation and testing items would be evidence of having appropriate safety controls and validation practice:

- Apply validation classification levels to all system items and apply according safety controls
- Ensure sequential progression of verification and validation methods from test bench and off-road testing before progressing to public road environment or reduced monitoring

The assessments could involve the following activities:

- Reviewing documents and evidence: For example, assessing their quality, completeness, consistency, conformance, and accuracy.
- Reviewing datasets and procedures used to gather and verify them: For example, ensuring that training datasets for ML systems do not introduce any biases, and any data characterising road user behaviour is of sufficient quality and quantity.
- Inspection/witnessing: For example, that resources & facilities are available, witnessing a demonstration of tooling capabilities
- Evaluations: For example, checking specifications, undertaking risk assessment

5.2.3.1 Definition and appropriateness of ODD for Deployment Domain

In the earlier HASS Specification and Safety Check (see 5.2.1) it was identified to check that the ODD is specified according to defined, structured, traceable, and appropriate formats which could include conformance to defined standards. A current applicable standard is BSI PAS 1883⁴⁸. Currently, ISO are developing an ODD taxonomy (ISO/AWI 34503⁴⁹) and ASAM are developing an OpenODD standard.

An ODD may be defined from different perspectives, i.e., manufacturers, tier-1 suppliers, test houses, regulators, local authorities etc. From an assurance process perspective, aspects to be demonstrated include:

1. Agreement and awareness between relevant stakeholders on the ODD definition.
2. ADS awareness of the defined ODD attributes in real-time
3. Completeness, and accuracy of the ODD definition. This could include, for example the definition and application of a process to define and test the ODD.

Even if a stakeholder uses a taxonomy standard such as BSI PAS 1883, it is important to highlight that an ODD definition shall be extensible in order to include special requirements of the deployment ODD which may be specific to the deployment area, e.g. a bridge with glass side-walls. BSI PAS 1883 defines units for rainfall attribute as mm/hr. However, a manufacturer may want to define additional sub-attributes for rainfall such as: light rain, medium rain, and heavy rain. Similarly, a local authority may want to add various agent types in the dynamic element's attributes e.g. pedestrians, e-scooters etc.

⁴⁸ <https://www.bsigroup.com/en-GB/CAV/pas-1883/>

⁴⁹ <https://www.iso.org/standard/78952.html>

5.2.4 System Safety case / evidence evaluation

The objective is to assess that the safety case is met and has sufficient arguments and evidence

It is assumed that regulatory requirements would be defined (see Section 3.3) and that an ADSE should develop a system safety case that argues these requirements are satisfied for the HASS, and has generated evidence to justify achievement of the arguments.

Section 4.2 of this report discuss practice for system safety case which could be referred to for establishing objectives to be demonstrated and conformance criteria. This may include that the safety case is structured into Goals, Argumentation and Evidence and demonstrates achievement of the following objectives:

- Goals
 - Define safety goals that ensure full coverage of requirements
 - Relate to well defined and quantified performance indicators
 - Consider system safety concept and safety by design
 - Assess and manage relationship between processes that develop/validate system specifications (e.g. behavioural safety/intended functionality), with other processes those impact, and the safety argumentation (e.g. selection and structuring of performance indicators)
 - Assess and manage the potential interaction of residual risks between validation/development processes, ensure that characterisation of residual risks is robust.
 - Cover safety culture and validation strategy, such as relationship of whole system validation with individual system validation processes and test pillars, reduction of uncertainties, and evidencing
- Argumentation
 - Provides organised structured arguments according to a defined format
 - Ensure relational inputs and outputs of safety processes are identified and implemented
 - Ensure system wide risks and performance aspects are identified and assessed
- Evidence
 - Use evidence with auditable, defined formats

Assessment of the safety case argumentation and supporting documentation could include verifying that:

- The safety case is sufficiently developed, provides full coverage of the deployment specification, has sufficient argumentation for coverage of inputs and reduction of unknowns, and satisfies safety requirements
- Resources and tooling have appropriate quality controls
- Evidence is sufficient for supporting safety arguments
- Suitable practice and standards are applied

The following could be considered for the assessment of evidence:

- That it is sufficient for justifying safety validation targets
- That it has been used to sufficiently reduce unknown scenarios
- That key performance determinations are congruent with monitoring of on-road trials
- It includes sufficient quantity and quality of on-road trials monitoring to argue safety validation

A key factor for the system safety case to argue is the achievement of confidence in system and safety maturity. The demonstration of this could include an explanation of how this is measured, how confidence has developed over time, and what values were achieved to support release judgements. A common protocol could be developed and applied as part of the demonstration and assessment of system safety maturity, for example based on learning curve for safety validation and receiver operator characteristic for tactical performance.

Evidence of the HASS operating on-road for a significant period of time, as part of trials/advanced trials, without failures, such as the need to activate mitigations is important validator of safety maturity and validation. When applying such metrics to trials this raises the issue of whether this incentivises trialling organisations to take risks in not activating mitigations and/or restrict driving to less-challenging situations. However, for the context of approval, the ADSE achieving confidence the HASS is able to deploy safely without mitigations would be an obvious pre-requisite to a genuine application for an approval for unmitigated deployment.

A case should be presented, including on-road trial deployment that demonstrates a sufficient interval of critical event free (i.e. no human intervention) operation to provide safety confidence and show congruence of intervals/learnings from on-road testing with validations from other methods. It should be demonstrated that this validation evidence has good coverage of the HAV deployment specification, for example with respect to the ODD and ADS functionality.

Runtime access to ADS predictions or run-time manoeuvre would be beneficial, however this will not be easy to harmonize as there will be significant structural differences between different ADSEs. Instead, a regulator could focus on metrics that will be common across systems, but closely linked to e.g. prediction performance.

5.2.4.1

ODD Awareness⁵⁰

For safe operation of the ADS within its ODD, the ADS must be aware of any attribute that has been used as part of the ODD definition. Measurement of the ODD attributes is key to safety of the ADS (and it being aware that it is within its deployment ODD). This may be achieved via on board sensing (on the ADS equipped vehicle) or off-board sensing (e.g. road side weather station for information on weather attributes or a traffic management system for information on dynamic elements attributes) or via the safety operator (by visual monitoring). E.g. if an ODD definition includes rainfall in mm/hr, local weather station may share such information via V2X communication to the ADS to make the ADS aware of the current rainfall rate.

Furthermore, during trials, the monitoring of the ODD attributes may be performed by the safety operator or dispatcher. As the safety operator(s) has/have the ultimate responsibility for the safety of the trial, they may even decide to allow momentary or sustained deviation from the defined trial ODD by either taking control back or by allowing the ADS to continue to perform the Dynamic Driving Task (DDT) if in their judgement, the deviation from the ODD will not lead to a hazardous situation. The safety case creators need to ensure guidance for any such judgement is included in the safety operator training procedures.

For example, if the ODD for an ADS is defined as an area where no pedestrians are allowed and there is a sudden occurrence of a pedestrian at a distance (from the ADS equipped vehicle) during the trial, the

⁵⁰ (Zenzic, n.d.)

safety operator may either choose to take control or let the ADS perform the DDT based on the safety operator's judgement of risk and training (may be because the pedestrian is sufficiently far away and not in ADS trajectory). In another example of the trial, the safety operator may be the source of ODD attribute awareness. In the case of a SAE Level 4 motorway chauffeur system being trialled on GB roads. The ODD definition of the system under test (as per BSI PAS 1883) might include motorways with heavy and light traffic but only when there is no rainfall. While the ADS may be aware about the motorway characteristics (due to an inbuilt HD-map), it may not have the capability to detect rainfall. As part of the safety case for the trial, safety case creators may allocate the responsibility for monitoring the presence or absence or intensity of rainfall on the safety operator, to make the judgement if the ADS is within or outside its defined ODD.

However, as we move towards unmitigated deployment, there is a need to develop means for the vehicle to make the decision about both ODD awareness and handing control back (or performing the Minimal Risk Manoeuvre), without human intervention.

Even in commercial deployment of HAV, some of the information about ODD attributes may be received by off-board sensors/sources. For example, if traffic density or traffic flow can cause a limitation for a HAV's performance, it will be defined as part of the HAV's ODD. However, the information about the traffic flow or traffic density may be shared with the HAV from a traffic management system via over the air communication. It is important to also note that such an information architecture would put additional requirements on the connectivity between the HAV and the traffic management centre on type of connectivity (e.g. 5G, 4G, LTE etc.), latency etc. There will subsequently be additional attributes of the ODD.

5.2.5 Scenario generation

The objective of this step is for the Technical Service to generate scenarios that provide a good challenge and coverage for subject HASS for use in the subsequent Testbench, Physical and fidelity check steps.

Section 3.4.9 of this report discusses the generation, formatting and storage of scenarios which could be applied as a reference. This includes a key principle that the generation of relevant scenarios is a function of both ODD and HAV behaviour (see **Figure 24**) that is defined as part of ADS functionality.

A challenge to the generation of scenarios for the purpose of system approval is that the selection needs to be limited to a relatively small number. This is in contrast to the generation and testing of scenarios for system development which typically involves a significant financial and resource investment committed over a period of several years. The selection of scenarios must therefore be discerning if the scenario tests are to be effective at supporting assurance.

Identified needs for scenario generation include:

- Being relevant to ODD, and HAV behaviour
- Cover a variety of control, environmental, and system attributes/challenges
- Weighted towards edge & corner cases
- Weighted towards risk exposure

A further challenge is considering how defined scenarios relate to assessment process and outcomes. It is easy to create scenarios where it is impossible for the HAV to avoid a harmful outcome, which raises the question of how to calibrate scenario parameters, so they are appropriate for a defined expected outcome and vice-versa. This is a complex area interrelated with the tools and process applied to test assessment. For example, if an adversarial assessment type is applied this would typically involve an

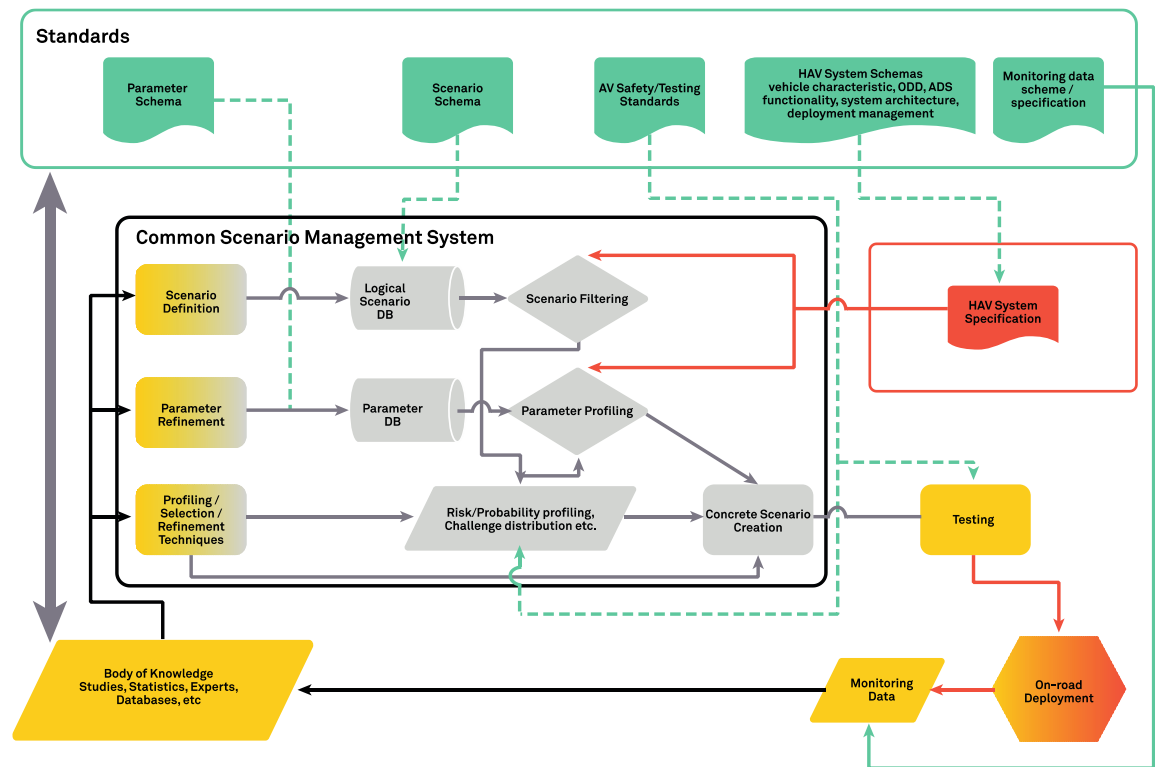
adversarial system automatically generating concrete scenarios. It will require several cycles of applying and refining scenario generation and assessment methodologies to establish good practice solutions. However, a general principle that may be applied is relating the scenario parameter values to probability distributions and severity outcomes obtained from collection of real-world data.

It would be beneficial to maintain databases for scenarios, which could be continually developed and improved over time. It is suggested that database lists of logical scenarios may be used for the purpose of scenario generation, however there are benefits in dynamically generating concrete scenarios which some degree of variation to guard against the HASS being trained to perform well for test conditions tests rather than for real world operational conditions. In-use monitoring pre or post approval assessment, could be a further source of data for assessment.

An example process for the generation of scenarios that aims to address these challenges is shown in **Figure 28**. In this process the generation of scenarios is supported by a common scenario management system which includes databases for logical scenarios and parameter profiles:

- Scenarios are generated to be relevant and effective for the HASS specification
 - Logical scenario library is filtered to be relevant for the operational characteristics, ODD and ADS functionality
 - Parameters are profiled to suit the operational characteristics, ADS functionality and system architecture
 - Concrete scenario creation is influenced by risk probability evaluations based on deployment management, and assessment knowledge/ techniques and understanding of the deployment domain
- The process is continually improved by:
 - Continual development of body of knowledge from monitoring data and knowledge gained from previous assessments and HAV operations
 - Continual development of Scenario/Parameter databases and techniques from body of knowledge
 - Linking to Standards

Figure 28: Example scenario generation process



It may be appropriate to apply different objectives to scenario generation for different test mediums:

- Testbench testing may have limitations on fidelity of sensor/perception modelling, however, is an ideal medium for high severity edge case scenarios and higher volume testing
- Physical testing does not have modelling limitations, however, presents challenges for scenarios with high safety risk and replicating exact control/environment conditions
- In addition to assessing HAV performance, an objective of scenario testing is to assess testbench.
 - Some scenarios may be generated with the purpose of assessing particular tooling capabilities (see Section 5.2.6)
 - To verify the fidelity of the evaluation toolchain by comparison some scenarios should be common across testbench and physical test mediums (see Section 5.2.8)

5.2.5.1 Testing an ODD

As discussed earlier, the ODD definition influences the entire V&V process and the safety case evidence generation. Therefore, it is essential to check if the defined ODD is actually the true ODD of the HAV.

There are two types of tests that need to be performed to establish the accuracy of the ODD definition: 1) Exiting an ODD test, 2) Entering an ODD test.

Exiting an ODD

Both for advanced trials and for commercial deployment, as part of the deployment check, it is important to perform tests at the boundaries of the defined ODD. It is important to establish that a take-over or a minimal risk manoeuvre is triggered as the HAV approaches its ODD boundary. A minimal-risk manoeuvre may differ based on the ODD and operational context of the HAV.

Entering an ODD

One of the situations that manufacturers need to prevent is the misuse of an automated system in an ODD for which it is not designed for. Therefore, there is a need to test for incorrect engagement of automation. In this type of tests, the HAV needs to be tested for attributes which differ from the values in the defined ODD of the HAV. A test should be considered as a "pass" as long as automation was not engaged while the attribute values were different from those in the ODD definition.

For both types of the tests, one may decide to use a combination of simulation and road test to provide the evidence for testing an ODD. For example, while snow in the UK in the winters is not common, it still is a realistic possibility. To perform an ODD test in real-world to show that the HAV identifies snowfall as an out of ODD condition, may be unrealistic. However, the same can be done via simulation-based evidence. This further highlights the requirement for validating the simulation toolchain as the evidence based on simulation will be of safety critical in nature.

5.2.6 Testbench testing

The objectives of testbench testing are to:

- Verify testbench capabilities
- Assess and verify HAV behaviour performance in scenario tests

The proposed methodology is that the Technical Service would have generated a set of scenarios (see Section 5.2.5) which the ADSE would input to their testbench for the Technical Service to witness and assess results. Furthermore, the Technical Service could request for selected scenarios from the ADSE's own database, or toolchain features to be demonstrated, these would be identified from understanding gained from the previous Audit and Inspection step (see Section 5.2.3). Performance assessment of the test involved in this step is discussed in Section 5.3.

The following could be considered for the verification of testbench capabilities:

- Hardware and software modelling and simulation testing tooling
- Test automation and management facilities. For instance, demonstration of automated performance assessment capabilities

Section 3.4.9.8 may be referenced for appropriateness of a test tools capabilities and fidelity based on the parameters of the scenario it is testing, and the role in safety argumentation of the evidence it generates.

The following could be considered for the verification/assessment of HAV performance in scenario tests:

- Scenarios created from the scenario generation step (see Section 5.2.5). These would need to be described and exchanged to the ADSE, Section 3.4.9.4 discusses formats that could support this
- Further scenarios could be selected from the ADSEs own database
- Scenarios could assess and/or characterise the performance of HAV subsystems/components
- Scenarios could be weighted towards boundary / Edge case assessment
- The ADSE could further demonstrate their scenario generation process and their capabilities to execute them on their testbench.
- An advantage of testbench testing is potential for higher volume, controllability, and repeatability.
- Results may need to be shared with Technical Service to facilitate performance assessment (see Section 5.3), potential standards to support format and exchange of data are discussed in Section 3.4.8.

5.2.7 Physical testing

The objectives of physical testing are to:

- Validate HASS perception performance
- Assess and verify HAV behaviour in real world-controlled scenarios

The proposed methodology is that the Technical Service would witness the whole HASS performing in test scenarios in a controlled environment such as test proving grounds.

This should be supported by collecting data on telematics unit and video recordings for to support performance analysis and provide records to support assessment conclusions. It would be beneficial for the Technical Service to access some data from the HASS, for example object detection data from the live dataset, and/or actuator data from the CAN bus, however this could be objectionable to ADSE's due to sharing of potentially IP sensitive data.

The following could be considered for the validation of HASS perception performance

- It is preferable to have access to the ADS's perceived object map so this may be compared with ground truth to validate perception performance. However, in the absence of this, it is possible to create scenarios where the correct perception of objects can be measured by a response in HAV behaviour.
- The tests should be appropriate for the HASS specification attributes such as ODD. The fidelity and coverage of the tests would be proportional to the accuracy of the ground truth and scope of facilities such as types and quantities of test targets.

The following could be considered for the verification/assessment of HAV performance in scenario tests:

- Scenarios created from the scenario generation step (see Section 5.2.5). These would need to be described and exchanged to the ADSE, Section 3.4.9.4 discusses formats that could support this
- Further scenarios could be selected from the ADSE's own database
- Scenarios could assess competence or level of performance of whole HASS
- It is challenging to perform tests where there is risk of collision. This can be managed to an extent with soft targets etc however scenarios could be weighted towards assessing appropriate tactical decision making in scenarios that can be managed before becoming critical, rather than focussing on dynamically critical scenarios
- An advantage of physical testing is potential for higher volume, controllability, and repeatability
- Variability of test execution could be used to an advantage to better reflect real life operational challenges, and guard against systems being trained for test conditions. However, consistency/fidelity of test does need to be managed not to cause unfair assessment outcomes
- Data may need to be shared with Technical Service to facilitate performance assessment (see Section 5.3), potential standards to support format and exchange of data are discussed in Section 3.4.8

5.2.8 Fidelity check

The objective is to validate the fidelity of the toolchain and physical test/trial data.

The methodology for this would be to compare the results recorded from a physical test scenario with the same scenario performed in toolchain mediums.

The following could be considered:

- The usual progression of testing, for risk management purposes, is to undertake toolchain testing first, then follow with physical testing. However, for the case of fidelity checking scenarios it may be advantageous to perform physical testing first and use the ground truth measurement of this to define the toolchain fidelity check scenarios. The reasoning for this is that it may be challenging to replicate the ground truth, such as the movement of test targets, in physical testing whereas this should be straightforward in a capable testbench
- Exact congruence between test mediums should not be expected. Techniques should be applied to align results where variance is not critical to fidelity assessment, such as initialisation offsets
- The assessment of variance could focus on determining whether this is explainable and whether errors have been accounted for to ensure a robust system safety case

5.3 Performance Assessment

For testbench and Physical testing described in Section 5.2 an assessment of the vehicle's performance is required. This involves relating measurements taken during a test to the requirements which define a vehicle's behaviour post-deployment. The evaluation of system safety case evidence, or ongoing monitoring and performance may also involve performance assessments.

5.3.1 Performance Indicators

Section 3.2 of this report categorises performance indicators into four types (event occurrence, performance models, formal specifications and informal specifications) and provides examples of each type. Performance indicators play a key role as a metric used in performance assessments providing traceability between requirements and test outcomes.

A recent report by the RAND Corporation (discussed fully in Appendix B, Section 8.17) splits the challenge of assessing safety into three parts:

- Safety as a measurement, how data can be used to understand the level of safety offered by a system
- Safety as a process, the methods, culture and actions which can help to ensure safe outcomes
- Safety as a threshold, addressing the question of "how safe is safe enough"

To assess "safety as a process", one could evaluate whether the tests which have been done, and the results of those tests, give confidence that the ADS will be safe when deployed. Various standards (e.g. UL 4600, BSI PAS 1881, ISO 26262) recommend that the ADS developer should create a structured argument in the form of a safety case to illustrate how their development processes achieve safety. This safety case will, amongst other things, need to argue that the developer's test results provide evidence that requirements are met. Practically, this means that the tests need to collect data which relates to the performance indicators defined in the requirements. For example, if a performance indicator "rate of collisions per operating hour" is used as part of a requirement, it is clear that tests need to be capable of producing an estimate of that rate.

The strength of the argument linking safety test results to requirements is vital for managing the risk that unexpected outcomes will occur after deployment. However, no test process is perfect, so there will always be an element of risk associated with new deployments. Organisations may choose to have a graded approach to deployment, starting first with lower risk deployments and moving onto higher risk ones once more evidence has been gathered.

5.3.2 Assessment techniques

Several techniques may be applied to assess the performance of a HASS in a scenario test. This is a complex area which is interdependent on the definition and calibration of the test scenario as discussed in Section 5.2.5.

Another work package of CertiCAV, separate to this report, has undertaken work to demonstrate how the framework can be applied to testing of vehicle behaviour in simulation. A separate report has been published for that work package with accompanying source code (Connected Places Catapult, 2021a), (Connected Places Catapult, 2021b). This includes:

- Analysis of the types of requirement which can meaningfully be assessed using an automated testing framework
- Examples of how raw data from a simulator can be used to make these assessments
- A discussion of the practical challenges involved in designing a software framework to implement these principles

5.4 Application to trial deployments

It is intended that a shorter-term system of approval may be drawn out from the long-term framework assurance to support the assurance of short/medium term trial deployments (see Section 1.2.1), these could be considered as 'Advanced Trials'. Formulating this process would be subject to the following considerations:

- The output of the process may be different given that is likely to be associated with more conditional system use and involving ongoing monitoring and review which is typical for a trial. For example, the process could be regarded as an acceptance test, rather than a technical approval

The deployment is likely to be mitigated (see Section 2.1.3.3) to manage the risks associated with deploying systems that are not of production quality and maturity. This typically involves human supervision and reliance on performing DDT intervention/fallback. This involves:

- Areas of safety assurance only relevant to automation below level 4 including: driver training and alertness, takeover control, and response times
- An assessment of what the current validation status of the system is, what hazards this presents and that the mitigations applied adequately control the risks
- On a technical level, this may require assurances that the mitigations are themselves sufficiently reliable, including safety drivers always alert, HMI sufficient, handover requests, simple and robust electronic pathway for emergency stop and/or takeover
- Trials associated with system development would involve ongoing changes to the system software/hardware, and potentially changes to ODD or operational characteristics. This pivots the safety assurance towards change control and process quality and away from assurance being on controlled versions. The exception to this would be if the trials are purely for validation purposes

Clarifying the above considerations is dependent on what an advanced trial constitutes. For instance, whether a human safety driver is on-board and whether their controls are traditional or novel.

The following objectives could be considered for the assurance of trials:

- Checking there is a suitable limitation of risk associated with the public trials
- If applicable, checking there is suitable system and organisational development to achieve longer term goals
- The trials may involve a degree of public sponsorship as part of longer-term strategy to deliver public services or develop industry. If so, it may be applicable to checkpoint their state of development and review their development plan

With reference to the CertiCAV approach described in Section 1.2.1, drawing out the assurance process for trials from a framework developed for longer term approvals has the advantage of providing learning through experience to develop and accelerate the long term approval framework. This adaptation could account for the objectives/considerations above and the following points:

- Establishing the current maturity of system safety, and the current level of safety validation achieved for each system process/item, and:
- Assessing whether mitigations are adequate to cover the uncertainty/risk associated with incomplete validation
- Adjusting the assurances processes to account for the trial considerations and established development maturity and validation
- Independent test coverage, for example in the quantity and detail of scenario generation and testing may be lower than for production system approval, this may be mitigated by enhanced monitoring and/or reduced scale of deployment
- There is likely to be a greater reliance on operational safety, such as monitoring processes, due to the relative uncertainty of system safety
- Involving co-operative processes between stakeholders to manage change control
- Trials could be undertaken in several phases, with checkpoints of testing and acceptance taking place between development steps

5.4.1 Definition of advanced trial

UK CAV literature and practitioners sometimes refer to 'advanced trials', however this is currently not a well-defined term. Therefore, this section provides some discussion to help contribute to the definition of the term Advanced Trial and show the potential relationship with system safety.

The trialling of automated vehicles is currently covered by a code of practice published by UK Government on 6 February 2019⁵¹. Here it is defined that an automated vehicle trial needs to meet the following legal requirements

- A driver is present, in or out of the vehicle, who is ready, able, and willing to resume control of the vehicle; and
- The vehicle is roadworthy; and
- Appropriate insurance in place

Various guidance is provided, including that operators should speak with the road and enforcement authorities, develop engagement plans, and have data recorders fitted. Furthermore, the code of practice discusses advanced trials stating that more advanced trials may currently be outside of the law and may require support and facilitation from the Department for Transport to proceed. It does not appear to be explicit on the distinction between a trial and an advanced trial, but it appears to be implied that an advanced trial is the use of a vehicle that is not facilitated by existing regulations and would therefore require support from authorities.

A vehicle constructed or adapted for a specific purpose and which, owing to its construction or specific purpose may not meet vehicle construction and use requirements, may be recognised as a special vehicle in Statutory Instrument 'The Road Vehicles (Authorisation of Special Types)'⁵². In which case it is stipulated that 'A relevant vehicle that is used on roads must not carry any load or transport goods or

⁵¹ (Centre for Connected and Autonomous Vehicles, 2019)

⁵² (The Road Vehicles (Authorisation of Special Types) (General) Order 2003) article 36

burden⁵³. This does not specify whether this extends to carrying passengers, however it may be implied that testing of vehicles does not extend to providing passenger carrying service.

The Law commissions Automated Vehicles CP 3⁵⁴ states that a UIC must be in or in direct sight of the vehicle. Otherwise, it is classes as a No User-In-Charge (NUIC) vehicle which should be subject to approvals and operated by a licenced operator.

Furthermore, irrespective of existing legal or practical definitions it may be relevant to consider practical areas of risk, including:

- Due to lower system maturity and incomplete validation there may be unforeseen faults and/or unidentified risks
- Involvement of vehicle control over the air introduces functional safety and cybersecurity challenges. The implications depend on the system of control
- Implementing direct actuator control OTA places a challenging burden on validating the safe and secure performance of all associated systems. However, it provides a fallback option in the event of failure of onboard tactical control systems
- Restricting over-the-air control to path planning guidance requests removes risks of OTA control system failure, however the HAV safety is then reliant on validated safe fallback performance on the onboard control systems
- Depending on the deployment specification, safe HAV operation may rely on the safe performance of systems under the responsibility of different organisations. For example:
 - OTA communications over mobile network telecommunications infrastructure
 - V2X data from roadside units operated by traffic authority
- The established legislation that allows on-road testing of vehicles involves a degree of trust in manufacturers. However, if trialling organisations do not have an established background in automotive development there is an increased risk of managing trials to a lower competency, with immature processes to manage validation and risk, managing operational risk, and have a less developed risk culture.
- Existing vehicle construction rules relating to driver controls and visibility are mature and are generally considered safety critical
- Novel vehicle construction or types of control may impact the operator's situational awareness, ability to perceive the environment, or their control of the vehicle

⁵³ (The Secretary of State for Transport, 2003) schedule 11

⁵⁴ (The Law Commissions, 2020)

5.5 Summary findings

This section discussed a process of independent assessment associated with the approval of a HASS.

In Section 5.2 we outline activities at a high level including audit/inspection, road testing, testbench and physical testing. However, refining and developing a more comprehensive low-level guide may be useful to ensure consistency of the assurance process. It will take time and experience to develop a matured system of independent assessment. The content in this section may be considered as a starting point for engagement and development.

We have identified the need to implement processes to capture learnings from current and future consultation/approval activities to facilitate:

- Continuous improvement of system approval activities
- Development and improvement of standards/practice

We have identified three key areas to be covered

1. Learning: There is a need to develop and manage knowledge relating to HASS-related audit and assessment activities. RDW present an approach for this, including lists of inspection/audit questions in their Vehicle Safety and Security Framework ⁵⁵
2. Test Scenarios: There is a need to develop database for logical scenarios and parameter distributions
3. Techniques & Tooling: There is a need to develop capabilities to support assessment activities. This includes modelling tooling, and assessment techniques including test selection and performance assessment

Other key observations for consideration are that:

- Given the finding in Section 5.3.2 that separate test evidence from the perception and behaviour subsystems of an ADScan reduce the quantity of evidence needed, standardisation activities in this area would aid assessment practice. For example, a standard interface between the perception and behaviour subsystems would allow assessors to more easily confirm that expected results have been obtained, and that expected scenarios have been tested.
- Guidelines on what independent tests should be carried out as part of the assurance process are needed. A common stakeholder view is that a short public road "test drive" of the HAV would be appropriate, but other tests may also be cost-effective for increasing confidence in the evidence presented by the ADSE (see Sections 5.2.2 and 5.1.3).
- If independent assurance relies on common, standardised scenarios, then there are advantages to creating these scenarios by an expert group representing a range of stakeholders and by following a best-practice process (Section 5.2.4.1).

⁵⁵ (RDW, 2019)

6 Acknowledgements

Firstly, we would like to thank the project sponsors at IVS for their support of our work, including not only regular meetings despite their busy calendars, but moreover technical conversations, ideas, and reviews via email.

The project Steering Panel, which included representatives from the VCA, CCAV, and Innovate UK as well as CPC and IVS, has provided a regular drumbeat for us to demo our outputs. The panel members have shown unflagging interest in our work and have freely contributed their knowledge to improve our thinking.

Our Advisory Group mailing list contacts, and all the wider CAV community, have followed developments during the project keenly, participated actively in our workshops, and been willing to speak to us whenever we reached out.

Finally, we would like to extend special thanks to the wider reviewers, who generously donated significant amounts of time to provide comments on a draft that was both over 100 pages long and untidy/incomplete in many places.

7 Works Cited

- Aptiv, Audi, Baidu, BMW, Continental, Daimler, FCA, Here, Infineon, Intel, Volkswagen. (2019). *Safety first for automated driving*. Retrieved from <https://www.daimler.com/documents/innovation/other/safety-first-for-automated-driving.pdf>
- ASAM. (2017, July 26). *ASAM OpenSCENARIO*. Retrieved from <https://www.asam.net/standards/detail/openscenario/>
- ASAM. (n.d.). *OpenDRIVE*. Retrieved from <https://www.asam.net/standards/detail/opendrive/>
- Augustin, D., Schucker, J., Tschirner, J., Hofmann, M., & Konigorski, U. (2019, July). A Simulation-Based Reinforcement Learning Approach for Long-Term Maneuver Planning in Highway Traffic Scenarios. *VDI AUTOREG 2019 Volume 2349*.
- Blumenthal, M. S., Fraade-Blanar, L., Best, R., & Irwin, J. L. (2020). *Safe Enough: Approaches to Assessing Acceptable Safety for Automated Vehicles*. RAND Corporation. Retrieved from https://www.rand.org/pubs/research_reports/RRA569-1.html
- British Standards Institute. (n.d.). *CAV Roadmap*. Retrieved from British Standards Institute: <https://www.bsigroup.com/en-GB/CAV/cav-resources/download-cav-roadmap/>
- BSI. (2020). *Connected and Automated Vehicles Vocabulary*. Retrieved from BSI Group: <https://www.bsigroup.com/en-GB/CAV/cav-vocabulary/>
- CAA_A. (n.d.). <https://www.caa.co.uk/Our-work/About-us/Our-role/>. Retrieved from <https://www.caa.co.uk/Our-work/About-us/Our-role/>
- Centre for Connected and Autonomous Vehicles. (2019, February 6). *Code of Practice: Automated vehicle trialling*. Retrieved from GOV.UK: <https://www.gov.uk/government/publications/trialling-automated-vehicle-technologies-in-public/code-of-practice-automated-vehicle-trialling>
- CETTRAN. (n.d.). *Singapore Publishes Technical Reference for AVs*. Retrieved from <https://cetran.sg/tr68/>
- Connected Places Catapult. (2020, March 31). *MUSICC repository*. Retrieved from <https://gitlab.com/connected-places-catapult/musicc>
- Connected Places Catapult. (2021a). *CertiCAV Software Framework: Implementation and evaluation*. Retrieved from <https://cp.catapult.org.uk/news/implementing-and-evaluating-safety-metrics-for-cavs-report/>
- Connected Places Catapult. (2021b, March 31). *CertiCAV source code*. Retrieved from GitLab: <https://gitlab.com/connected-places-catapult/CertiCAV>

- Connected Places Catapult and National Physical Laboratory. (2020, July). *Performance Testing for Sensors in CAV*. Retrieved from CPC publications: <https://cp.catapult.org.uk/wp-content/uploads/2021/01/Performance-Testing-for-Sensors-in-Connected-and-Autonomous-Vehicles-Feasibility-Study.pdf>
- Connected Places Catapult. (n.d.). *MUSICC*. Retrieved from <https://cp.catapult.org.uk/case-studies/musicc/>
- Department for Transport. (2018). *Transport Statistics Great Britain: 2018*. Retrieved from <https://www.gov.uk/government/statistics/transport-statistics-great-britain-2018>
- Department for Transport. (2019, September 26). *Reported road casualties in Great Britain, annual report: 2018*. Retrieved from <https://www.gov.uk/government/statistics/reported-road-casualties-in-great-britain-annual-report-2018>
- Department for Transport. (2019, August 20). *The Highway Code*. Retrieved from gov.uk: <https://www.gov.uk/guidance/the-highway-code>
- Department for Transport. (2020, September 30). *Reported road casualties in Great Britain: 2019 annual report*. Retrieved from [assets.publishing.service.gov.uk: 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)
- Department for Transport. (2020, September 10). *Road Traffic Estimates: Great Britain 2019*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/916749/road-traffic-estimates-in-great-britain-2019.pdf
- Dosovitskiy, A., Ros, G., Codevilla, F., López, A., & Koltun, V. (2017). CARLA: An Open Urban Driving Simulator. *1st Conference on Robot Learning*. arXiv:1711.03938. Retrieved from <http://proceedings.mlr.press/v78/dosovitskiy17a/dosovitskiy17a.pdf>
- EASA. (n.d.). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R0947&from=EN>. Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R0947&from=EN>
- ETSI. (2018, August). *ETSI TS 103 301 V1.2.1*. Retrieved from ETSI.org: https://www.etsi.org/deliver/etsi_ts/103300_103399/103301/01010201_60/ts_103301v010201p.pdf
- FiveAI. (2018, November). *Certification of Highly Automated Vehicles for Use on UK Roads: Creating An Industry-Wide Framework for Safety*.
- Fortellix. (n.d.). *Open M-SDL*. (Fortellix) Retrieved from <https://www.fortellix.com/open-language/>
- Gromke, C. (2019, September). *Aerodynamic loads on a cyclist while overtaking by a vehicle*. Retrieved from <https://www.gala-ev.org/images/Beitraege/Beitraege2019/pdf/7.pdf>
- Hallerback, S., Xia, Y., Eberle, U., & Koester, F. (2018). Simulation-Based Identification of Critical Scenarios for Cooperative and Automated Vehicles. *SAE Intl. JCAV 1(2):93-106*. doi:10.4271/2018-01-1066
- Health & Safety Executive. (2001). *Reducing risks, protecting people*. HMSO.
- Health and Safety Executive. (1999). *The cost to Britain of workplace accidents and work-related ill health in 1995/96*. Retrieved from <https://www.hse.gov.uk/pubns/priced/hsg101.pdf>

- HM Government. (2017). *The Key Principles of Cyber Security for Connected and Automated Vehicles*. London. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/661135/cyber-security-connected-automated-vehicles-key-principles.pdf
- Hyden, C. (1987). *The development of a method for traffic safety evaluation*. Retrieved from ictct.net: https://www.ictct.net/wp-content/uploads/SMoS_Library/LIB_Hyden_1987.pdf
- ICAO_A. (n.d.). <https://www.icao.int/airnavigation/Pages/default.aspx>. Retrieved from <https://www.icao.int/airnavigation/Pages/default.aspx>
- ICAO_B. (n.d.). <https://www.icao.int/safety/airnavigation/OPS/Pages/default.aspx>. Retrieved from <https://www.icao.int/safety/airnavigation/OPS/Pages/default.aspx>
- ICAO_C. (n.d.). <https://www.icao.int/safety/airnavigation/OPS/Pages/flsresp.aspx>. Retrieved from <https://www.icao.int/safety/airnavigation/OPS/Pages/flsresp.aspx>
- IEEE. (n.d.). Retrieved from WG: VT/ITS/ AV Decision Making: <https://sagroups.ieee.org/2846/>
- Karla, N., & Paddock, S. M. (2016). *Driving to safety: how many miles of driving would it take to demonstrate autonomous vehicle reliability*. RAND Corporation. Retrieved from https://www.rand.org/pubs/research_reports/RR1478.html
- Klamann, B., Lippert, M., Amersbach, C., & Winner, H. (2019, November 28). *Defining Pass/Fail-Criteria for Particular Tests of Automated Driving Functions*. Retrieved from <https://ieeexplore.ieee.org/abstract/document/8917483>
- Koopman, P., Kane, A., & Black, J. (2019). *Credible autonomy safety argumentation*. Retrieved from http://users.ece.cmu.edu/~koopman/pubs/Koopman19_SSS_CredibleSafetyArgumentation.pdf
- Legislation.gov.uk. (n.d.). Retrieved from Road Traffic Act 1988, Part 1, Section 2A: <https://www.legislation.gov.uk/ukpga/1988/52/section/2A>
- Menzel, T., Bagschik, G., & Maurer, M. (2018). Scenarios for development, test and validation for automated vehicles. *2018 IEEE Intelligent Vehicles Symposium*. arXiv:1801.08598.
- MISRA consortium. (2019). *Guidelines for Automotive Safety Arguments*. UK: MISRA C. doi:ISBN 978-1-906400-24-8
- Mobileye. (2017). *On a Formal Model of Safe and Scalable Self-driving Cars*. Retrieved from arxiv.org: <https://arxiv.org/pdf/1708.06374.pdf>
- Myers, R., & Saigol, Z. (2020). Pass-fail criteria for scenario-based testing of automated driving systems. *Submitted for publication*.
- Myers, R., Nichols, S., & Saigol, Z. (2021). *CertiCAV Software Framework: Implementation and evaluation*. Connected Places Catapult.
- Nanyang technical university, Singapore. (n.d.). *Centre of Excellence for Testing & Research of AVs*. Retrieved from <http://erian.ntu.edu.sg/Programmes/IRP/FMSs/Pages/Centre-of-Excellence-for-Testing-Research-of-AVs-NTU-CETRAN.aspx>
- NASA. (2011, November). *NASA System Safety Handbook, Volume 1*. Retrieved from NASA: <https://ntrs.nasa.gov/api/citations/20120003291/downloads/20120003291.pdf>
- Nvidia. (2019). *An Introduction to the Safety Force Field*. Retrieved from Nvidia.com: <https://www.nvidia.com/content/dam/en-zz/Solutions/self-driving-cars/safety-force-field/an-introduction-to-the-safety-force-field-v2.pdf>

- Parliament. (1835). *Highway Act 1835, Section 78*. Retrieved from legislation.gov.uk:
<https://www.legislation.gov.uk/ukpga/Will4/5-6/50/section/78>
- PEGASUS Projekt. (2019). *Pegasus Method: An Overview*. Retrieved from
<https://www.pegasusprojekt.de/files/tmpl/Pegasus-Abschlussveranstaltung/PEGASUS-Gesamtmethode.pdf>
- PEGASUS Projekt. (2019, May 14). *Scenario Description*. Retrieved from
https://www.pegasusprojekt.de/files/tmpl/PDF-Symposium/04_Scenario-Description.pdf
- Pütz, A., Zlocki, A., Küfen, J., Bock, J., & Lutz, E. (2017, June). *Database approach for the sign-off process of highly automated vehicles*. Retrieved from
<http://indexsmart.mirasmart.com/25esv/PDFfiles/25ESV-000060.pdf>
- R.Queiroz, Berger, T., & Czarnecki, K. (2019). GeoScenario: An open DSL for autonomous driving scenario representation. *IEEE Intelligent Vehicles Symposium (IV)*.
- RDW. (2019). *GRVA-01-40*. Retrieved from UN-ECE:
<https://unece.org/DAM/trans/doc/2018/wp29grva/GRVA-01-40.pptx>
- Reid, T. (2019, Jun 3). *Localization Requirements for Autonomous Vehicles*. doi:arXiv:1906.01061
- SAE International. (2018). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. On-Road Automated Driving (ORAD) committee.
doi:https://doi.org/10.4271/J3016_201806
- Saigol, Z., & Peters, A. (2018). Verifying automated driving systems in simulation: framework and challenges. *25th ITS World Congress*. Copenhagen.
- Schöner, H. (2020). *Challenging Highway Scenarios Beyond Collision Avoidance for Autonomous Vehicle Certification*. IFO-Consulting. Retrieved from
https://1drv.ms/b/s!Alg5jpeKONBqgYRCnAaQMelcG_oFOw
- Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2017, August 21). *On a Formal Model of Safe and Scalable Self-driving Cars*. (Mobileye) doi:arXiv:1708.06374 [cs.RO]
- The British Standards Institution. (2018). *BSI PAS 1885: The fundamental principles of automotive cyber security. Specification*. London: The British Standards Institution.
- The British Standards Institution. (2020). *BSI Connected and automated vehicles - Vocabulary BSI Flex 1890 V3.0:2020-10*. London: The British Standards Institution.
- The British Standards Institution. (2020). *BSI PAS 1881: Assuring the safety of automated vehicle trials and testing - Specification*. London: The British Standards Institution.
- The Law Commissions. (2019). *Automated Vehicles: Consultation Paper 2*. London: Law Commission. Retrieved from <https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jxou24uy7q/uploads/2019/10/Automated-Vehicles-Consultation-Paper-final.pdf>
- The Law Commissions. (2020). *Automated Vehicles: Consultation Paper 3*. London: Law Commission. Retrieved from <https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jxou24uy7q/uploads/2021/01/AV-CP3.pdf>
- The Police Foundation. (2020). *Roadcraft*. Retrieved from Roadcraft: <https://www.roadcraft.co.uk/>
- The Secretary of State for Transport. (2003, August 4). *The Road Vehicles (Authorisation of Special Types) (General) Order 2003*. Retrieved from <https://www.legislation.gov.uk/>:
<https://www.legislation.gov.uk/uksi/2003/1998/contents/made>

- The Stationary Office. (2016). *The Traffic Signs Regulations and General Directions 2016, Schedule 1*. Retrieved from Legislation.gov.uk:
<https://www.legislation.gov.uk/uksi/2016/362/schedule/1/made>
- Thorn, E., Kimmel, S., & Chaka, M. (2018). *A framework for automated driving system testable cases and scenarios (DOT HS 812 623)*. National Highway Traffic Safety Administration.
- TNO. (2018, July). *StreetWise*. Retrieved from
<https://publications.tno.nl/publication/34626550/AyT8Zc/TNO-2018-streetwise.pdf>
- Uber ATG. (n.d.). *Uber ATG Safety Case Framework*. Retrieved from Uber ATG Resources:
<https://uberatgresources.com/safetycase/gsn>
- UK Government. (2008). *Road Traffic Act 1988 Section 3ZA (as amended)*. Retrieved from
<http://www.legislation.gov.uk/ukpga/1988/52/section/3ZA>
- Ulbric, S., Menzel, T., Reschka, A., Schuldt, F., & Maurer, M. (2015). Defining and substantiating the terms scene, situation, and scenario for Automated Driving. *18th IEEE ITSC*.
- UNECE. (2019). *Safety Criteria Study on Innovative Safety Validation Methods of Automated Driving System*. Retrieved from <https://wiki.unece.org/download/attachments/87622683/VMAD-04-04%20Japan%20safety%20criteria%20study.pdf?api=v2>
- UNECE. (2020). *Proposal For a New UN Regulation on Uniform Provisions Concerning the Approval of Vehicles With Regard to Automated Lane Keeping Systems*.
- United Nations. (2013). *Uniform provisions concerning the approval of motor vehicles with regard to the Advanced Emergency Braking Systems (AEBS)*. Retrieved from [unece.org](https://unece.org/fileadmin/DAM/trans/main/wp29/wp29regs/2013/R131e.pdf):
<https://unece.org/fileadmin/DAM/trans/main/wp29/wp29regs/2013/R131e.pdf>
- United Nations. (2017). *UN Regulation No. 79*. Retrieved from United Nations Economic Commission for Europe: <https://unece.org/fileadmin/DAM/trans/main/wp29/wp29regs/2017/R079r3e.pdf>
- Zenzic and Ordnance Survey. (2019). *Geodata Report - analysis and recommendations for self-driving vehicle testing*. Retrieved from <https://zenzic.io/insights/geodata-report/>
- Zenzic. (n.d.). *Zenzic Safety Case Framework*. Retrieved from <https://zenzic.io/reports-and-resources/safety-case-framework/>
- Zhao, D., & Peng, H. (2017, May). *From the lab to the street: solving the challenge of accelerating automated vehicle testing*. Mcity whitepaper: arXiv:1707.04792. Retrieved from
https://mcity.umich.edu/wp-content/uploads/2017/05/Mcity-White-Paper_Accelerated-AV-Testing.pdf

8 Appendices

A Workshop & Expert Consultations

8.1 Workshop 1: Consultation Webinar

This workshop held on 21st July 2020 shared our initial understanding of the challenges of CAV assurance, and showcased some initial solutions to the challenges that were developed by the CertiCAV project. A total of 123 people attended with 92% of attendees remaining from start to end.

Figure 29: Workshop 1 attendance



Following the event, we held 10 small group feedback sessions with a total of 38 guest attendees. During these sessions we reviewed the content presented asked questions on the following topics.

- Safety criteria
- Scenario based validation methods
- Independent assessment process
- Any other thoughts, issues, or comments

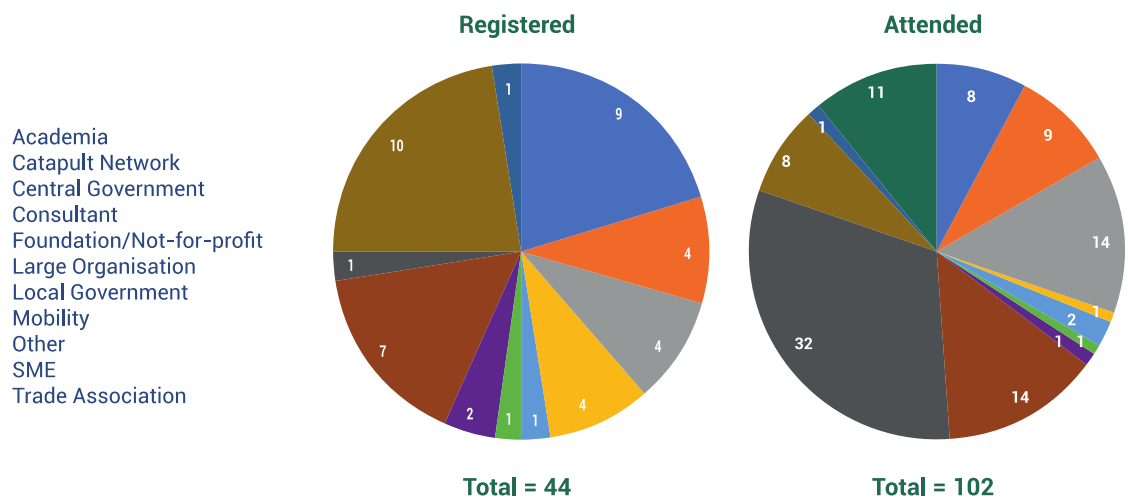
From these sessions we captured 42 key feedback points that were considered in the development of this report. In summary, these included the following observations:

- Clarify how CertiCAV fits in wider strategy – e.g. CAVPASS
- Identified other projects to review including: ALKS, VeriCAV, dRISK, BSI simulation workshop
- Validation that Insurance providers are a stakeholder in assurance
- Consider interaction between different evidence 'pillars'
- Consider checks / accreditation of simulation tooling
- Consensus/validation of referring to common evaluation models rather than proprietary
- Consider legal vs technical interpretations
- Quadrant diagram visualisation model of driving: An interesting approach to a complex issue however is confusing unless it is explained carefully

8.2 Workshop 2: HAV Assurance, Requirements, Goals, & Strategy

This workshop provided an updated view of HAV assurance challenges and solutions under development. This sustained the engagement of the target audience for CertiCAV with 102 people attending and 84% of attendees remaining from start to end. In response to the workshop we were contacted by two technology developers and had follow up discussions with them.

Figure 30: Workshop 2 attendance



8.3 Aviation Interview Study

8.3.1 Introduction

The assurance of HAVs shares many of the challenges faced by commercial and general aviation with the introduction of Unmanned Aerial Systems (UAS), for example:

- The involvement of highly complex safety critical electro-mechanical systems.
- The safety validation of complex software control systems.
- The involvement of simulation tooling and testbench in verification and validation activities.

There may be learnings to be brought across from the aviation domain as a mature and internationally regulated domain; to investigate this, we interviewed international aircraft and unmanned drone stakeholders. With a view to capture relevant technical policies, regulations, standards, or codes/practices, which would support the framework, the survey was based on the following focusing questions:

- What are the key principles/objectives of the assurance process?
- How do the organisations involved (suppliers, developers, manufacturers, inspection bodies, regulators) share the burden of assurance, and how do the regulators ensure integrity and trust for the whole the chain?
- How can independent certification bodies access sufficient evidence? And what should be acceptance (pass/fail) criteria?
- What technical solutions should be demonstrated to satisfy the required assurance levels for complex automation?
- What aspects of the assurance system work well e.g. for complex systems, and what aspects are becoming increasingly challenging to scale to increased complexity?

Section **8.3.2** below discusses key recommendations captured during the interviews. Then Section **8.3.3** provides a brief overview of the existing relevant regulations that may be of interest for HAV assurance. Finally, in Section **8.3.4** we provide conclusions from this interview study.

8.3.2 Survey Outputs

Based on survey and interviewees' view, there are several areas to address in creating a safety framework for HAVs. However, the top priorities are related to the principles and key considerations of such a system.

In particular, the interviewees expressed that there is a need for an international organisation to set overarching minimum regulations that would ensure the safety of HAVs. These regulations should define assurance levels and ensure that the benefits outweigh risks within each potential operational domain. Furthermore, this organisation must use principles of road safety from all participating countries and therefore support international (cross border) operations. The global outlook of this organisation would help develop the initial thinking around policy and guidance, which could be taken forward by other entities and experts such as associations and international standards organisations.

The survey indicates that the principles for the HAVs should be based on:

- Safety assurance
- Transparent and traceable procedures and processes
- Evidence – Risk-Based Certifications
- Best-practice approach
- Competent assurance methods
- Delegation of authority
- Qualified and Competent Operators/ Personnel and skills competency matrix including toolsets used (e.g., finite element analysis, etc.)
- A safety framework for HAVs might be based on the current international framework with the need of amending or creating new (if the integration complexity is high) regulations to adopt such vehicles with the cooperation and consultation of related expert groups

In the UAS domain, the survey captured the following the main considerations necessary to ensure adequate safety:

- Regulatory categories should be created, based on the use case and the size, maximum altitude, and maximum take-off weight of the vehicle, as well as its operational complexity/risk. These factors would be captured in several documents, such as the MAA Regulation 1600, Commission Implementing Regulation (EU) 2019/947 and CAP722. Hence, the level of assurance of safety-critical aspects would be dependent upon the risk of each category.
- The above regulatory categories should consider the environment of operation and the exposure of 2nd and 3rd parties in that environment.
- In the case of category submission by an organisation for a UAS, there is a need for a categorisation safety checklist for the organisation, the UAS, and its operational procedures.
- The risk assessment of each category should be evidence-based through tests for specific deployment domains and environmental circumstances. This will also demonstrate the performance of several safety-critical components, such as operating performance under contingencies, software code quality and so forth. Note that categories with higher risk will require more independent testing and more safety evidence to be submitted.
- Implementation of an independent evaluation for each safety case during testing.
- Use of widely accepted industry standards.
- Use of an operation-specific safety equation to measure the risk-to-life, accident rate/likelihood estimations, risk of damage to property, etc.
- Use of appropriate technology to mitigate conflicts and provide separation of vehicles and other users.
- The vehicle manufacturers must be certified and proven entities.
- Use of specific documentation including manuals and records that will be reviewed independently by the approval issuer.
- Provide redundancy in UAS systems to avoid a single point of failure.
- Provide insurance for all operations, as this provides a further safeguard and helps ensure only robust and safe operations are permitted.

8.3.3 Relevant Regulation

In the aviation sector, the overall governing organisation is the International Civil Aviation Organization (ICAO). One of the objectives of this organisation is aviation safety. As stated in (ICAO_A, n.d.) the organisation is constantly striving, in close collaboration with the entire air transport community, to further improve aviation's successful safety performance while maintaining a high level of capacity and efficiency. This is achieved through:

- The development of global strategies.
- The development and maintenance of standards, recommended practices, and procedures applicable to international civil aviation activities.
- The monitoring of safety and air navigation trends.
- The implementation of targeted safety and air navigation programmes to address deficiencies.
- An effective response to the disruption of the aviation system created by natural disasters, conflicts, or other causes.

The same resource states that in all of its highly coordinated activities, ICAO strives to implement practical and achievable measures to improve safety, capacity and efficiency in all sectors of the air transport system to ensure that aviation's achievement of a remarkably safe and efficient transport network continue to serve a fundamental role in supporting global social and economic priorities.

The ICAO includes the Operational Safety Section (OPS), which is responsible for the development of standards, recommended practices, procedures and guidance material related to the operation and certification of aircraft (ICAO_B, n.d.). A detailed list of responsibilities of the OPS, such as airworthiness of aircraft, training policy and standards, flight procedures, flight safety and human factors, and safety management systems is given in (ICAO_C, n.d.).

ICAO maintains close technical level collaborations with the breadth of the air transport community. This includes drafting technical proposals and ensuring these are developed in collaboration with relevant wider expert groups. The outputs of this process are regulations and standards that are mature enough for adoption and implementation within the member states.

In order to enable remotely piloted aircraft operations, as has been mentioned in previous studies, three main ICAO Annexes (licensing/competencies, airworthiness, and flightworthiness) might need to be amended from the current global regulations.

At a lower level, there are several layers of regulations, standards and requirements provided by more regional authorities, such as European Aviation Safety Agency (EASA) regulations for EU aviation. The purpose of these is to define rules, procedures, acceptable means of compliance and guidance material that specify key principles, conditions (e.g., for obtaining certificates), compliance, pricing, etc. for different systems, features, services, and stakeholders. To integrate UAS operations, EASA is currently working on the development of a regulatory framework for UAS traffic management, and has defined three operations categories (open, specific, and certified) using a risk-based approach. The rationale of this process is that the rules and procedures applicable to UAS operations should be proportionate to the nature and risk of the operation and adapted to the operational characteristics of the unmanned aircraft concerned, and the characteristics of the area of operations (such as the population density, surface characteristics, and the presence of buildings) (EASA, n.d.).

At a national level, there is a requirement for an aviation authority/regulator to define and apply an appropriate and proportionate regulatory and certification system for civilian-use aircraft to ensure they are safely and efficiently operated. In the UK, the Civil Aviation Authority (CAA) has harmonised with most of the aviation regulation and policy across the world (as set by ICAO and EASA) to ensure consistent levels of safety and consumer protection. Areas that are not covered by EU Regulations are

regulated by the Air Navigation Order (ANO) that is the statutory instrument made under the Civil Aviation Act (1982) and is overseen by the CAA (CAA_A, n.d.). The CAA works to ensure that:

- The aviation industry meets the highest safety standards.
- Consumers have a choice, value for money, are protected and treated fairly when they fly.
- Through efficient use of airspace, the environmental impact of aviation on local communities is effectively managed and CO2 emissions are reduced.
- The aviation industry manages security risks effectively.

The current UK framework and ANOs support the civil operation of unmanned aerial vehicles especially in Visual Line of Sight (VLOS). However, there is more work that is underway towards a higher level of automation, such as for Beyond Visual Line Of Sight (BVLOS), via initiatives supported by the Innovation Team at the Department for Transport such as the Government Drone Pathfinder and Open-Access UAS Traffic Management systems. Other organisations, such as the Department for Business, Energy & Industrial Strategy (BEIS), Department for International Trade & Department for International Development, Aerospace Technology Institute (ATI), UK Research and Innovation (UKRI), and more industry group and associations also run initiatives.

A full list of relevant standards bodies from the aviation domain that could help inform HAV regulation includes:

- | | |
|------------|--|
| • ASD-STAN | • EUROCONTROL / SESAR JU research projects |
| • AUVSI | • GUTMA |
| • CAA | • ICAO |
| • EASA | • JARUS |
| • EUROCAE | • NATO |
| | • RTCA |

8.3.4

Conclusion

The purpose of this study was to collect information from the aviation domain to support the development of an assurance framework for HAVs. The interviews conducted as part of this study have laid the groundwork for future work on topics such as principles and considerations, assessment methods, and overall criteria. The study has also described a high-level trust chain that is developed from the international to regional and then national level for the aviation domain, and this trust chain currently supports aerial automated vehicles and could be implemented for HAVs as well.

This study has found several key principles, considerations and assessment processes that might be implemented directly to a safety framework for the autonomous road vehicles sector. However, further work is needed to capture performance metrics, key "pass/fail" criteria, different safety assessment processes, and verification and validation procedures that could support or be mapped against the autonomous road vehicles sector.

Overall, this study identified that both aviation and autonomous road vehicles sectors could benefit from the exchange of ideas and knowledge transfer in several common research areas.

8.3.5

Acknowledgement

We would like to offer special thanks to the following organisations for joining in the survey and providing insightful information and comments.

- | | |
|---------------------|----------|
| • Altitude Angel | • Boeing |
| • ANRA Technologies | • CAA |

8.4 WMG Interview Study

8.4.1 **Method:**

We used a semi-structure interview study to gather stakeholder input. Semi-structured interviews provide the ability to have richer data from the interviewees (as compared to surveys and questionnaires). Additionally, they allow the flexibility to examine topics in different degrees of depth (as per interviewees' interest and background). The interviews were transcribed, and the text was sanitized to remove any proprietary mentions. A coding analysis was performed on the sanitized text and themes and categories were identified from the various interview answers.

In order to prevent any bias, the interviewees were allowed to talk freely while answering the questions and were not prompted for any answers. Participant interviews were transcribed into text and were later coded to perform thematic analysis. Key themes were identified from the transcribed text.

8.4.2 **Participants:**

Participants were selected from a diverse demography cutting across the automotive. Eleven participants representing OEMs (Original Equipment Manufacturers), Tier 1/2 suppliers, regulators and academic/research organizations/start-ups working in the area of automated driving were interviewed. Each participant had over 10 years' experience in the field of testing and development of systems in the automotive industry, especially ADAS and automated driving. To ensure independence of the interviewees, participants were recruited from different countries including the Australia, Canada, USA, and Germany.

8.4.3

Interview design:

The following table illustrates the guiding (main) questions asked. Based on the interviewee's responses, the follow-up questions were asked.

Guiding Question	Question 1:	Question 2:	Question 3:
	What are the key aspects of safety assurance for CAVs deployment?	How will a certification framework benefit the development and large-scale deployment of CAVs?	What are the essential requirements for a CAV certification framework?
Check what was mentioned by interviewee	Appropriate safety assurance for the automation level and level of maturity of the technology.	Key benefits How it will help the industry progress (safety, cyber security, common standards)	Global best practices Consistency Considerations for development of a new framework (if starting from scratch)
	Role of simulation, controlled and public road trialling in V&V.	How it can influence customer/public confidence in CAVs	Existing frameworks – positives/negatives
Follow-up question	How can a certification framework address this for different levels of automation/services/ODDs?	What are the issues that a certification framework will need to address?	What are the learning points from development of similar frameworks?
	How can safety assurance be evidenced accurately and consistently within the safety case?	What are the main standards to certify against?	Is self-certification a viable option for this technology? Compare against type approval If not, who should do the certification? Specific body? Independent certification?

Guiding Question	Question 4:	Question 5:	Question 6:
	It is accepted that simulation will play a key role in the certification of CAVs. What level of evidence is needed to trust simulation results?	Scenario based testing is widely accepted as the way forward for certification. How would scenarios be used in certification?	Between Audit and test execution, how would you choose certification approval methods?
Check what was mentioned by interviewee	Quantity, type, and level of fidelity/reliability of simulation required Key scenarios to simulate and how to match this with the ODD Edge cases Law of diminishing returns (at what point has enough simulation been done (from a certification perspective). Iterative process?))	Relationship between the Operational Design Domain and certification. Scenario database	Audit at what level? Fully life cycle? EuroNCAP is not an audit but a spot test. If test execution is chosen, how should the decision between real-world and simulation test be made? Could a hybrid be chosen?
Follow-up question	How can this level of fidelity/reliability be proven? How can the repeatability/consistency of simulation for certification be ensured?	How do you pick which scenarios within the database for certification approval?	Would OEMs/developers be open to audit?

8.4.3.1

On question 1:

One of the primary requirements from all stakeholders was the need for reproducibility of the safety assurance process while ensuring the scenarios used as part of the process represent a real-world situation. Certain regulators mentioned the need to ensure that the safety assurance process was compliant to standards. Another important point raised multiple times was the incorporation of edge cases/ near miss scenarios or common risk situations as part of the assurance process. However, the rigour of the process needed to be calibrated to the ODD of the deployment.

8.4.3.2

On question 2:

One of the major benefits that was suggested around having a robust certification framework was that it will help address some of the concerns around the liability. It was suggested that a safety score like that of Euro NCAP that public could understand would help create more trust among public for HAVs.

8.4.3.3**On question 3:**

Once again, reproducibility was highlighted as one of the essential requirements for a CAV certification framework. Another recurring theme was the need for version control in the certification process. Stakeholders envisioned that version control would be needed for both scenarios as well as the tools/toolchains used as part of the certification framework. The need to adopt standards as a part of the certification framework was also highlighted. However, certain stakeholders also highlighted the lack of relevant standards to cover the entire certification framework.

8.4.3.4**On question 4:**

One interesting aspect which was highlighted was the fact that proving ground tests have limited use for higher levels of automation as they lack the detail, especially for systems with complex ODDs like that of city driving. In these cases, proving ground tests will mostly be for braking and collision detection. Thus, the role of simulation was even more important for complex ODDs.

From a trust perspective, it was suggested that there is a need to create a standard set of key performance indicators (KPIs) for simulation performance. While there is a need to use KPIs to compare real-world trace with simulation trace to get confidence in simulation, it was suggested that such KPIs need to be determined by the function being tested. For example: an HMI test will have different set of KPIs for the simulation as compared to a simulation setup testing HAV's perception functionality.

Regulators interviewed seem to have limited experience with simulation based testing and it will require the development of new internal processes for the regulator to be able to use simulation. Until then, they will be guided by the manufacturers.

8.4.3.5**On question 5:**

It was suggested that scenario databases will enable scenario sharing and is a necessity to prevent re-invention of the wheel. There was a suggestion that organisation have a "moral responsibility" to share scenarios. However, while some stakeholders suggested that scenario databases should focus on "common driving situations" and "common manoeuvres", others suggested edge/corner cases should also be included. There was a common understanding that all scenarios should be mapped to their ODDs. However, there is a need to check how a system copes with invalid values or out of ODD situations. Additionally, it was suggested that scenarios in a database used for certification need to be safety related. Certain stakeholders also mentioned the need for a quality assurance if a scenario database is used as part of a certification framework.

Another theme that emerged from the interviews was about the selection of scenario parameters for the testing/spot checks. It was suggested that this selection needs to be based on the analysis of real world data and its distributions.

From regulatory perspective, it was suggested that scenarios needed to be scored based on complexity (ODD, manoeuvres etc.) and a database needs to enable a powerful search based on ODD and the complexity score.

8.4.3.6**On question 6:**

There was a common understanding that OEMs are unlikely to hand over IP voluntarily. Additionally, every OEM is likely to have their own ODD and may not be willing to use methods/processes for another (even if similar) ODD.

While it was suggested that there will be a need for spot checks (scenario based testing in virtual and real-world environment), a concern regarding toolchain compatibility was raised. Here again, the need and implementation for standards was highlighted a potential way forward which could enable a bi-direction API setup.

One unique but important viewpoint was the need from an audit perspective to understand how the test scenarios were arrived at (generation process or motivation). From a regulatory perspective, it was suggested an audit process could be immensely supported if standard documentation requirements could be defined.

B State of the Art

8.5 Law Commissions Consultations

The Law Commission of England and Wales and the Scottish Law Commission have been asked by CCAV to undertake a far-reaching review of the legal framework for automated vehicles, and their use as part of public transport networks and on-demand passenger services.

This work has produced three consultation papers, of which the third and final paper, at the time of writing, is currently published for consultation feedback.

The publications and further details are available on the project webpage <https://www.lawcom.gov.uk/project/automated-vehicles/>

In this section we have tabulated key proposals from the 1st and 2nd completed consultations, noting the reported conclusion of the consultation feedback and the CertiCAV influence and observations relating to these proposals.

8.5.1 Consultation 1

Proposal	Consultation Feedback	CertiCAV influence & observations
Proposal: ADS submitted for approval by an ADSE, with legal responsibility to ensure systems are safe	Strong agreement	Assumption for RASIC & licencing/ assurance model
Proposal: new safety assurance scheme to authorise ADS. Unauthorised ADS prohibited. Any non-authorised ADS is prohibited. ADS added as a modification to a type approved vehicle should go through a UK approval process	Generally supported. Safety assurance must consider through-life issues. Self-certification not enough on its own.	Assumptions -ADS will require an approval -Independent verification is required
Possibility: May be a need for a national accident investigation branch. Alternatively, a facility for providing specialist expertise to police forces.	Some support - only issue unites the different road user groups! (not the police, SMMT or lawyers though). Possible use as short term support for police.	In use aspect is out of scope. However, capture the recommendation in background/RASIC as the approval documentation/evidence would be control documents that investigators refer to.

Mentioned as a possibility: new reporting requirements to support safety assurance scheme. Mention of new data set to be retained in an accident.	Mixed. Developers requested more information about what data they should retain from accidents.	In use aspect out of scope, however, we may undertake assessment of data from on-road trials which would have similarities in data reporting. No mention of collecting equivalent baseline data from manual vehicles - e.g. if a requirement to report near misses or minor collisions is introduced
Proposal: By default, all automated vehicles should have a user in charge. Vehicles without a user in charge will need special approval and to be run by a licenced operator. User in charge not intended to include remote operator. Note that this excludes valet parking if there is a need to use public roads to reach a car park.	Generally supported.	Assuming that we're considering the licencing of ADS without UIC then we could assume the proposal that there will be a licenced operator.
Statement: Government has indicated that only vehicles capable of achieving a minimal risk condition will be listed as capable of safely driving themselves.	N/A	Assessment to include MRC
Proposal: User in charge has no responsibility for monitoring DDT when ADS engaged. However, demand to take over after an imperfect minimal risk condition is reached (e.g. stopped on red route). Should therefore be fit to drive if required.	Mostly in favour but surprisingly mixed - people like having a driver to blame. A few gave examples of edge cases (e.g. handover to automated system while driving towards pedestrians)	Out of scope assuming we're looking at L4.
Proposal: User in charge (UIC) is responsible for issues like maintenance, insurance, and children's seatbelts	Generally supported, though some suggest owner/keeper of vehicle responsibility	This'd be licenced operator based on the L4 assumption.
Proposal: The UIC is responsible for driving once they have accepted a handover. However, the ADS may still be held responsible for crashes if the hazard existed before the	No clear conclusions. Concern that unplanned hand overs could create a hazard.	Out of scope assuming we're looking at L4. Interesting implication that manufacturers might prevent UIC from taking

human took over (reference to US approach).		control in an emergency (in case they crash)
Possibility: owners may contract with a service provider if they wish to operate without a user in charge. This organisation would take over many HARPS operator responsibilities.		Out of scope assuming we're looking at L4.
Proposal: Responsibility for safe execution of DDT rests with ADSE	Mostly supported, with some exceptions. Need to deal with patterns of incidents.	Reinforces first point.
Statement: A DHC with precise rules for every instance is not possible		Input into whatever assumptions we make about role of DHC
Possibility: Driver may only be allowed to participate in secondary activities if the vehicle is designed to reach MRC (i.e. SAE level 4). Other less restrictive options also available.	Clear that secondary activities are OK at L4, debate at L3	Out of scope assuming we're looking at L4.
Statement: Regulators and developers need to collaborate on developing digital road rules (but the form of this is yet to be determined).	General agreement. Need for a balance of safety and usefulness. Suggestion from Five that the code should define what should be achieved, not how. May be some cases where CAVs have different rules (e.g. platooning) Output: Government should consider setting up a new forum for application of rules to self-driving vehicles.	Input into whatever assumptions we make about role of DHC
Possibility: Organisation responsible for authorising ADS should also have responsibilities for marketing materials, market surveillance (e.g. recalls) and roadworthiness tests	Majority, but not total, agreement - e.g. SMMT thought existing agencies (DVSA) should remain responsible.	Out of scope
Possibility: New MOT style tests for automated vehicles - may need to be overseen by same organisation as above		Out of scope, perhaps capture the role that documents generated by approval may input into
Proposal: Regulator should apply sanctions to the ADSE when its ADS acts in a way which would be an offence for a human driver	Widespread agreement. TSC response that it should be about patterns, not individual failings.	Care needs to be taken how this is applied. Human drivers commit offences with good intentions and enforcement discretion is applied.

		It may be imprudent to penalise an ADSE for implementing collision avoidance manoeuvres that may violate traffic rules.
'Envisage': A government agency will be responsible for monitoring that whatever safety standard is set is achieved (e.g. statistically). Proposal: the accident rate should be compared to human drivers. Statement: Regulatory tools already exist to require recalls.	Broad support, some suggestion that results should be published. Difficult to draw direct and valid comparisons, needs specialist skills. Vital to understand "types and reasons" for collisions, not just number (TSC response)	Mention that whatever criteria are defined may then be monitored.
Proposal: no review of civil liability. Causation of a collision (for civil liability) is left to the courts to determine Statement: Product liability for manufacturers means that they are responsible by default for damage caused by their products. However, there is a "state of the art" defence if they could not reasonably have discovered the fault. Courts are likely to follow the safety standard set by the regulators. This means that civil liability is unlikely to be an important form of accountability for design defects in the UK.	Mixed. Many (especially insurance) would like clearer guidance in law. This might go much further than RSS causation - e.g. a cyclist frightened by a vehicle changing lanes erratically.	Consideration for setting of safety criteria. This is related to the issue of not being able to determine the "correct" outcome for any scenario. Risk based approach solves verification problem but not legal one. Need to be careful when making statements like "do not cause a crash". Out of scope (should be covered by insurance)
Some rules will exist for data preservation in the event of an accident	Comprehensive detailed description provided	Out of scope
Possibility: Responsibility for driving in a prohibited area rests with the user.	Mixed, strong arguments that it should lie with the ADSE	Out of scope

Proposal: User in charge (where present) responsible for duties following accident	May be unreasonable if involved in secondary task	Out of scope
Proposal: User in charge (where present) must follow instructions from a police officer (after the vehicle has stopped)	Difficulties around secondary tasks	Out of scope
Statement: Interfering with roads or vehicles is already illegal. Possibility: Causing death or serious injury could be added as an aggregating factor.	Not intended to cover repairers/developers.	Out of scope
Possibility: New offense of death or serious injury by developer wrongdoing (e.g. hiding test results)	Majority support. Might be difficult to show causation. Question about whether this applies to deliberate omission only or general technology failings.	
Possibility: audits for training data, ethics policies to be published and publishing ethics policies	Many supporters approved of transparency, but others questioned value. Many thought ethics guidance should come from the government (e.g. whether the safety of people should be prioritised over animals).	If there is some specification on what the ethics policies are then this could be assumed as input to independent assessor.
Possibility: Some other information may need to be published	Themes in responses: Maximum transparency - e.g. release all testing an assessment outcome to the general public. Collision data - to apportion blame, carry out investigations and collect statistics. Capabilities and limitations of the system - operator info, reasons for updates, safety info (e.g. Euro NCAP results). Nothing at all, type approval should cover all requirements - supported by BMW UK	If there is some specification on what is to be published, then this could be assumed as input to independent assessor.

8.5.2

Consultation 2

Proposal	Consultation Feedback	CertiCAV influence & observations
Proposal: single national scheme for HARPS operator licencing [CQ1]	Agreement. Many replies highlighted need for local involvement. Should be consistent with international norms.	
Proposal: national scheme of basic safety standards for HARPS [CQ2]	Strong agreement.	Associated with Deployment specification
Proposal: HARPS operator is a business which carries passengers for hire or reward using HAVs on a road without a human driver or user in charge [CQ3]	Generally supported. Some would go further - apply to all vehicles. Perhaps difference between passenger and freight should be removed. Could extend to "road or other public place"	Need to make sure there isn't a gap for non-passenger vehicles, non-freight vehicles (e.g. self-delivering hire car)
Possibility: Managers for operators should demonstrate that they are competent and of good repute [CQ7]	Mostly agreed but unclear how to measure. Some want current system of exams extended, others prefer self-certification or safety case. Five proposes requirement for the operator to have a (aviation or rail style) safety management system.	
HARPS operators should maintain, insure, supervise vehicles, and report collisions/other incidents [analysis 2.35]. Legal obligation to ensure roadworthiness and have facilities of maintenance. [CQ9]. Are 'users' for the purpose of insurance and roadworthiness offences [CQ10]	Agreed. Mixed views on responsibility for protection from abuse by other passengers. WMG suggested that ADSE should be responsible for insurance.	
User in charge may be outside the vehicle, but must have direct line of sight [CQ3]	Some concern that this is too wide - limit instead to parking only, or with line of sight of stopping place	
No exemptions from operator licencing for community groups [CQ5]	Just over half agree	
Possibility: Secretary of state can make exemption for trials with fee-paying passengers [CQ6]	Most agree but some concerns	
Proposal: HARPS operators have obligation to insure, supervise, report accidents, and take reasonable steps to safeguard passengers from harassment [CQ11]	Mostly agree or partly agree, but with caveats. Some propose owners or ADSE should be responsible for insurance. Remote supervision is complex. May be a need for a duty to remove defective vehicles. Some	

	suggestions that accident reporting should be extended to near misses. Safeguarding passengers is controversial.
Possibility: HARPS operator to report non-collision untoward events and contextualising data (e.g. miles travelled) [CQ12]	Much support, also opposition on grounds of commercial confidentiality. Miles travelled is easy to gather but not necessarily very meaningful.
Proposal: Legislation will set out broad duties and be supplemented by statutory guidance [CQ13]	Strong support allows flexibility. Some CertiCAV is intending to comments about a need for clarity of inform statutory expectations. guidance
No firm proposal on which regulatory body is responsible for HARPS licencing, but lots of suggestions noted [CQ15]	
No firm proposal on freight [CQ16]	

8.6

8.7 An overview of Systems Theoretic Process Analysis (STPA)

Systems Theoretic Process Analysis or STPA is a hazardous event identification method which is based on the accident causality model called STAMP (Systems Theoretic Accident Model and Processes) which in turn is grounded in systems theory and control theory (REF: STPA Handbook). Whilst traditional analysis methods believe that accidents happen due to the component failures, STPA believes that accidents are more than a chain of events and they occur not just because of component failures but most likely when external disturbances or dysfunctional interactions among system components are not adequately handled by the control system. STPA can potentially identify causes not identified by other methods, especially those concerning software bugs, human-computer interaction, missing requirements. One of the key benefits of STPA is that it allows the person performing the analysis to identify the causal factors of the hazards and their corresponding requirements that if implemented, would prevent the hazard from occurring. Therefore, it supports the identification of preventive actions for a hazard and not just its downstream mitigation (as in other methods). STPA is a four step process:

8.7.1 Step 1: Define purposes of the analysis

This step includes four parts:

1. Identify losses

- A loss involves something of value to stakeholders. Losses may include a loss of human life or human injury, property damage, environmental pollution, loss of mission, loss of reputation, loss or leak of sensitive information, or any other loss that is unacceptable to the stakeholders.

2. Identify system-level hazards

- A hazard is a system state or set of conditions that, together with a particular set of worst-case environmental conditions, will lead to a loss.
- A typical hazard contains:
- <Hazard Specification> = <System> & <Unsafe Condition> & <Link to Losses>

3. Identify system-level safety constraints

- A system-level constraint specifies system conditions or behaviours that need to be satisfied to prevent hazards (and ultimately prevent losses)
- System-level safety constraints are usually identified as the negation of the corresponding system-level hazards. A typical safety constraint contains:
- <Safety Constraint> = <System> & <Condition to Enforce> & <Link to Hazards>
- Refine hazards (optional)

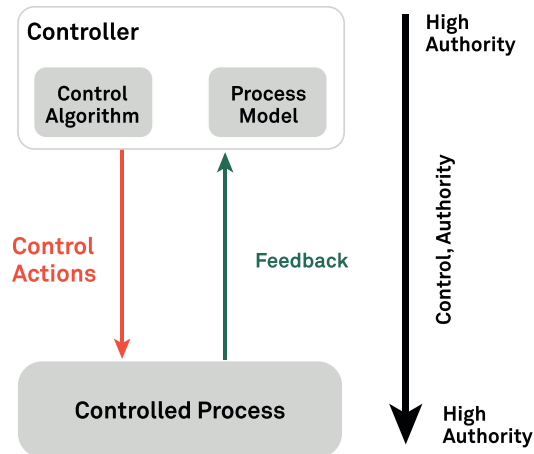
B.1.1.1.1 Identify loss scenario

- The causal factors that can lead to the unsafe control actions and to hazards

8.7.2 Step 2: Model the Control Structure

A control structure is a system model composed of nested generic control loops as illustrated in **Figure 31**. The control structure includes one or more of the following components: Control Algorithm, Process Model and Controlled Process. These components are connected by colour coded arrows with red arrows representing Control Actions and green arrows representing Feedbacks.

Figure 31: Generic Control Loop



In general, a controller may provide control actions to control some process and to enforce constraints on the behaviour of the controlled process. The control algorithm represents the controller's decision-making process, i.e. it determines the control actions to provide. Controllers also have process models that represent the controller's internal beliefs and assumptions (i.e. its view of the outer world) used to make decisions. Process models may include beliefs about the process being controlled or other relevant aspects of the system or the environment. Process models may be updated in part by feedback used to observe the controlled process.

In some situations when modelling the control structure, it is important to decide whether the signal is a control action or a feedback as either of them could change the states of the system. To clarify, we can re-define the control actions and feedbacks as following:

- **Control Action:** The signal which dictates the system/module on how to behave is a Control Action. A control action can be remembered as 'this is what you should do'
- **Feedback:** The signal which defines the state of the system/environment or the signal that is a manifestation of the external world (outside the system boundary) is a Feedback. A feedback can be remembered as 'this is what it is'

For example, a driver provides 'brake' control action to the vehicle based on the feedback it received from HMI (Human Machine Interface) regarding the 'current vehicle speed'. Here, driver is telling the vehicle that 'you should slow down' and HMI which is part of the vehicle is telling the driver that 'this is the current speed'.

8.7.3 Step 3: Identify Unsafe Control Actions

There are four ways a control action can be unsafe as illustrated in Figure 4, which are:

- Not providing the control action leads to a hazard
- Providing the control action leads to a hazard
- Providing a control action but too early, too late, or in the wrong order
- The control action lasts too long or is stopped too soon

An Unsafe Control Action (UCA) can be triggered by any of the conditions defined in Figure 4. There is a defined structure for specifying a UCA. Every UCA contains five parts:

<Source> <Type> <Control Action> <Context> <Link to Hazards>

E.g. UCA-1:

Auto- SW provides lane change cmd when a vehicle is in adjacent lane [H-1, H-2]

The '<Link to Hazards>' part acts as an index number which will be used throughout the STPA process for traceability purposes.

To prevent the UCA-1 mentioned above, we can specify a controller constraint [C-1] as follows:

C-1: Auto- SW must not provide lane change cmd when a vehicle is in adjacent lane [UCA-1]

Each controller constraint (C-1) can be traceable to an unsafe control action (UCA-1), and each UCA can be traceable to one or more hazards, which can be the cause for one or more losses. This ensures the traceability from requirements to all system artefacts which can be easily maintained. This will enhance the system's maintainability and evolution.

A sample table containing different types of UCAs are shown in the . Note that the 'linked hazards' are removed from this table as this example is only for showing different types of UCAs.

Table 10: Table demonstrating Unsafe Control Actions (UCA's)

Control Action	Not providing causes hazard	Providing causes hazard	Providing (too early)/(too late) causes hazard	(Stopped too soon)/(Provided too long)
lane change cmd	Auto- SW does not provide lane change cmd when the ego vehicle is about to collide with another vehicle ahead	Auto- SW provides lane change cmd when there is another vehicle in the adjacent lane	Auto- SW provides lane change cmd too late when the ego vehicle is about to collide with another vehicle ahead	N/A – Because the lane change cmd is an instantaneous signal. (NOTE: steering torque is the continuous signal)

8.7.4 Step 4: Identify Loss Scenario

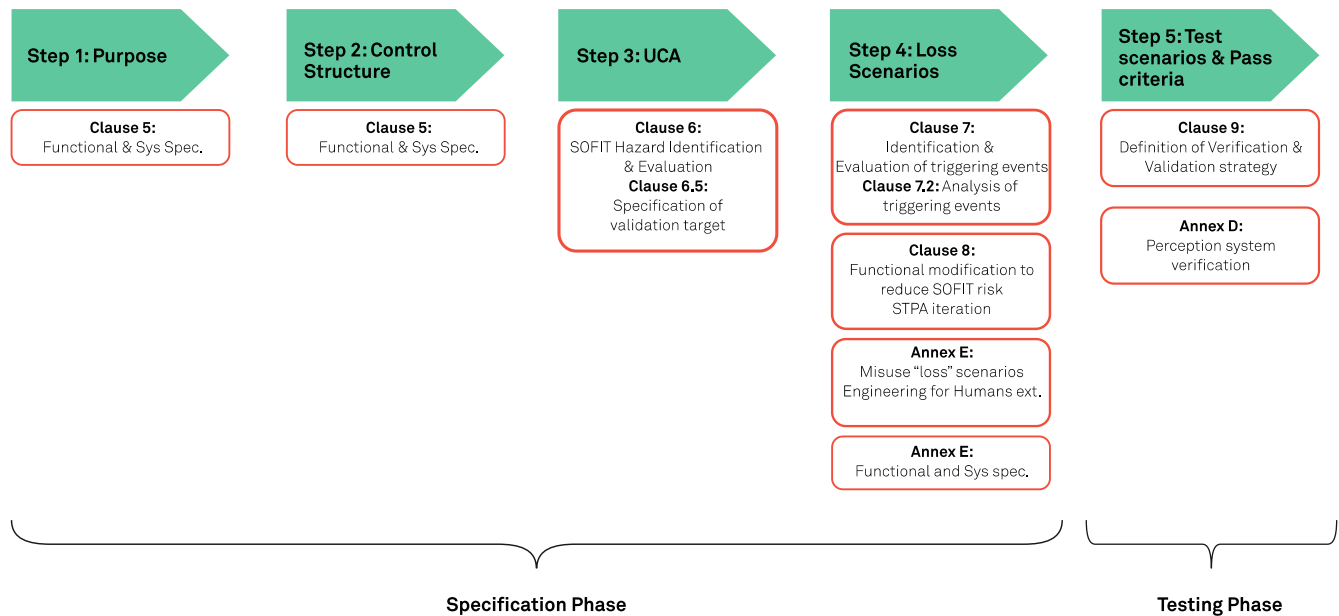
A loss scenario describes the causal factors that can lead to the unsafe control actions and to hazards.

There are two types of loss scenarios that must be considered:

- Type A Loss scenarios: Why would the UCA occur?
- Type B Loss scenarios: Why would control actions be improperly executed or not executed? (given correct control action has been commanded)

8.7.5 STPA in the SOTIF context

The figure below illustrates the mapping of STPA steps to the various clauses in SOTIF (ISO 21448). The mapping can be classified into two phases - specification phase and the testing phase. Step 1-4 of the STPA contribute to the specification phase. Step 1 which involves the defining the purpose of the analysis (losses, hazards, ODD, system), together with step 2 (which involves creation of the control structure of the system) align themselves well with SOTIF clause 5 (functional and system specification). STPA step 3 which involves identification of the Unsafe Control Actions (UCAs) which is essentially a hazard identification step, aligns with SOTIF clause 6 (SOTIF hazard identification and evaluation). STPA step 4 involving identification of loss scenarios provides alignment with various SOTIF clauses - clause 7 (identification and evaluation of triggering events), clause 8 (functional modification to reduce SOTIF risk, Annex E (misuse scenarios), Annex F (functional and system specification). The extension to traditional STPA (step 5) can be used to identify test scenarios, parameters, and their corresponding pass criteria.

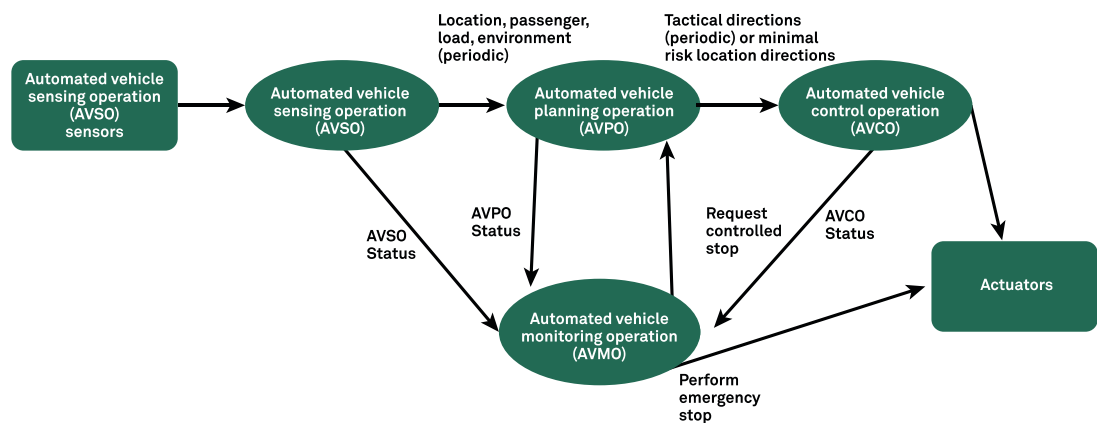


8.8 BSI PAS 1880 :2020 Guidelines for Developing and Assessing Control Systems for Automated Vehicles

This PAS provides guidelines for the high-level design of AV control systems that are intended to ensure deployments are safe, secure, and effective.

It presents a possible AV control system architecture as illustrated in **Figure 32** below

Figure 32: AV Control system schematic



Objectives are defined that developers of AVs would need to meet to comply with the PAS. Further clauses provide guidance on how these objectives are to be met. These objectives and guidelines relate to the:

- Design of the AVSO, AVCO, AVPO and AVMO functions shown in the AV control system schematic
- Definition of the 'Mission' which includes characteristics of what is being carried, type of routing (static/dynamic), and configuration limits
- Definition and scope of the ODD
- Demonstration of safety, security, and effectiveness of the AV
- That evidence should be provided that the defined objectives are met

Annexes provide informative discussion of:

- Examples of evidencing compliance with the objectives
- Possible ways of implementing the AV control system functions by means of automated vehicle units
- Use of Artificial Intelligence techniques in AVs
- an illustrative relationship of the use of AVs in an earthmoving/mining context and some of the related standards that apply or are under development
- A list of standards that might be of interest to readers of the PAS

The CAP is aligned with the requirements defined in this PAS where practicable, however, terminologies applied differ for the purpose of clarity and aligning with wider set of applicable standards.

8.9 BSI PAS 1881: Assuring the safety of automated vehicle trials and testing – Specification

This PAS specifies requirements for operational safety cases for automated vehicle trials and development testing in the UK to demonstrate that trialling and testing activities can be undertaken safely and securely.

It describes that 'Safety assurance for automated vehicles can be categorized into two interdependent areas: system safety and operational safety. System safety is achieved through ensuring adequate functional safety, safety of the intended functionality (SOTIF) and cybersecurity.'

The focus of this PAS being on operational safety and 'does not cover the system safety of the vehicle (e.g. functional safety, safety of the intended functionality (SOTIF) and cybersecurity assessments) but does rely on their outputs.'

It describes a 'Safety case':

- In the Introduction as: a structured argument supported by a body of evidence that demonstrates that the safety risks have been identified, managed, and reduced as low as reasonably practicable (ALARP)
- In the terminology as: a structured argument, supported by evidence, intended to justify that a system and activity is acceptably safe for a specific application in a specific operating environment

Clause 4 defines compliance requirements for safety case:

- to 'document and maintain a safety case for automated vehicle trials and testing to ensure and demonstrate that risks to all affected parties, including vulnerable road users, are assessed and reduced ALARP to an acceptable level throughout the lifecycle of the testing and trials.'
- to complement safety processes and procedures implemented by testbeds
- be owned by the trialling organization.

Clause 5 defines compliance requirements for the contents of the safety case. These relate to information that shall be documented in a safety case and what it shall demonstrate. It does not define practice for developing and structuring an argument that justifies the achievement of defined safety specifications or needs.

8.10 Draft PAS 1882:20 20

This PAS specifies requirements for data collection, curation, and storage during automated vehicle trials and advanced trials in the UK in relation to the information collected or received by the system.

It describes an initial set of information that may be required to conduct an incident investigation in the case of an incident involving an automated vehicle during its trial operation. It builds on the requirements of the PAS 1881 that focusses on safety case requirements, by providing information requirements for incident investigation.

The core of this PAS focusses on the creation of an Information Management Plan (IMP), which sets out the information management activities that are to be in place for incident information for a given trial. Furthermore, it requires a trial operator to have an incident response plan, which illustrate the policies and procedures for undertaking an internal incident investigation which may be of three types: 1) criminal investigation, 2) civil investigation, and 3) safety fact-finding investigations.

The IMP developed by the Trialling Organisation (TO), specifies the relevant data to be collected and its management for an automated vehicle trial. An IMP is required to have a control loop model with control actions and feedback between control loop entities. While near miss events are important events to be captured, there is a lack of consensus on what constitutes a near miss. Therefore, the IMP is required to include the definition of the near miss by the trial operator for the purpose of the trial. As part of the IMP, this PAS requires data to be recorded 30 seconds before and after the triggering of the incident.

The type of data that is required to be collected, curated, and stored is split into the following sub-categories:

- ODD information (includes ODD status, ODD attributes, and risk status)
- Vehicle communications and ADS specific information
- Vehicle dynamic and operational information
- Vehicle level of risk as perceived by ADS
- Location data

8.11 Performance Testing for Sensors in Connected and Autonomous Vehicles: Feasibility Study

This is a report commissioned by CCAV and produced by CPC and National Physical Laboratory to investigate the requirements, infrastructure and facilities needed to enable reliable CAV environmental sensor testing and validation, which can support a safety evaluation and assurance programme.

The report identifies three barriers to the accurate and comprehensive testing of CAV sensors:

- Testing sensor performance in different weather conditions is recognised as the largest gap in current testing facilities. In order to test performance in different weather scenarios and combinations there is a requirement for clear, standardised definitions for all weather types. Presently these definitions are missing with the exception of some Met Office definitions for rain
 - As a result of this simulation environments do not have defined characteristics for different weather types, meaning that they are unable to accurately test the performance of sensors for autonomous driving systems
 - Understandably manufacturers are protective of their IP. However, a reluctance to share sensor data makes it impossible for the sensors to be accurately modelled in simulation environments and for sensor performance to be validated
 - There are no standardised metrics for measuring sensor performance within an autonomous/ADS context, this is impeding the creation of accurate modelling environments and therefore maintaining a significant risk for vehicle OEMs, autonomous driving systems developers, regulators, and the public
- Without standardised metrics it is hard for ADS developers to self-validate their system's reliability; and it also stifles earlier stage investment in emerging companies to this sector

Other observations made in the report are that:

- There is currently a lack of availability of sensor models sufficient for full fidelity simulation testing.
- There is demand for a common approach on the characterisation of different conditions affecting sensors: in particular, it would be useful to have definitions of weather types, and how that affects sensor performance.
- Interference between multiple vehicles equipped with active sensor systems operating in close proximity is an open challenge, as is the possibility of malicious interferers aiming to blind or deceive the sensor systems
- Current CAV perception sensors testing facilities are unable to provide complete testing scenarios required for the automotive industry. To adequately test the performance of sensor systems, through their development and integration cycle, there needs to be a diverse ecosystem of testing facilities; from bench top and compact chambers for development, calibration, and validation; to large centres for functional verification when integrated. There is a particular weakness in the ability to emulate weather conditions.

8.11.1 Conclusions

Whilst it is established that there is a lack of testing infrastructure for sensors, the specification of such facilities is dependent on how industry specifies, validates, and manages the safety integrity of sensors. This indicates a priority to develop the technical framework, including characterising sensor

performance, sensor models for simulation, validation testing and supply chain management before developing physical testing infrastructure.

8.12 VeriCAV

VeriCAV is a 2019-20 CAVSIM project develop an integrated platform to allow Automated Driving Systems (ADS) to be tested in simulation. It is aimed towards the internal verification and validation process undertaken by system developers. Recognising the need to test AD systems against a very large quantity of scenarios, the VeriCAV project aims to address the challenges of efficient and effective testing of ADS in simulation.

However, there are common areas of interest with CertiCAV. They've developed behaviour quality models that may be adapted for application to independent tests and metrics of which could relate to regulatory requirements.

VeriCAV advanced the state of technology in three key areas: firstly, by creating "smart actors" to control the road users in a simulation, leveraging AI and cognitive research. Smart actors not only ensure the system-under-test is exposed to realistic behaviour from other road users, but also mean many random variations of a core test can be executed. The second innovation was to demonstrate how heterogeneous components can be integrated into a single platform, building on industry standard protocols. Finally, VeriCAV developed methods to automatically generate and select the test scenarios, to automatically score the performance of the system-under-test in scenarios via a "test oracle", and to use these performance scores to optimise the exploration of the test space. These remove the reliance on human expertise to set up and analyse tests, making it possible to easily and inexpensively run orders of magnitude more tests, and also ensure consistency and repeatability of performance assessment. A summary of the process is shown in Figure 33.

Figure 33: Illustration core test execution process within the VeriCAV project



8.13 dRISK

dRISK is a consortium that claims de-risking autonomy by training and testing AVs on way more edge cases, for dramatically improved smoothness and safety. They have received a £3million grant from the CAVSIM programme to undertake project work through to March 2022.

According to information on their website <https://drisk.ai/> they are integrating a massively diverse set of data and expert curation into the ultimate data resource for AVs. They claim to have invented a way to "map" all edge cases into a comprehensive taxonomy - a knowledge graph of risk - which acts as both an edge case database and evaluation space.

They claim this patented technology trains, tests and validates AVs on comprehensive sequences of edge cases derived from their massive knowledge graph. The result is a step increase in AV responsiveness and safety. They believe that along with partners, they will develop this technology into the first true driver's test for AVs.

We convened a meeting with members from dRISK to discuss what influence or role their product/project may have in the future of HAV assurance. They were not able to share any art at that time to illustrate their work. A subsequent search however revealed their patent

<https://patentimages.storage.googleapis.com/69/2e/26/1672ec9b5e6286/US20200081445A1.pdf>

This sets out a methodology for deriving edge case scenario for AV across a comprehensive input space that is optimised to create a complete and challenging set of training data by application of graph theory implemented with neural networks.

8.14 PD ISO/PAS21448:20 19, Road Vehicles – Safety of the intended functionality

This standard provides guidance for the design, verification and validation measures needed to achieve the Safety Of The Intended Functionality (SOTIF). Where SOTIF is defined as the absence of unreasonable risk due to hazards resulting from functional insufficiencies of the intended functionality or by reasonably foreseeable misuse by persons.

Safety relevant topics covered by SOTIF include:

- Performance limitations or insufficient situational awareness
- And in conjunction with ISO 26262:
 - Reasonably foreseeable misuse such as incorrect HMI
 - Impact from car surroundings such as other users, physical and digital infrastructure, sensory conditions

Therefore, SOTIF focuses on safety associated with performance of perception. The definition of scenario used in SOTIF a "description of the temporal development between several scenes in a sequence of scenes" is centric to the perception of scenes. However, several of the objectives, principles and methodologies defined in SOTIF may be useful for application in a wider context.

The normative content of SOTIF sets out objectives to be applied to relevant product development processes. There is further content setting out practice for achieving the objective, however this content is informative. As such, compliance to the SOTIF standard requires arguments to be provided that all objectives have been achieved but does not place requirements the arguments themselves.

- Functional and system specification
- Development of the use case for the system comprising of relevant scenarios, identification of triggering events within scenarios that may lead to harm
- Identification and Evaluation of hazards caused by the intended functionality
- Specification of acceptable validation targets
- Identification and evaluation of hazardous events
- Identification and Evaluation of triggering events
- Determining whether specific conditions that initiate potentially hazardous reactions are acceptable
- Functional modifications to reduce SOTIF related risks
- Development and evaluation of measures to control SOTIF related risks
- Definition of the verification and validation strategy
- Determination of argument for SOTIF, evidence required for supporting it, and the procedure for evidence generation
- Verification of the SOTIF
- Determination that the system behaves as expected for known scenarios of hazards and misuse
- Validation of the SOTIF
- Determination that the system does not cause an unreasonable level of risk in real life use-cases including unknown unsafe scenarios
- Methodology and criteria for SOTIF release
- Review activities and consider findings of SOTIF activities to evaluate acceptability of risks

This process and objectives are directly applicable to the assurance of the safety of an ADSs perception.

The stated objectives, applied to the context of CertiCAV, cross reference to requirements in the specification of various aspects of the deployment specification, including the ODD and ADS functionality, to define the use-case and scene characteristics and relationships.

Furthermore, in the context of CertiCAV they cross reference requirements to further informative practice is defined as a reference for developing arguments for the objectives.

However, this version of SOTIF is intended to be applied to systems automated to Level 0-2, where a human driver is responsible for the sustained conduct of the driving tasks and automated driving functions may temporarily take over driving control to enhance safety or performance. SOTIF therefore is intended to provide practice that ensures that the implementation of automated intended functionalities does not introduce unreasonable safety risk. It is not intended to provide practice that ensures the implementation of intended functionalities is sufficient for providing safety for the sustained conduct of the dynamic driving task without a human driver being responsible and available.

Further consideration is therefore required in the application of SOTIF to the scope of the CertiCAV framework:

- Ensuring implementation of functionality is sufficient for safe performance of the sustained dynamic driving task
- Automated system performance of fallback operation / unavailability of human driver
- Full automated responsibility for conduct of driving task, i.e. unknown unsafe scenarios and both false positive and false negative actions are unacceptable

It should be noted that SOTIF was developed for, and is applied to, a wider system where the top level intended behaviour of driving functions is externally explicitly defined by other standards or technical regulations. For example, in the case Adaptive cruise control system the intended behaviour of the driving function is defined in ISO 15622:2010. For CertiCAV use-case the intended behaviour implemented by the ADS would be too complex and interdependent on the individual deployment specification to be explicitly defined in a common standard and further process and methodologies are required to manage this.

However, the following practice defined in SOTIF is considered relevant for wider application to assurance:

- SOTIF presents a classification of driving scenarios into four areas, which allows a visualisation of SOTIF-related hazardous event model. This system may be useful for application to wider scope scenario discussion in CertiCAV. However, further consideration would be required for management of unsafe scenarios, and the potential interaction of residual risks with residual risks from other areas of safety: The four areas are:
 - Known safe scenarios
 - Known unsafe scenarios
 - Unknown unsafe scenarios
 - Unknown safe scenarios
- Several of the process and methodologies relating to the strategy and implementation of Verification and Validation (Clause 6-12) may be applicable to the general safety assurance and Quality Management of an ADS.
- Clause 6 covers the identification and evaluation of hazards caused by the intended functionality. This is however, largely limited to defining that a systematic Hazard Identification and Risk evaluations to determine that a validation target is met. The topic of developing and verifying an intended functionality for a driving behaviour function is not discussed.
- Of particular relevant Clause 6.5 discusses the specification of validation targets relating them to regulations, statistics, and functional requirements. It is suggested that a target could be an accident rate based on statistics, or follow rationales of: 'globally at least as good', or 'as low as reasonably practical'. Annex B provides an example for determining the validation target and strategy for an AEB system which involves the analysis of traffic statistics and definition of test scenarios.

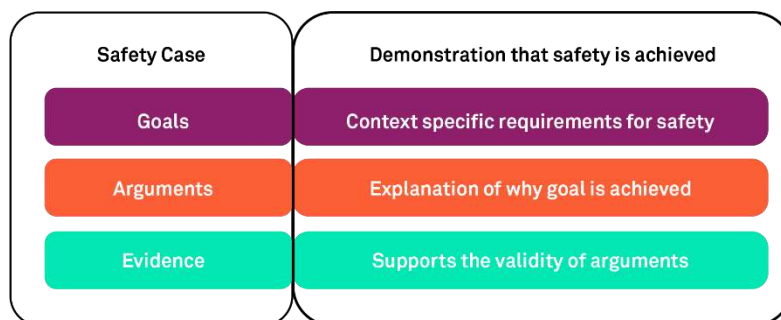
8.15 UL 4600

UL 4600 is a standard covering safety principles, risk mitigation, tools, techniques, and lifecycle processes for building and evaluating a safety argument for vehicles that can operate in an autonomous mode. It does not intend to provide a development process or methodologies, rather it references other processes where appropriate. Its annex informatively references ISO 26262 and ISO/PAS21448

It has similar assumptions on vehicle automation to CertiCAV. Operation is assumed to occur without human supervision and without expectation of human intervention in performing and supervising the dynamic driving task and other normal system operations based upon the current item state and ability to sense and otherwise interpret the operating environment. Human contributions to safety outside of normal operation are considered (e.g., maintenance), as are interactions with humans who are not operating the item (e.g., pedestrians).

Its approach is to require a goal-based safety case that in effect encompasses a complete safety assurance argumentation. The safety case is structured into claims, arguments, and evidence that the vehicle plus other support contributing to safety is acceptably safe. This outline structure is shown in **Figure 34** below.

Figure 34: UL 4600 safety case structure



The standard defines requirements and guidelines for formalising the style and formatting of the safety case in terms of syntax, semantics and relating to development, consistency, and completeness of argumentation.

This includes involving Safety Performance Indicators (SPIs) in the safety case. These are categorised in terms of their role in the safety case of.

- Safety metrics
- Risk mitigation effectiveness
- Interactions with subsystems, ODD, and the environment
- Fault and failure recovery
- Safety culture

This contrasts usefully with CertiCAV's discussion on different types of Performance Indicators. For example, UL 4600 conformance requires safety metrics to include:

1. Incident rates
2. Rule violation rates
3. Minor collision rates
4. Minor injury rates
5. Major injury rates
6. Fatality rates

CertiCAV discusses the difference between Performance Indicator types of event occurrence and informal specification; and therefore, the need for specifications relating to traffic rules may need to be formalised to facilitate detection of rule violation rates.

The sets out requirements and guidelines for safety case coverage in the following categories

- Interaction with humans and Road Users
- Autonomy functions and support
- Software and system engineering processes
- Dependability
- Data and networking
- Verification validation and test
- Tool qualifications
- Lifecycle
- Maintenance

The autonomy functions and support section cover ODD, sensing, perception, machine learning, planning/prediction/trajectory, and actuation/timing performance

There is a section setting out conformance assessment and monitoring methodology.

8.16 TR68: Singapore's technical Reference for AVs

TR68 is Singapore's technical reference for Autonomous Vehicles (their equivalent of a PAS). Version 1 was published January 2019⁵⁶; and their intention is to develop and publish a version 2 in the future. Their aim is to develop these into Singapore Standards over several years; however, they also recognise the need to align with international standards as and when they're developed.

TR68 is divided into four parts that are outlined below in terms of relevance to CertiCAV.

Part 1, Basic Behaviour:

- The principles for applying driving code of practice to ADS performance requirements
- Specified some relevant behaviour requirements, for example achieving a safe state, and lateral clearance distances
- The principles for managing contraventions to driving rules. How ADSE may argue justification case for programming driving policies that allow for contravention of rules in certain situations, and principles to follow including minimal violation, data, and continual improvement model
- However, it should be noted that various requirements are not rigorous. For example, saying that the vehicle should read signs, and various 'rules' that do not have full specifications such as "An AV shall only perform an overtaking manoeuvre when it is safe to do so" but what is 'safe'?

Part 2, Safety:

- Principles and objectives for safety management
- Safety and quality management processes
- Safety argumentation
- Appropriate safety goals

Part 3, Cybersecurity:

- The content is not relevant to CertiCAV

Part 4, Vehicular data types and formats:

- Defines the wider eco-system for AV stakeholders - worth reviewing as a verification exercise
- Provides some example use case for data exchange. However, these use cases are not directly relevant to CertiCAV
- The significant portion of content is the definition of data structures. If working in this area SAE/ISO standards for data structures should also be considered

⁵⁶ (CETTRAN, n.d.)

8.17 RAND – Approaches to assessing acceptable safety for automated vehicles

This report by the RAND corporation splits the challenge of assessing safety into three parts:

- Safety as a measurement – how data can be used to understand the level of safety offered by a system
- Safety as a process – the methods, culture and actions which can help to ensure safe outcomes
- Safety as a threshold – addressing the question of "how safe is safe enough"

There is also a discussion of how best to communicate with the public about safety.

8.17.1 Safety as a measurement

RAND classifies measures of safety as either "leading" or "lagging". Lagging indicators strictly relate to crashes, so only provide results after harm has already occurred. Leading indicators are pre-crash measures which can be used as a proxy for harmful outcomes. The report notes that some lagging indicators can also be used as a proxy for more serious ones. For example, the occurrence of damage-only crashes can indicate that future injury crashes are likely.

Lagging indicators are by design based on events which occur rarely. This means that statistically valid judgements cannot be made until an ADS has been deployed at scale for some time (and, if it is a poorly performing ADS, a substantial number of people have been harmed). Therefore, industry is currently focused on leading indicators. Three types of indicator are considered:

- Disengagements. These are difficult to use to make meaningful judgements, as their occurrence is highly dependent on things such as the type of testing, company policies, and safety driver training. The definition of a disengagement differs between developers, further complicating comparisons
- Legal infractions. RAND reported that there was little enthusiasm for using these as a leading measure within the industry, as the safest course of action sometimes involves legal infractions and the relationship with collisions is unclear (though we note that this also applies to some proposed 'roadmanship' metrics, such as harsh braking). The report also highlights conflicts between what the law says should happen and the expectations of other road users. RAND's authors are based in the USA, and it is interesting to note that the examples selected for this point (e.g. frustrating other road users by stopping at all stop signs) do not translate perfectly to UK driver expectations
- A broad category of indicator called "Roadmanship" which is the current area of focus for most developers. This attempts to measure good driving, rather than simply the absence of crashes. Many indicators can be used to measure roadmanship, some of which are listed in the report, but overall, the measure lacks definition

CertiCAV is focused on assurance of HAVs pre-deployment. This means that the distinction between leading and lagging measures is less helpful – for example, it is not clear whether a high-speed collision in simulation should be considered a leading or lagging indicator. However, the principle behind the approach is highly relevant: in a realistic amount of statistically representative testing, the most harmful events will not be observed often enough to assure safety. This means that pre-deployment testing will need to make inferences about serious outcomes based on less serious events.

There are a few other ideas in this section of the report which are highly relevant to CertiCAV:

- Leading measures are not enough to provide certainty about lagging measures on their own. While leading indicators are likely to correlate with lagging ones, the relationship is not set in stone. Certainty about safety can only be achieved using lagging indicators. This highlights that pre-deployment testing cannot be guaranteed to detect all unsafe systems.
- Most stakeholders interviewed felt that the industry now has enough common knowledge to standardise some leading measures. However, these may need to be supplemented with application or ODD specific measures
- There is a suggestion that it will not be acceptable to expose one group of road users to an excessive risk, even if the risk overall created by a deployment is acceptable
- Lagging indicators will need to be validated over a longer period, which means that software updates are likely during the validation period. A previous RAND report is referenced suggesting that a moving average measure should be used, though this should still be reset if the ODD or measurement techniques change. One possible challenge is defining what is meant by an ODD change: as with the ADS, minor updates are constantly made to road networks

8.17.2 Safety as a process

A key theme of this section is the need to gain public trust. It is suggested that this will not be achieved simply by making statistical comparisons (which, in any case, may not be possible to demonstrate robustly) but by showing a genuine effort to make an ADS as safe as possible.

The report suggests the following key points relating to safe processes:

- Standards have a role to play, but the exact role is not yet clear. As is expected at this point in development, many standardisation projects are in progress. Key questions remain around what should be standardised and whether the resulting standards should be mandated by regulation
- It is suggested that standards are generally most effective once developers in different organisations are starting to converge on similar solutions
- Some interviewees suggested standardising the set of assumptions which may be made about other road users
- Many in the industry feel that regulatory standards (at least in the USA) react too slowly for the current state of the industry and risk holding back progress. Voluntary standards were perceived more favourably. Interviewees with a regulatory perspective felt that it was the role of the government to set a minimum standard for safety which should be exceeded where good practice is followed
- It is not realistic to expect all ADS software to be error free. What is important is that errors are identified and fixed quickly
- Safety management systems are seen as an important tool. These can include features such as an independent safety department, safety training for staff, and use of data to drive decisions. There is broad agreement that safety starts at the top and requires engagement from company CEOs
- One interviewee suggested requiring companies to publish a safety philosophy. Companies which made impossible or misleading claims would have provided evidence that they should not be trusted

- In the aviation and rail industries there is a culture of sharing information about what went wrong, so that the whole industry can learn from it. This is not yet true in the automated vehicle domain. RAND suggest that the comparatively small number of industry players is a factor in this. From a CertiCAV perspective, we think it is important to note that the culture of openness in other industries is neither automatic nor universal. Regulators have a key role in creating and enforcing such a culture

Many of the points in this section can be seen as a way of demonstrating an effort to minimise risks to a level which is "as low as reasonably practicable". Minimising risks in this way has historically been a key part of UK health and safety regulation.

8.17.3 Safety as a threshold

This section addresses the question of how a threshold for acceptably safe should be defined. It considers several types of definition, including a discussion of how to relate definitions to the potential of the technology and goals such as Vision Zero. However, the discussion of how comparisons should be made to a human driver is especially relevant. Key findings include:

- It is natural to make statistical comparisons based on whole-population metrics, but these do not match how people experience risk. A minority of drivers engage in high risk behaviours, such as driving at excessive speed or whilst drunk. For someone who does not do these things, getting into a vehicle which is as safe as the average driver represents an increase in risk. Options for improving this measure include:
 - Comparing the ADS to a specific, relatively low risk driver profile
 - Setting a target which requires a percentage improvement on the human reference
 - Basing comparisons on accident rates excluding certain risky behaviours
- Baseline data on the safety of human drivers is usually only available for lagging indicators. As discussed earlier, making comparisons using these metrics alone is problematic for pre-deployment testing. Naturalistic driving studies are starting to change this, but for a meaningful comparison the data must relate to the ODD of the automated vehicle. This is because some areas are inherently more dangerous to drive in than others

There is also some discussion around the practicalities of 'driving tests' for automated vehicles. These could have a role in assessing an ADS against defined competencies. However, several limitations of this approach are noted:

- Automated vehicles are a 'black box' to the test (the examiner cannot see their internal workings). In contrast, examiners make judgements of a human's competence based on the process which they follow (e.g. checking mirrors) as well as the outcome (e.g. not coming into conflict with another vehicle)
- Newly qualified human drivers are still relatively high risk. When a human passes the test, there is an expectation that they will improve with experience, which does not translate well to automated driving
- There is no proof that displaying behavioural competencies on tests will result in safe outcomes in the real world

While there are drawbacks, this form of testing is widely used by manufacturers and is likely to play some role in safety assurance.

8.17.4**Conclusions**

- Where a statistical judgement is required, it cannot be made solely based on serious or fatal accident occurrence. Other indicators need to be used as a safety proxy
- As a result of this, it cannot be known for sure whether a vehicle will meet statistical safety targets for some time after deployment
- Statistical evidence will not be enough to win public trust on its own. The industry will also need to convince people that it takes safety seriously and is minimising the risk associated with automated vehicles. Following appropriate safety processes can help with this
- Standards and safety management systems both have a clear role to play. There needs to be an emphasis on finding and fixing problems effectively, rather than expecting them never to occur.
- There is a clear role for a standard relating to the assumptions which can be made about other vehicles
- Strictly following driving rules is sometimes in conflict with the expectations of other road users
- Sharing information about adverse events can help lead to industry wide safety improvements, but there is currently a reluctance amongst developers to do this
- Comparisons to human drivers are challenging, both practically (collecting appropriate data) and philosophically (picking an appropriate driver or group of drivers to compare against)
- Scenario-based 'black box' testing does not solve the problem of assessing automated vehicles on its own

C Example HASS specification

8.18 Introduction

This section shows an outline example of a HASS specification for a HAV parking valet system. This example does not attempt to apply the CertiCAV framework completely, e.g. with the full set of attributes identified in Section 2.4. Nor does this example attempt to cover the level of detail that a real HAV system developer would apply. The purpose of this section is to provide a general outline of what a HASS specification may look like for discussion and development purposes.

8.19 Parking Valet, example HASS specification

This sets out a notional HASS specification for the use case of a HAV parking valet system.

For this system a HAV manufactured by an OEM, is owned by an individual who uses an automated car parking valet service provided by a car park operator. In this case the car park operator would therefore be acting as the LFO even though they do not own the HAV.

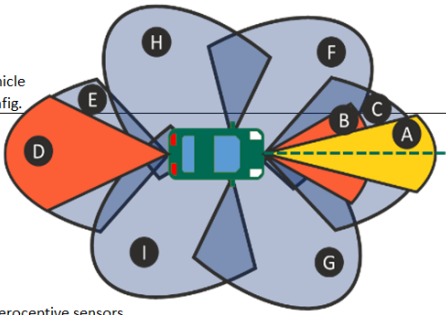
Cat.	Attributes	Definitions
Vehicle Operational Characteristics	Outline Physical attributes	Category M passenger car e.g 4080mm len, 1560mm wi, 1640kg.
	Chassis dynamics	8m turning circle, 110kW power unit, 9m/s/s maximum braking
	Vehicle use	5 seats, one for driver. No occupants on board when in automated mode.
	Deployment operation	Storage and retrieval of unoccupied car in multi story car park
	Vehicle config.	Vehicle must not have any trailer attachment except specified roof rack with no cargo. Cannot have any active error/service codes Must have original wheels and specified tyres. Front windscreen must not have tint allowing less than 70% transmittance Side windows must not have tint allowing less than 50% light transmittance
	Exteroceptive sensors	 + A: Lidar, azi 0deg, +/-15deg, 60m 0 B: Radar, +/-25deg, 40m - C: Camera, +/-40deg, 55m - D: Radar, azi 180deg, +/-20deg, 55m - E: Camera, azi +50deg, +/-40deg, 55m - F: Camera, azi +120deg, +/-40deg, 55m - G: Camera, azi -50deg, +/-40deg, 55m - H: Camera, azi -120deg, +/-40deg, 55m - I: Camera, azi 180deg, +/-40deg, 55m
	Cabin sensors	Internal infrared occupancy sensors
ODD		Concrete building structure, localisation waypoint codes painted on corners, pillars and walls Asphalt carriageway Concrete kerbs Bollards with reflective strips Ramps at 8% grade Linemarkings
	Scenery	Artificial lighting
	Sensory conditions	
	Dynamic elements	Stationary vehicles only

Figure 35: Example HASS specification, HAV parking valet, part 1

Cat.	Attributes	Definitions
ADS functionality	Object and Event detection specification, capabilities and limitations.	Detects and classifies: parked vehicles, localisation waypoints, roadmarkings, ramp transition points. Detects the presence of any object larger than 10x10x10cm in the driveable area that is not part of the scenery map. Does not allow entry into automated mode if cabin sensors detect occupant. Detects performance loss of ADS systems / HAV actuators or unexpected limit of ODD which triggers MRM.
	Localisation specification, capabilities, and limitations.	Localises to global precision of 10cm using a combination of inertia/dead reckoning and computer vision assisted by localisation waypoints
	Data communications: role of wireless data in functionality of ADS	Wi-Fi is used for initialisation check and strategic control.
	Tactical Behaviour / Manoeuvres	Follows fixed path in lane Parking manoeuvre into empty parking bay Parking manoeuvre in pick up area Controlled stop if any obstacle or moving object detected
	Fallback / MRM	Turn hazard lights on and controlled stop.
	Longitudinal motion control	Maximum forward/reverse speed 3.6m/s, maximum service acceleration 3m/s/s. emergency deceleration limit 9m/s/s Anti-lock braking
	Lateral motion control	Yaw rate limit 5 deg/s electronic stability control
	Non-motion vehicle control	In automated mode headlights are on throughout. Activates hazard lights if fault condition detected. Sounds horn if possible collision event detected. Turns off ignition/lights on successful exit of automated mode Turns on ignition/lights on entry of automated mode. Unlocks doors under certain fault conditions only
HASS Architecture	Define the architecture of the HASS, including off-board or nomadic electronic systems that are not integrated into the HAV but may influence the ADS's performance of DDT and therefore HAV behaviour through data communications. For example:	
	•Vehicle/Fleet management systems; including vehicle routing, and monitoring. Certain routing restrictions may be in place to manage safety and capability limitations of ADS functionality. •Global Navigation Satellite Systems (GNSS) or other systems used to help vehicle localisation •Communications equipment/services such as C-ITS •Tele operation systems •Mapping / databases	
Deployment management	Entities involved in the deployment and their roles, how the vehicle is operated and managed.	ADSE: OEM manufacturer LFO: Car Park operator UIC/User: Vehicle driver
	How the vehicle is operated and managed	a - Parking 1. The vehicle driver parks the vehicle in a designated bay, ensures the car is empty of occupants and compliant with other requirements and then locks the vehicle. 2. Vehicle driver requests vehicle to be parked through mobile app. 3. LFA's RSU communicates with vehicle for initialisation check (no fault condition, software versions, servicing up to date, sufficient battery level) 4. If initialisation check is successful, and path is clear the RSU communicates strategic mission to vehicle, path and parking destination 5. Vehicle turns on ignition and executes the mission
	Change control management procedures	Agreements in place between ADSE/LFO on management of any physical changes and release of new software version. Vehicle must have latest releases to activate automated mode.
	Safety monitoring, incident detection	Roadside optical/infrared camera to check vehicle is not occupied.
	Through life management	Inspection schedule for car park with maintenance/intervention. Vehicle inspected as part of service schedule.
	Operational conditions assessment	Environment is well controlled, any activation of fallback investigated for potential ODD exit.

Figure 36: Example HASS specification, parking valet, part 2

8.20 HASS specification relationship to system safety case and system assessment

The specification of the HASS has a direct influence on the formation of the system safety case, and subsequently on system approval activities.

For instance, in this parking valet example the HAV driving behaviour does not to interact with any dynamic actors. Classification of objects by ADS perception is limited to static objects that affect path planning, including: Parked vehicles, localisation waypoints, road markings, and ramp transition points. All other object and event detection is limited to detection of potential obstacles larger than 10x10x10cm, performance loss, or out of ODD bounds and triggering MRM which is in turn limited to a controlled stop. This reduces a significant part of the system safety case to demonstrating that the HAV detects obstacles and responds by braking irrespective of the type of object or other situational factors. The system approval activities are consequently constrained due to the relative simplicity of the safety case requiring less evidence and a relatively small number of scenarios required to cover the range of the systems behaviour and operational conditions.

Conversely, a HASS operating at higher speeds and with more complex driving behaviour, would need to demonstrate reliable classification and behavioural response to dynamic objects, and appropriate control of MRM. This makes the structure and argumentation of the system safety case substantially more complex, for example needing to validate reliable and appropriate dynamic object classification, and appropriate behavioural response to the large number of potential scenarios. This in turn expands the system approval activities, having to cover a more complex safety case, validation toolchain and more expansive test coverage.

Thomas Webster

Thomas.Webster@cp.catapult.org.uk

Siddartha Khastgir, Robert Myers, and Zeyn Saigol

Visit our website

cp.catapult.org.uk

Follow us on Twitter

@CPCatapult

Email us

info@cp.catapult.org.uk