# Safe Drive Initiative: SafeDI scenario-based AV policy framework – technical implementation guidance

WHITE PAPER

NOVEMBER 2020

**Cover:** Getty Images/Kinwun

**Inside:** Getty Images/4X-image

# Contents


3 1 Introduction

4 1.1 About this guide

4 1.2 Framework overview

5 2 Technical implementation guidance

5 2.1 Establishing a centre of excellence (CoE)

6 2.2 Determining ODD taxonomy

6 2.3 Defining the deployment ODD

7 2.4 Identifying interim milestones

9 2.5 Choosing scenarios for the interim milestones

11 2.6 Generating qualitative scenario library

12 2.7 Parameterizing scenarios

13 2.8 Prioritizing scenarios

13 2.9 Selecting tests for each milestone and scenario

15 2.10 Allocating tests across modalities

17 2.11 Determining success/failure/advancement criteria

18 2.12 Verifying the AV backup system

19 3 Reference terminology and definitions

19 3.1 General terms

19 3.2 Scenarios and related terms

20 3.3 Defining verification methods

21 Contributors

22 Endnotes

# 1 Introduction

# 1.1 About this guide

This paper is part of a series from the Safe Drive Initiative, proposing a high-level policy framework to enable a regulator to implement an operational safety assessment that allows AV companies to operate without a safety driver. The SafeDI framework synthesizes the core technical knowledge required to understand the validation process and allows regulators to customize it to their jurisdiction by evaluating the key steps, design choices and policy levers.

This technical implementation guide provides supporting information to policy-makers to help them prepare their own policy framework by summarizing the most important supporting material, reference standards and other knowledge required to create a scenario-based AV safety assessment. It is not intended to be exhaustive, but provides a range of recommended knowledge to assist policy-makers in creating their own safety assurance process.

The document is intended to be a living resource and will be updated periodically to reflect the latest standards and additional research as appropriate.

# 1.2 Framework overview

In the prior publication, *SafeDI Scenario-Based AV Policy Framework: An overview for policy-makers*, the World Economic Forum proposed a process that provides a regulator with the necessary knowledge to implement an operational safety assessment of an AV.

TABLE 1 **Framework implementation process**

| Prepare | Define | Measure | Execute |
|---|---|---|---|
| Create a centre of excellence (CoE)<br><br>Convene necessary stakeholders, define the end goal and fill out the details of the process | Define the test ODD<br><br>Establish the required behavioural competencies for the AV, define the geographic areas and parameters for each interim milestone | Specify on-road, controlled environment and simulation tests, and determine success/advancement criteria | Conduct tests and collect data from AV providers as necessary, improving the safety assurance process as needed |

The first step in the process is to appoint a centre of excellence (CoE) with the goal of convening the necessary stakeholders, including local, regional and national regulators, academic researchers, industry representatives and municipal leaders. The members of this group will jointly establish the desired end state (i.e. define the deployment operational design domain [ODD]), and are to be consulted in the development of the required competencies for AV systems – both how AVs should behave and how AV developers need to demonstrate safe performance.

Via the CoE, regulators and AV providers will then work together to define a series of interim milestones, establish a set of qualitative scenarios and create agreed-upon tests of the AV's competency. These tests first open up small areas for on-road testing, slowly increasing the test ODD to match the desired deployment ODD when the AV's competency to operate safely in the defined scenario library has been sufficiently proven. The CoE should also determine metrics for the AV system's successful performance and criteria for advancing the system to the next milestone.

Once the process is established, the CoE should oversee tests and collect data from ongoing development to continuously improve the validation process.

The process does not require significant technical knowledge to conceptually understand it; rather it relies on a group of technical experts who mediate between regulators and AV developers and work out the finer details of the safety assurance programme. In this document, technical implementation guidance is provided to support these steps so that the regulator can facilitate the high-level process.

# 2 Technical implementation guidance

## 2.1 Establishing a centre of excellence (CoE)

Regulators will need to collaborate with multiple stakeholders due to the highly technical nature of AV development. The CoE convenes technical expertise to establish and conduct AV safety assurance testing.

### Potential composition

The CoE should bring together multiple stakeholders including, but not limited to, the following representatives:

- Regulatory representatives from the jurisdiction under consideration
- Policy representatives who are experts in regional/national government regulation to ensure that lower- and higher-level regulations do not conflict
- AV providers who are experts in AV technology
- Researchers to provide an external perspective on technology, and who can propose additional validation methods

### Primary activities

The CoE should work with its constituents to develop the desired framework for the jurisdiction. These responsibilities may include:

- Defining possible deployment ODD(s)
- Determining interim test milestones (i.e. subset of deployment ODD to allow access to each stage of testing and validation)
- Establishing behavioural competencies and/or qualitative scenarios the AV could be exposed to and safely navigate within the deployment ODD
- Working with AV providers to convert qualitative scenarios into concrete scenarios (as this step is AV hardware dependent)
- Prioritizing scenarios to test earlier vs. later in the approval process (as a function of criticality and exposure to each scenario)
- Choosing appropriate tests and success criteria to demonstrate safe behaviour within each scenario and milestone
- Working with any local leaders as necessary to enforce regulations

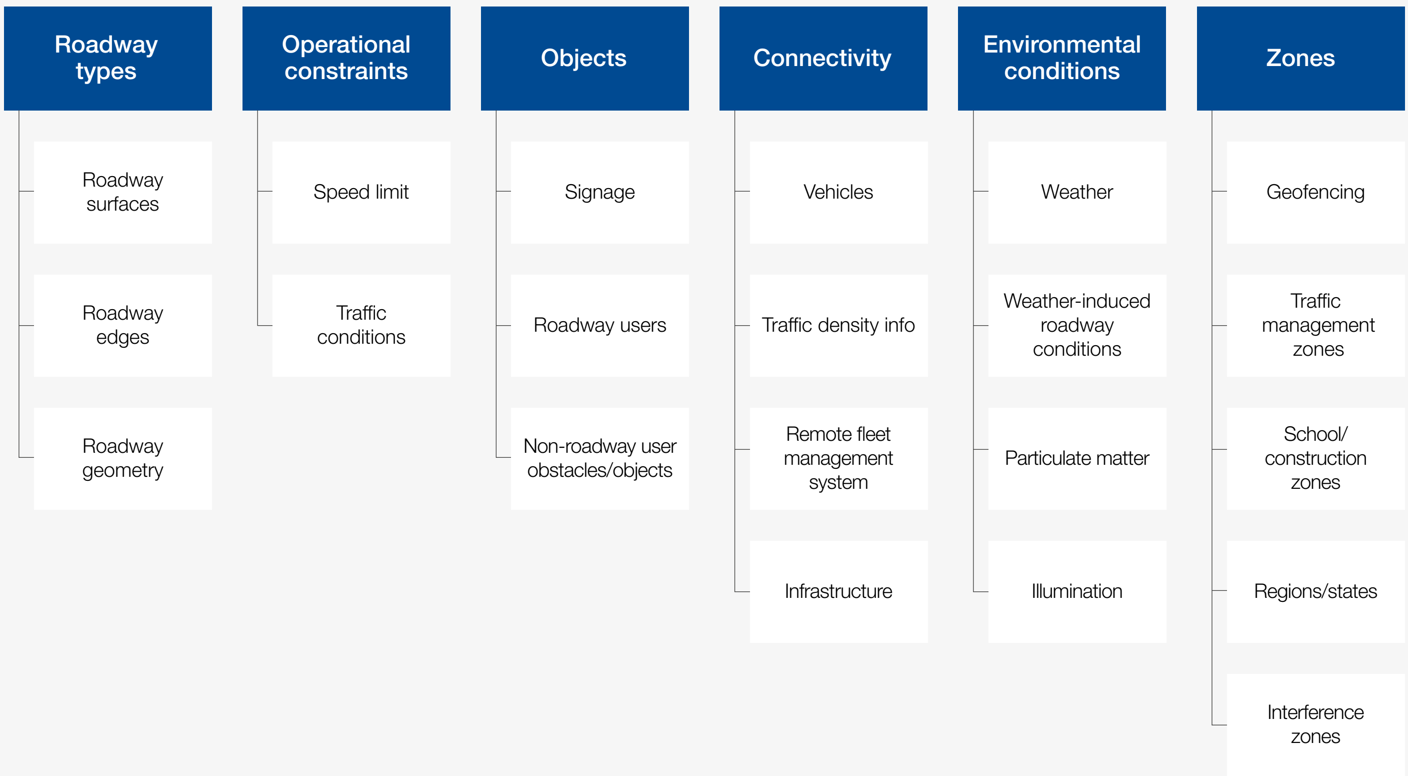
## 2.2 Determining ODD taxonomy

Under this proposed framework, regulators would certify vehicles to operate in one ODD at a time. An ODD defines the attributes of a vehicle's surrounding environment.

Defining the ODD is important, as the operational requirements for a suburban campus environment differ significantly from a dense urban city. An AV's operational requirements will also differ between cities and countries depending on various unique elements. For example, an AV in San Francisco will have to know how to behave around street cars in the fog, while an AV in Dubai will need to operate in extreme heat.

The ODD must be described with a logical taxonomy to enable it to be uniformly understood by AV companies. For example, the National Highway Traffic Safety Administration (NHTSA)[1] defines six categories to define for an AV: physical infrastructure, operational constraints, objects, connectivity, environmental conditions and zones (Figure 1). Other organizations such as the British Standards Institution (BSI)[2] have also produced ODD taxonomies. Regulators should choose the taxonomy most appropriate to their jurisdiction and describe the ODD at a qualitative level; quantitative ODD parameterization will remain the responsibility of the AV developer.

FIGURE 1 **Scope of NHTSA ODD taxonomy**




**Source:** Adapted from data provided by National Highway Traffic Safety Administration (NHTSA)

## 2.3 Defining the deployment ODD

The CoE should set the objectives for the policy framework through consultation with the AV developers interested in deploying in their jurisdiction, other government stakeholders and their community.

An example of this end objective could be: "*Create a testing and deployment programme that enables safe driverless operation of AVs on the streets of City X.*"

To provide context, an example city is shown in the following sections.

FIGURE 2 | **Example map of City X**

**Source:**
Openstreetmap, licensed under CCBY-SA2.0

In this example, we assume that the deployment ODD represents the largest geographic area in which the AV developers wish to operate, which also encompasses the broadest set of parameters for weather, time of day, traffic and other operating conditions. The deployment ODD should be a large enough geographic area operational window to support the necessary business case (e.g. robotaxi or delivery service). This is represented by ODD C in Figure 4.

Before defining the interim steps, the CoE should define and parameterize the deployment ODD in which the AV developer intends to commercially operate. Any defined ODD parameter should also be something of which the AV is aware (e.g. road type) or otherwise monitorable (e.g. weather), in accordance with the ODD taxonomy. Being aware of the ODD – so that the AV recognizes when it is operating within its specified parameters, and when it leaves those parameters and should enter a safe fallback state – is essential to safe operation.

## 2.4 Identifying interim milestones

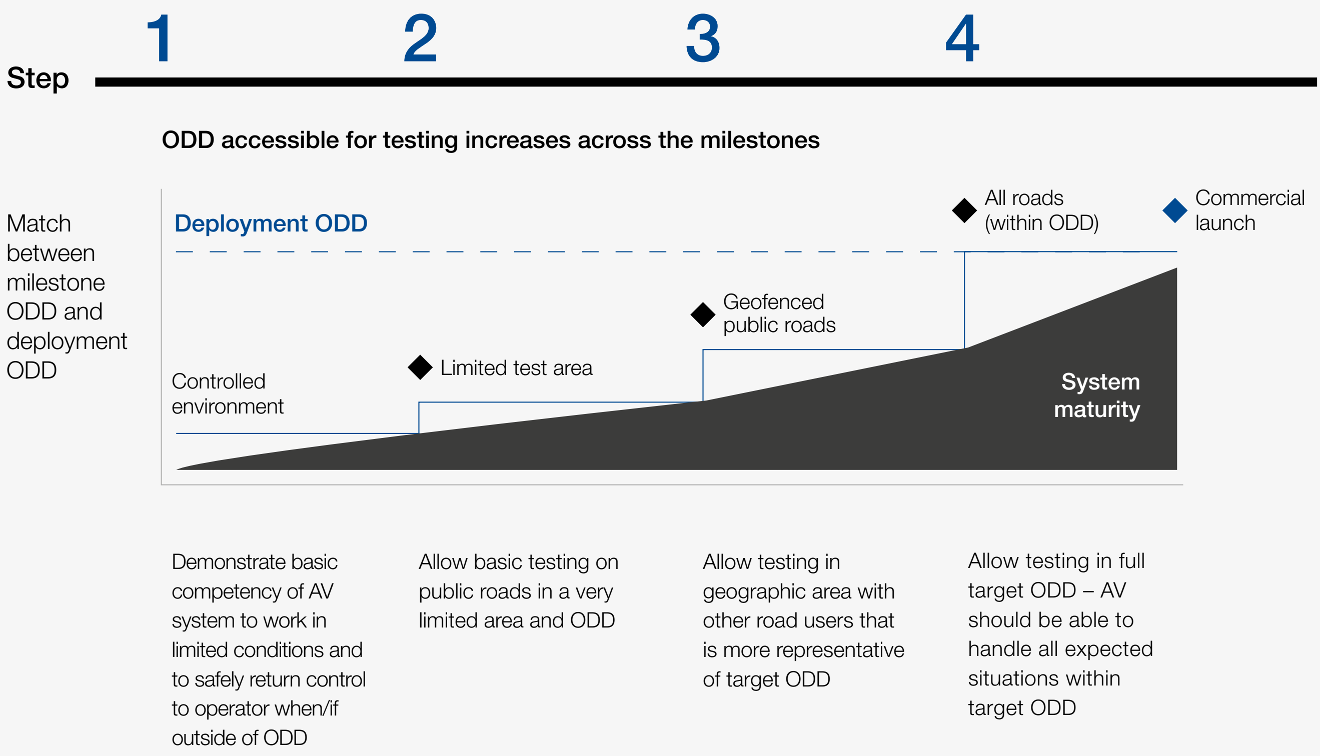
To ensure safe development and testing, AV developers should begin testing in low-risk areas, eventually expanding the test area to encompass the full deployment ODD.

While the number of interim milestones can vary, Figure 3 illustrates a four-stage process from initial test to commercial deployment. Each milestone's ODD gradually expands, eventually matching the target ODD.

A general process for defining the interim milestones is as follows:

1. Start with parameterized deployment ODD as the end state for Milestone 4, which may include certain parameter constraints such as inclement weather conditions
2. Identify a controlled area (which could be a test track, proving ground or closed streets) for basic competency demonstration in Milestone 1
3. Identify a limited geographic area that is a representative microcosm of the jurisdiction but provides a low-risk test bed for Milestone 2 (e.g. limited hours of operation in an area with few pedestrians and good infrastructure)
4. Identify a larger, more representative area for Milestone 3. This area should broadly reflect the deployment ODD, but may still have some limits (e.g. certain difficult streets) to support safe testing

FIGURE 3 | **Four example interim milestones**




**Source:** World Economic Forum / McKinsey & Co. Analysis

In this example application, we propose four milestones, expressed in terms of the deployment ODD:

– **Milestone 1:** A pre-deployment assessment of the basic competency of the AV, conducted in a controlled environment such as a test track, proving ground or private road, or possibly through the use of simulation. Successfully passing this assessment allows AV operators to progress to the next assessment.

– **Milestone 2:** Successfully passing this assessment allows for testing on public roads in a simple, defined ODD (Deployment Area A, with some limitations on time of day, weather conditions etc.), with limited levels of traffic, such as a business park, industrial district or campus.

  Illustrative areas for each of these milestones, and their respective ODDs, are overlaid onto the example map of "City X" in Figure 4.

  – Deployment Area A is an industrial park, with low daily vehicle traffic and few vulnerable road users (VRUs)

  – Deployment Area B is a subset of the roads in City X, and comprises a mix of road types, traffic throughput, infrastructure and VRUs that can be considered representative of the rest of the city

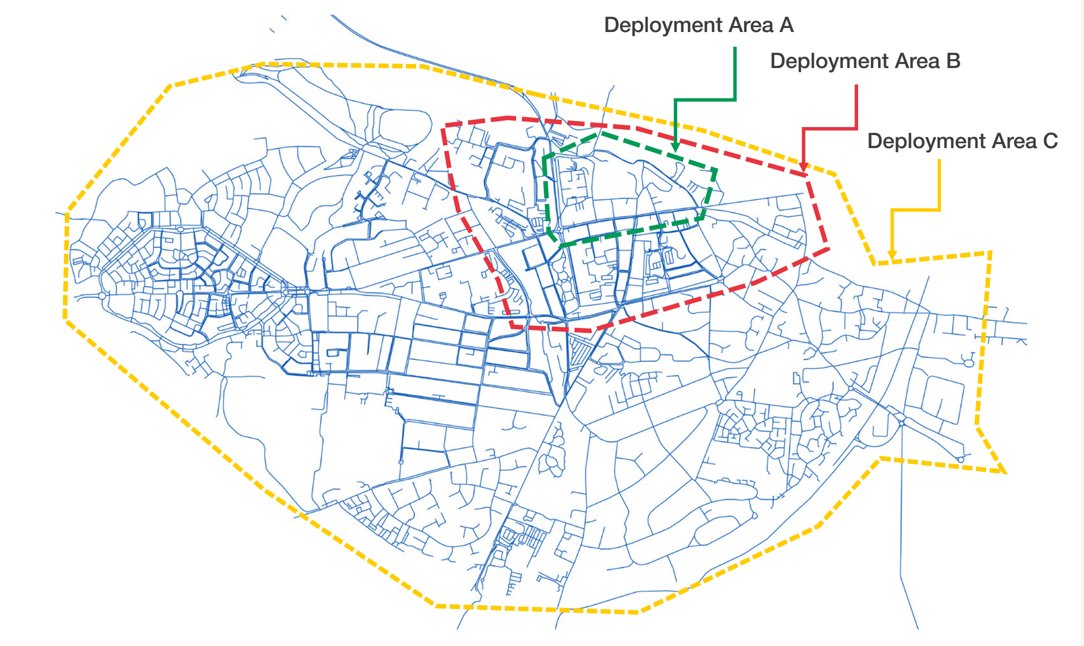
  – Deployment Area C is the deployment ODD, granting access to the full city.

  Milestone 2 should also open up the other parameters of the ODD, such as time of day, weather conditions and traffic permissions, to represent increased complexity.

– **Milestone 3:** Successfully passing this assessment allows for testing in a defined geographic area (Deployment Area B, with fewer restrictions on other ODD parameters), which is largely representative of the full city's driving environment in terms of road composition, infrastructure, traffic and VRUs. This is still a subset of the full environment to allow for reasonable monitoring by the regulator.

– **Milestone 4:** Successfully passing this assessment allows for unrestricted operation in the full deployment ODD (Deployment Area C, with few/no restrictions on other ODD parameters). The AV operator is permitted to remove the safety driver and/or launch a commercial service.

FIGURE 4 | City X incremental ODDs




**Source:** Openstreetmap, licensed under CCBY-SA2.0

## 2.5 Choosing scenarios for the interim milestones

Regulatory bodies should be responsible for defining behavioural competencies or qualitative scenarios – that is, high-level, naturalistic descriptions of situations an AV may encounter. These scenarios should be detailed enough that they can be parameterized, for the purposes of assessment within each ODD. The scenario parameterization process should be conducted by the AV developer, as the implementation of such parameterization in simulation and on-road testing will depend on the specific hardware and software used in the AV platform.

There are two approaches to generating a scenario set: a logical approach and a data-driven approach.

### Logical approach

A logical approach builds upon existing research on scenario-based AV safety and considers the needs of the operating environment broadly. As a starting point, there are a range of public scenario sets and behavioural competencies that have been published by regulators, such as NHTSA, road safety bodies such as Euro NCAP and even by AV developers including Waymo and Voyage. For example:

- NHTSA: A Framework for Automated Driving System Testable Cases and Scenarios[3]
- Waymo: Basic Behavioural Competency Testing[4]
- Voyage: Open Autonomous Safety Scenarios[5]
- CETRAN: Scenario Categories for the Assessment of Automated Vehicles[6]

These existing lists can inspire an initial set of scenarios that should correspond to Milestones 1 and 2 to demonstrate basic operational safety prior to deployment on the road.

Having selected from these initial scenario sets, the CoE should now consider what challenges the AV will encounter in each incremental ODD. Example questions to consider when deriving scenarios include:

- What types of road users are found in each ODD?
- What kinds of intersections and traffic infrastructure will the vehicle encounter?
- What are the speed limits in each ODD?
- What is the range of weather conditions in that operating environment?
- What time of day will the AVs be allowed to operate in that environment?

By working through a series of targeted questions, the CoE should be able to set a broad set of initial requirements for each ODD that can be expressed as a series of qualitative scenarios.

At this stage, it is necessary to outline the differences in classes of scenario. In this framework, we refer to three categories of scenarios:

– **Qualitative scenario:** An abstract description of a scenario in natural language, with definitions of the traffic situation, driving environment, other vehicles and road users, and environmental conditions – e.g. vehicle in traffic on a three-lane roadway on a summer's day.

– **Logical scenario:** A qualitative scenario that has been parameterized, including possible value ranges for each parameter; it may also include probability distributions for certain parameters – e.g. lane width 2.3–3.5 metres (m), traffic speed 0–30 kilometres/hour (km/hr), temperature 10–40°C.

– **Concrete scenario:** A logical scenario with specified values for each parameter. Such a description is grounded in its environment (context, with its parameters) and includes ego vehicle goals – e.g. lane width 2.3 m, traffic speed 30 km/hr, temperature 23°C.

Examples of each type of scenario are described in Table 2.

TABLE 2 **Example scenario categories**

| | Qualitative scenario | Logical scenario | Concrete scenario |
|---|---|---|---|
| **Example** | Vehicle in traffic on a three-lane curved motorway in the summer | Lane width: 2.3–3.5 m<br>Traffic speed: 0–30 km/hr<br>Temperature: 10–40°C<br>... | Lane width: 2.3 m<br>Traffic speed: 30 km/hr<br>Temperature: 23°C<br>... |
| **Selection methodology** | Selection based on criticality and frequency | Ranges to be selected based on evaluation of real-world driving data and assumptions based on physics and regulation | Selected based on simulation of logical scenarios and identification of critical parameter sets |
| **Input for selection process** | Clustering of real-world driving data<br>Accident reports<br>Insurance claims | Evaluation of real-world driving data<br>Driving dynamics model | Simulation of logical scenarios with all datasets |

**Source:** World Economic Forum / McKinsey & Co. analysis

The regulator should not seek to specify detail any deeper than the logical scenario level. To generate concrete scenarios, the exact parameters of each qualitative scenario should be defined by the AV developer, as the parameters will vary depending on the specific hardware and software used by the AV platform.

## Data-driven approach

Following an initial listing exercise to determine the logical scenarios, it may be possible to draw upon traffic data, where available, to further determine critical scenarios for each deployment ODD, especially those that are specific to a particular city or region. For example:

– According to reported collision data, what are the most frequent types of accident that occur in each deployment ODD?

– What is the typical traffic flow through the ODD, and how does that vary throughout the day?

– What were the circumstances of collisions involving vulnerable road users in that ODD? What was the outcome of these collisions?

– What unique road elements in the jurisdiction (e.g. transit modes, road users, infrastructure design) should the AV comfortably manage?

Such data can also be used to prioritize the scenarios that represent the collisions occurring in that ODD, and to further identify challenging traffic scenarios that the AV providers should satisfy to provide a degree of safety assurance.

## 2.6 Generating qualitative scenario library

Data-driven and analysis-based approaches can be combined as a basis for creating a scenario library specific to the jurisdiction and deployment ODD. Regulatory bodies should be responsible for defining *qualitative scenarios,* that is, high-level, abstract descriptions of scenarios that are capable of being parameterized. However, that parameterization process should be conducted by the AV developer, as the implementation of such parameterization in simulation and on-road testing will depend on the specific hardware and software used in the AV platform.

As part of this project, the World Economic Forum has partnered with our autonomous vehicle community to create Safety Pool, led by Deepen.ai, to generate an independent scenario library and data exchange to benefit AV companies and regulators alike. While Safety Pool can be used to provide reference scenario data for the purposes of this type of scenario-based assessment, it is also intended to function as a data brokerage between AV developers. Learn more at https://www.safetypool.ai/.

Other scenario libraries also exist that may provide an indication of the structure and coverage of such a scenario library, including Voyage's Open Autonomous Safety Scenarios[7] and the UK's government-sponsored MUSIC Consortium.[8]
As described in Table 3, regulators may choose whether to depend on pre-built scenario libraries vs. developing one themselves. Regardless of the source, the scenario library should be iteratively refined throughout the approval process.

TABLE 3 **Approaches to scenario library creation**

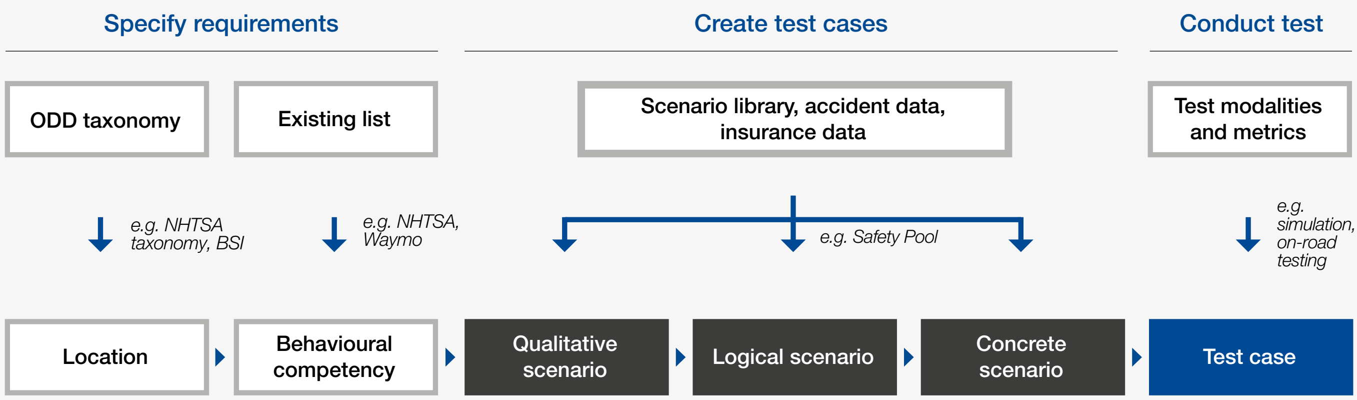
| | High touch | Medium touch | Low touch |
|---|---|---|---|
| **Description** | Regulator develops own scenario library from intended deployment environment | Regulator builds on existing scenario library to define competencies<br><br>(e.g. adopt Safety Pool scenario library) | AV developers provide scenario library/competencies they have used to demonstrate performance |
| **Benefits/challenges** | + Data set is drawn from deployment environment and hence fully representative<br><br>– Expensive, time-consuming, may have limited scenario coverage | + Reduced cost, builds on existing initiatives for scenario-based assessment, broader scenario coverage<br><br>– Unlikely to be significantly representative of deployment environment, suitable only for broad demonstration of competence<br><br>– Requires customization following pre-set taxonomy | + Lowest cost and effort to CoE<br><br>+ Generally preferred path for AV developers<br><br>– Places onus on AV companies to provide safety assurance, reduces insights to CoE and regulator<br><br>– Less independent verification possible of AV performance within specific regulatory jurisdiction |

## 2.7 Parameterizing scenarios

Qualitative scenarios should be defined by regulators, but AV developers are responsible for deriving parameterized scenarios that can be implemented and tested. Each scenario should be parameterized independently, though there may be similarities across scenario sets (e.g. parameters for pedestrians will be similar across multiple scenarios).

In general, the AV developer will identify key parameters for each scenario (e.g. road width, distance to pedestrian etc.) and consider the range of possible values for each parameter to which the AV may be exposed within the deployment ODD. These parameters and values may be based on ODD parameters, accident data, knowledge of scenario dynamics in this jurisdiction and/or existing parameters described in scenario libraries. This process will create *logical scenarios*. Choosing specific parameters to test within each range will create *concrete scenarios. Test cases* can then be created by defining the specific test parameters and criteria. Additional detail about this process is shown in Figure 5.

FIGURE 5 **Life cycle of a scenario**




| | Location | Behavioural competency | Qualitative scenario | Logical scenario | Concrete scenario | Test case |
|---|---|---|---|---|---|---|
| **Definition** | A particular context in which the behaviour will need to be performed, within the parameters of ODD taxonomy | The ability of an AV to operate in traffic conditions it will regularly encounter | A formal (but abstract) description of a traffic situation within the AV's ODD | A qualitative scenario that has been parameterized, including possible value ranges for each parameter | An instance of a logical scenario with specified values for each parameter | A concrete scenario that has been implemented in a test modality, along with specific metrics that determine success or failure |
| **Example[1]** | The intersection of Main St and Central Ave | Turn left | Vehicle in traffic on a three-lane curved motorway in the summer | Lane width: 2.3–3.5 m<br>Traffic speed 0–30 km/hr<br>Temperature: 10–40°C<br>… | Lane width: 2.3 m<br>Traffic speed: 30 km/hr<br>Temperature: 23°C<br>… | |

☐ Inputs
■ Scenario type

[1]Example based on Pegasus Project https://www.pegasusprojekt.de/files/tmpl/PDF-Symposium/04_Scenario-Description.pdf (link as of 29/9/20).

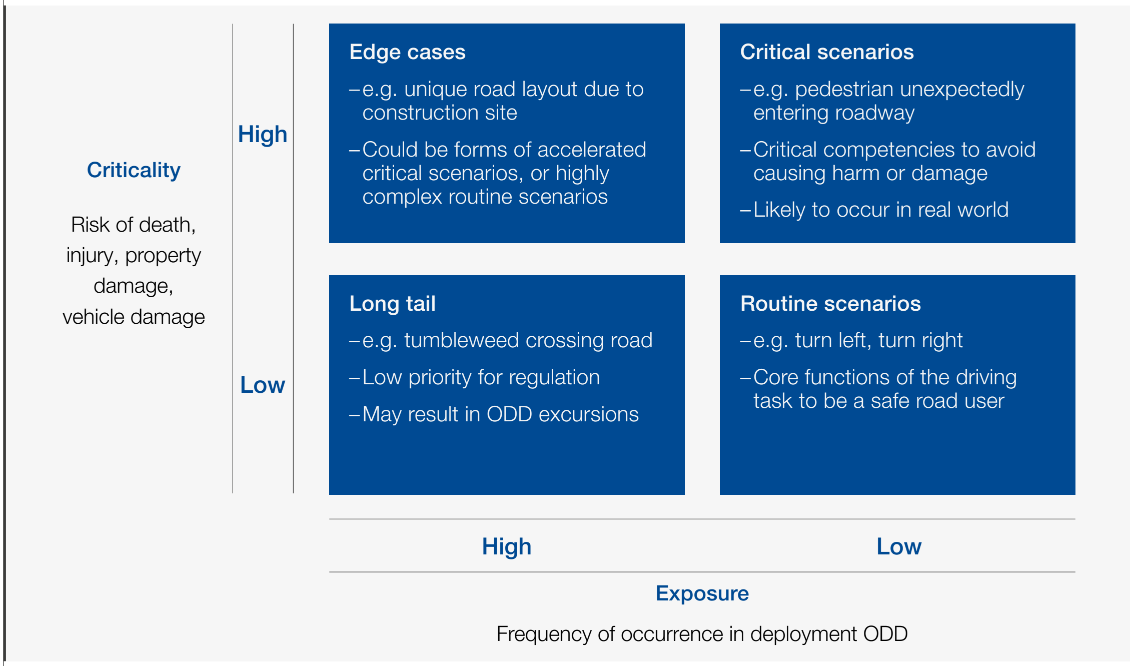
**Source:** World Economic Forum / McKinsey & Co. Analysis

## 2.8 Prioritizing scenarios

As it is not possible to test all scenarios simultaneously, the resulting concrete scenario list should be prioritized based on a risk assessment informed by two factors: exposure and criticality. Scenarios that occur frequently and present a high risk to other road users (e.g. a pedestrian crossing the road) should be prioritized for testing earlier in the safety assurance process.

Estimating criticality and frequency would be most accurately captured from naturalistic driving data. However, in the absence of such information, these factors can be estimated by evaluating data such as accident reports, considering a variety of metrics including death rate, injury rate and typical property damage for a given scenario. These evaluations need not be a highly quantitative, and can be adjusted throughout the process based on information gathered during testing.

FIGURE 6 **Categorizing scenarios**




**Source:** World Economic Forum / McKinsey & Co. Analysis

Additionally, the assessor should seek ways to introduce entropy (such as unusual variables, objects or events) to scenarios for assessment wherever possible, to increase the robustness of the assessment. This could be true for tests evaluated across all modalities of simulation, on the road or in a controlled environment.

## 2.9 Selecting tests for each milestone and scenario

For each scenario, the CoE and AV provider should work to determine which type of test environment (e.g. simulation, controlled track or on-road in ODD) is the most appropriate in order to make a reasonable and safe demonstration of the AV system's performance.

There are three general types of tests for each scenario:

– **Simulation:** Highly parameterized digital twins of roads that can be used to test AV systems under many different parameter values and estimate the vehicle's expected response

– **Test track:** A controlled environment that mimics a city infrastructure and can be used to simulate real-world driving in a lower-risk situation. Can enable verification of simulation scenarios in a physical environment with minimal risk.

– **Naturalistic on-road testing:** Using public roads for testing in regular traffic, typically supervised by a safety driver or remote operator. This is necessary to prove the safety of the vehicle in the real world, but also presents the most risk to the public.

There are trade-offs in the use of each testing environment. For example, it may be possible to test complex edge cases with severe potential outcomes for vulnerable road users only in simulation. Moreover, controlled test environments offer a high degree of repeatability, but lack the true complexity of the real world.

TABLE 4 Guiding principles for test selection

| Test type | Definition | When to use | When not to use |
|---|---|---|---|
| **Simulation** | Highly parameterized digital twins of roads that can be used to test many different parameter values and estimate the vehicle's expected response | In scenarios with many different parameters that can vary widely<br><br>In very rare or catastrophic situations such as collisions with other vehicles | When mapping between simulation and real world is not accurate<br><br>Where on-road testing has already demonstrated sufficient competency |
| **Controlled environment** | A controlled area such as a test track that mimics city infrastructure and recreates real-world driving scenarios in a low-risk context | When comparing realism of simulation to real world in a low-risk setting<br><br>For critical events that the AV will likely encounter, but which are dangerous to first test in naturalistic setting<br><br>To evaluates initial competency of AV system | In highly complex scenarios requiring exact timing and multiple road users to be a useful test case |
| **Naturalistic driving** | Using public roads and regular traffic for testing, traditionally supervised by a human such as a safety driver or remote operator | To test high-frequency scenarios, both low-risk and high-risk (once system has demonstrated sufficient competency in simulation) | When testing very rare situations the AV may not encounter during the validation process<br><br>In potentially catastrophic situations<br><br>Situations requiring tight control over parameter values |

The degree of involvement taken by the regulator can vary at this stage, as illustrated in Table 5.

TABLE 5 Potential levels of involvement of the regulator in defining test conditions

| | High touch | Medium touch | Low touch |
|---|---|---|---|
| **Description** | Regulator specifies exact tests and metrics to conduct for each parameter of each scenario<br><br>Regulator introduces forced entropy variables into scenarios to randomize aspects of the situation | Regulator specifies a few key scenario tests (e.g. must test emergency stop for pedestrian on a test track) and metrics (e.g. ODD excursions), leaving other scenarios open for AV developer to define successful outcomes | AV developers can report successful tests using whichever method they choose |
| **Benefits/ challenges** | + Provides greatest clarity on how each scenario is being evaluated<br><br>– Challenging for regulator to specify correct test to run for each scenario<br><br>– Time-consuming to evaluate each scenario (and parameters for each scenario) individually | + Regulator can ensure key scenarios most important to public safety are sufficiently evaluated<br><br>– Reduced insights into metrics and performance indicators | + Quickest path to approval as AV providers can provide documentation of existing tests<br><br>– Relies on AV provider to provide assurance methods with suitable transparency |

# 2.10 Allocating tests across modalities

## Simulation-based assessment

The CoE will have to choose between specifying a selection of approved simulation environments and allowing AV developers to demonstrate performance in their existing simulation tools. Additionally, the CoE should create a reporting programme to allow the AV developers to submit evidence of their performance in simulation, and an audit process if necessary. Any such reporting programme must ensure that AV companies' intellectual property and sensitive business information remains secure, as well as ensuring that consumer privacy, if applicable, is maintained.

TABLE 6 **Approaches to simulation-based testing**

| | High touch | Medium touch | Low touch |
|---|---|---|---|
| **Description** | Select and specify approved simulation tool(s) for AV operators to use; establish structured reporting programme to demonstrate performance against specified scenario set | Allow AV operators to select own simulation tools; establish structured reporting programme to demonstrate performance against specified scenario set | AV developers self-report scenario performances from own simulation tools; CoE audits responses |
| **Benefits/ challenges** | + High degree of control for CoE, uniform reporting process, common assessment across simulation tools<br><br>– Requires considerable knowledge of simulation tools available and contract bid/approval process<br><br>– Limits innovation, as it assumes one approach to simulation and may discourage companies that have already developed proprietary simulation tools from testing in a given jurisdiction | + Creates uniform reporting process; low cost to AV developers and regulators<br><br>+ Does not require CoE/ regulator to invest in building simulation tools and/or in-house technical expertise<br><br>– Variation in simulation fidelity, limited transparency for regulator | + Easiest to implement<br><br>– Limited insights to CoE |

## Controlled environment testing

A controlled environment or test track is most useful for structured testing in a repeatable fashion. AV operators will continue to use structured testing to recreate scenarios throughout their development cycle to validate a range of functions from emergency stops to verifying the fidelity of their simulation for certain scenarios. In this programme, the CoE will have to choose between creating its own test environment, allowing for testing at a third-party site, or asking the AV operators to self-certify their own track-based testing.

TABLE 7 Approaches to controlled environment testing

| | High touch | Medium touch | Low touch |
|---|---|---|---|
| **Description** | Build own test track for operational safety assessment | Allow for testing at existing proving ground | Require AV developers to attest to track testing before on-road deployment |
| **Benefits/ challenges** | + Full control of design to create a representative test environment<br><br>– Most expensive<br><br>– Longest lead time | + Short lead time<br><br>– Limited number of specialist AV test sites globally<br><br>– May not be representative of deployment environment | + No cost to regulator<br><br>– Relies on AV provider to structure tests and report results with transparency |

## On-road testing

As with other assessment modes, the CoE can choose between a highly structured approach, with a practical assessment akin to a driving test, and a less-structured approach, such as observing AVs on the road or allowing for self-reporting by AV operators. The CoE could apply a combination of approaches in this domain.

TABLE 8 Approaches to on-road testing

| | High touch | Medium touch | Low touch |
|---|---|---|---|
| **Description** | Conduct a ride-along assessment of each AV | Observe and audit AV operators on the road | Require AV developers to submit evidence demonstrating a range of scenario performance in real- world traffic |
| **Benefits/ challenges** | + Provides first-hand experience of vehicle in deployment environment<br><br>– Requires evaluator training<br><br>– Risk to employees<br><br>– One-time evaluation may not be indicative of ongoing progress | + Scope for regulator to specify range of reporting and audit according to available resources and priorities<br><br>+ Allows for regular insights between milestones<br><br>– Reduced insight compared to ride-along assessment<br><br>– Requires CoE/regulator investment in technical expertise necessary to evaluate, as well as robust intellectual property and trade secret protections for AV companies | + Scope to specify a range of reporting types |

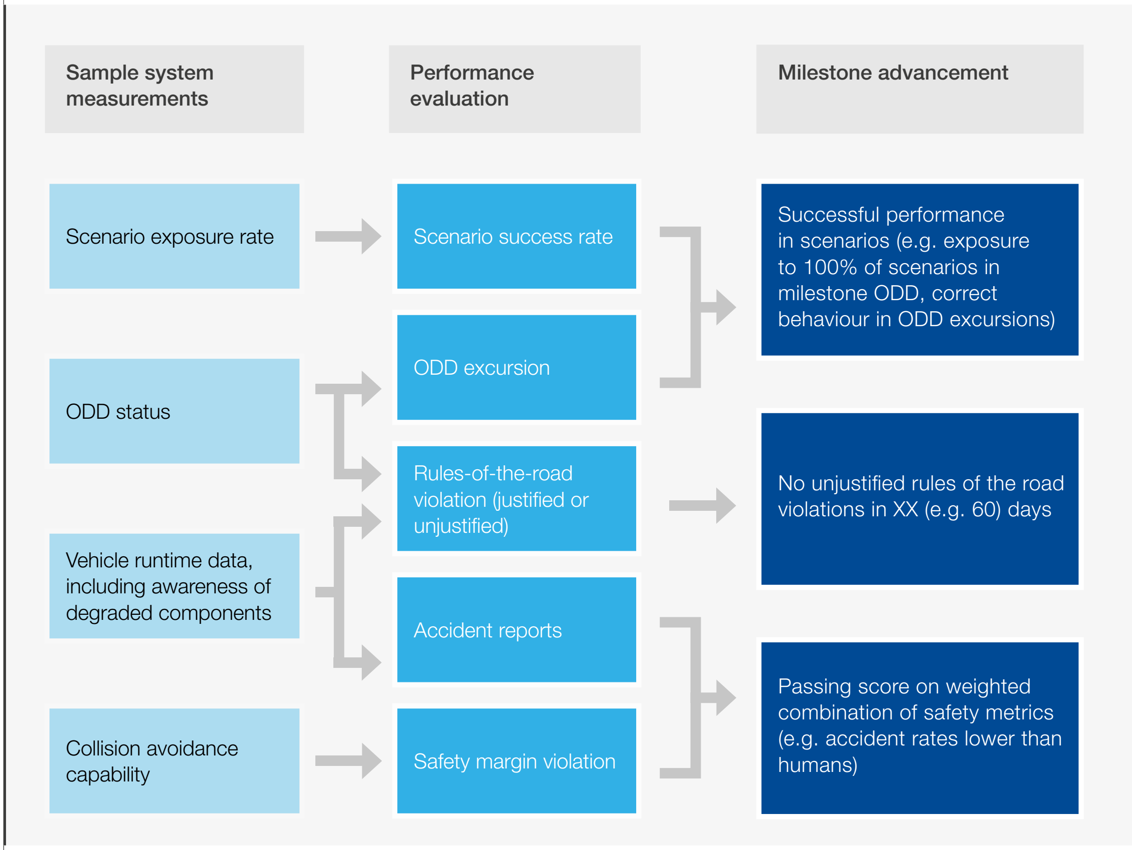
## 2.11 Determining success/failure/advancement criteria

Converting the vast amount of data collected during AV testing into useful metrics is challenging and also opens up AV developers to liability and public disclosure of companies' intellectual property rights and sensitive business information. As such, the decision about evaluation metrics is critical, and should consider not only **what** is being measured, but also **why** it is important to evaluate system performance, as well as **how** that measure informs advancement to each milestone.

Figure 7 summarizes numerous possible metrics to evaluate the system, infer performance and determine advancement to the next milestone.

FIGURE 7

**Evaluation metrics**




**Source:**
World Economic Forum / McKinsey & Co. Analysis

Additionally, there are a range of industry standards and technical guidance on the subject of safety indicators and metrics. Examples include:

- ANSI/UL 4600: Safety Performance Indicators[9]
- PAS 1880: Guidelines for Developing and Assessing Control Systems for Automated Vehicles[10]
- PAS 1881: Assuring the Safety of Automated Vehicle Trials and Testing – Specification[11]
- SAE: Collision Avoidance Capability[12]
- SAE: Driving Safety Performance Assessment Metrics for ADS Equipped Vehicles[13]
- PAS 1882: Data Collection and Management for AV Trials[14]
- PAS 1883: Hierarchical Taxonomy for Specifying an Operational Design Domain (ODD)[15]

One further metric to consider is ODD excursions. At a high level, an AV should be capable of measuring whether or not it is currently within its test or deployment ODD, as defined for the milestone or jurisdiction. This information may be a combination of what is sensed on board (e.g. road type, presence of road users, temperature), as well as off-board (e.g. incoming weather pattern, traffic density etc.). The exact inputs to measure and report this ODD should be chosen by the AV developer. This area is the subject of a number of ongoing research studies and there is not presently an industry-agreed approach to define or report this.

## 2.12 Verifying the AV backup system

Due to the complexity of the roadway, for the foreseeable future AVs will encounter unplanned or unexpected situations that will require the vehicle to fall back to a minimal risk condition (MRC) or trigger the intervention of a backup driver to bring the vehicle to a safe state. The exact backup mechanisms should be left up to the AV provider, as each AV platform comprises different components, and has different strengths and weaknesses.

In general, there are three states in which an AV could operate:

– **Normal operation:** The AV is operating sufficiently in the ODD. This state includes degraded performance, where individual components may fail but the system can still independently achieve its task.

– **Safe fallback state:** The AV is aware of a system fault, an imminent or prolonged ODD excursion, and seeks to transition to a safe state as a result. For example, this could initiate a minimal risk manoeuvre (MRM), with the goal of achieving an MRC, such as stopping safely at the side of the road.

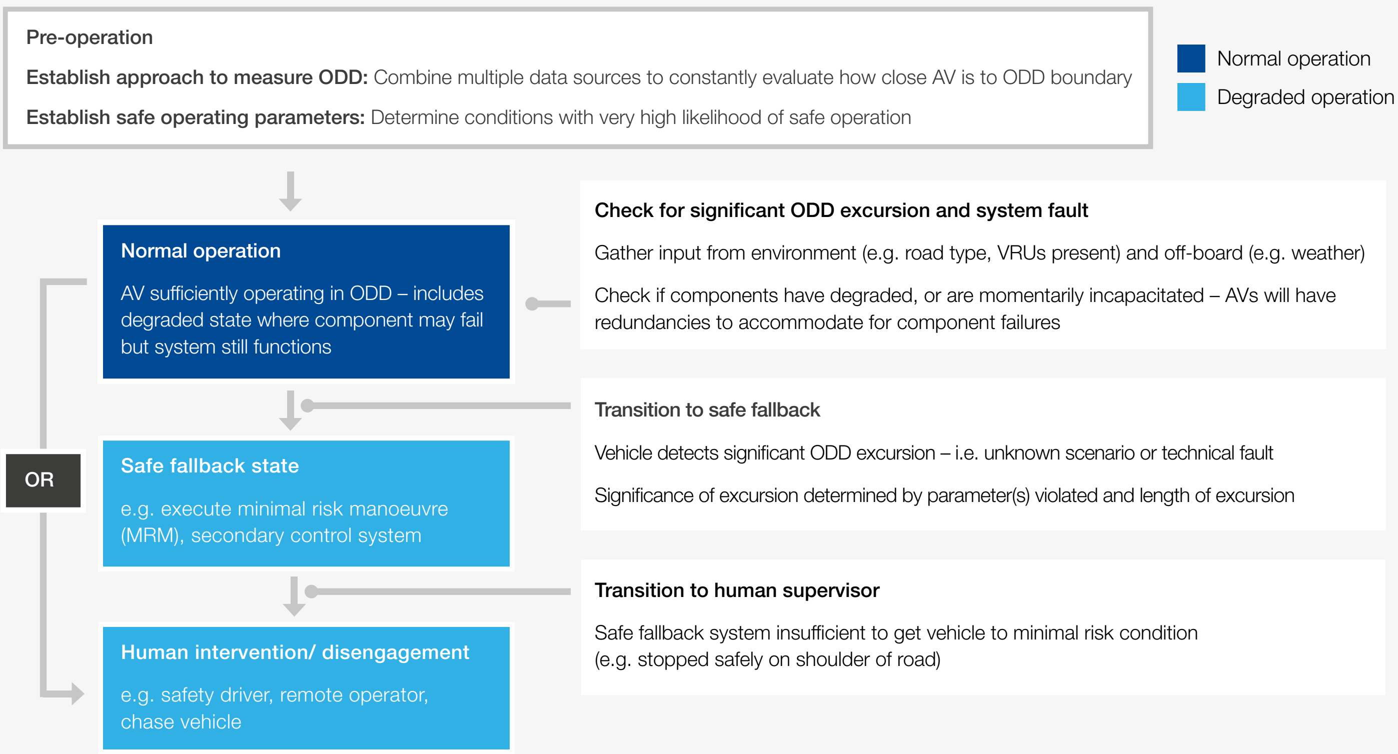
– **Human intervention:** Either handover of control to a safety driver (in-vehicle or a remote operator), or other human intervention to disengage the autonomous functionality.

AVs should return to a safe fallback state if they are aware that they have departed from their test ODD due to a new scenario or a significant external change such as weather. What constitutes a significant enough departure from the ODD is an ongoing research topic – an instantaneous ODD departure, such as a few drops of rain when the vehicle is not approved to drive in wet conditions, may not be enough to warrant a safe fallback state. The AV developer should monitor for some level of hysteresis for these situations.

To verify the backup systems, regulators should ask AV operators to detail potential triggers and define the expected fallback actions for the AV to execute. This may include a list of minimal risk manoeuvres that the AV can execute, such as pulling to the side of the road. Regulators may choose to require a demonstration of an AV's abilities in this regard by inducing faults in on-road test in challenging situations (e.g. a low connectivity area for a remote operator) in the course of an on-road scenario or requiring a demonstration during simulation or controlled course testing.

Human intervention systems need not be constrained to safety drivers in the vehicle. For instance, AV developers may be able to effectively demonstrate the use of teleoperation as a human backup system. At scale, a human operator may even supervise multiple AVs simultaneously, and regulatory policies should allow for innovation to flourish.

FIGURE 8 **Backup system hierarchy**




**Source:** World Economic Forum / McKinsey & Co. Analysis

# 3 Reference terminology and definitions

## 3.1 General terms

- **Automated driving system:** The hardware and software that are collectively capable of performing the entire dynamic driving task on a sustained basis, regardless of whether it is limited to a specific operational design domain.

- **Autonomous vehicle (AV):** A vehicle equipped with an automated driving system designed to function without a human driver as a Level 4 or 5 system under SAE J3016.

- **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.

- **Operational design domain (ODD):** A description of the specific operating domain(s) in which an automated driving system is designed to properly operate, including but not limited to roadway types, speed range, environmental conditions (weather, daytime/night-time etc.) and other domain constraints.

## 3.2 Scenarios and related terms

- **Scenario:** A traffic situation within the vehicle's operational design domain.

- **Behavioural competency:** A manoeuvre or function that an automated vehicle can demonstrate in various scenarios – for example: Turn Left or Emergency Stop.

- **Scenario-based assessment:** Evaluating a system based on its performance when exposed to a variety of predefined scenarios that correspond to its intended deployment ODD.

- **Minimal risk condition:** A condition of the autonomous vehicle or system to which either the user (safety operator) or the system itself brings the vehicle to reduce the risk of a crash when a given trip cannot or should not be completed. For example, a minimal risk condition might entail "bringing the vehicle to a stop in its current travel path" or "a more extensive manoeuvre designed to remove the vehicle from an active lane of traffic".

In this framework, we refer to three categories of scenarios:

– **Qualitative scenario:** An abstract description of a scenario in natural language, with definitions of the traffic situation, driving environment, other vehicles and road users and environmental conditions – e.g. vehicle in traffic on a three-lane roadway on a summer day.

– **Logical scenario:** A qualitative scenario that has been parameterized, including possible value ranges for each parameter; it may also include probability distributions for certain parameters – e.g. lane width 2.3–3.5 metres (m), traffic speed 0–30 kilometres/hour (km/hr), temperature: 10–40°C.

– **Concrete scenario:** A logical scenario with specified values for each parameter. Such a description is grounded in its environment (context, with its parameters) and includes ego vehicle goals – e.g. lane width 2.3 m, traffic speed 30 km/hr, temperature 23°C.

## 3.3 Defining verification methods

– **Simulation:** Highly parameterized digital twins of roads that can be used to test AV systems under many different parameter values and estimate the vehicle's expected response.

– **Test track:** A controlled environment that mimics a city infrastructure and can be used to simulate real-world driving in a lower-risk situation. Can enable verification of simulation scenarios in a physical environment with minimal risk.

– **Naturalistic on-road testing:** Using public roads for testing in regular traffic, typically supervised by a safety driver or remote operator. This is necessary to prove the safety of the vehicle in the real world, but also presents the most risk to the public.

# Contributors

The World Economic Forum's Safe Drive Initiative is a global, multistakeholder and cross-disciplinary initiative intended to help shape the development of successful autonomous vehicle policy and improve the safety of AV pilots. The project has engaged leaders from private companies, governments, civil society organizations and academia to understand AV policy, identify challenges and define principles to guide future policy solutions. The opinions expressed herein may not correspond with the opinions of all members and organizations involved in the project.

## Lead Author

**Tim Dawkins**
Project Lead, Automotive and Autonomous Mobility, World Economic Forum

We would like to thank our Safe Drive Initiative project community as well as the following contributors for their insights and contributions to this framework:

**Atul Acharya**
Director, Autonomous Vehicles Strategy, A3 Ventures

**Michelle Avary**
Head of Automotive and Autonomous Mobility, World Economic Forum

**Bryn Balcombe**
Founder, The Autonomous Driving Alliance

**Xantha Bruso**
Manager, Autonomous Vehicle Policy, AAA

**Nicola Croce**
Technical Program Manager, Safety Pool, Deepen.ai

**Thomas Form**
Senior Director Autonomous Driving, Volkswagen

**Kiran Jesudasan**
Transportation Systems Specialist, dRISK

**Mary Joyce**
Global Vice-President and General Manager, Mobility & Automotive Division, UL

**Martin Kellner**
Engagement Manager, McKinsey & Company

**Siddartha Khastgir**
Head of Verification & Validation – Intelligent Vehicles, WMG, University of Warwick

**Stefano Marzani**
Principal Specialist, Autonomous Driving, Amazon

**Dylan Moore**
Associate, McKinsey & Company

**Mohammad Musa**
Chief Executive Officer and Co-Founder, Deepen.ai

**Stefan Seltz-Axmacher**
Chief Executive Officer and Co-Founder, Starsky Robotics

**Sanjit Seshia**
Professor, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

**Emily Shao**
Associate Partner, McKinsey & Company

**Barnaby Simkin**
Autonomous Driving Standards Manager, NVIDIA

**Jann-Eve Stavesand**
Head of Consulting, dSPACE

**Chess Stetson**
Chief Executive Officer, dRISK

**Michael Watson**
Senior Controls Engineer, NVIDIA

**Ismail Zohdy**
Program Manager for Self-Driving Transport Strategy, Dubai Road and Transport Authority

# Endnotes


1. NHTSA: A Framework for Automated Driving System Testable Cases and Scenarios: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13882-automateddrivingsystems_092618_v1a_tag.pdf (link as of 7/10/20).
2. PAS 1883: Hierarchical Taxonomy for Specifying an Operational Design Domain (ODD): https://www.bsigroup.com/en-GB/CAV/pas-1883/ (link as of 7/10/20).
3. NHTSA: A Framework for Automated Driving System Testable Cases and Scenarios.
4. Waymo: Basic Behavioural Competency Testing: https://storage.googleapis.com/sdc-prod/v1/safety-report/Safety%20Report%202018.pdf#page=36 (link as of 7/10/20).
5. Voyage: Open Autonomous Safety Scenarios: https://oas.voyage.auto/scenarios/ (link as of 7/10/20).
6. CETRAN: Scenario Categories for the Assessment of Automated Vehicles: https://cetran.sg/wp-content/uploads/2020/01/REP200121_Scenario_Categories_v1.7.pdf (link as of 7/10/20).
7. Voyage: Open Autonomous Safety Scenarios.
8. Catapult Transport Systems: About the MUSICC Project: https://ts.catapult.org.uk/innovation-centre/cav/cav-projects-at-the-tsc/musicc/ (link as of 7/10/20).
9. UL 4600: Standard for Safety for the Evaluation of Autonomous Products: https://edge-case-research.com/ul4600/ (link as of 7/10/20).
10. PAS 1880: Guidelines for Developing and Assessing Control Systems for Automated Vehicles: https://www.bsigroup.com/en-GB/CAV/pas-1880/ (link as of 7/10/20).
11. PAS 1881: Assuring the Safety of Automated Vehicle Trials and Testing – Specification: https://www.bsigroup.com/en-GB/CAV/pas-1881/ (link as of 7/10/20).
12. Development and Application of a Collision Avoidance Capability Metric: https://www.sae.org/publications/technical-papers/content/2020-01-1207/ (link as of 7/10/20).
13. Driving Safety Performance Assessment Metrics for ADS-Equipped Vehicles: https://www.sae.org/publications/technical-papers/content/2020-01-1206/ (link as of 7/10/20).
14. PAS 1882: Data Collection and Management for AV Trials: https://standardsdevelopment.bsigroup.com/projects/2019-03951 (link as of 7/10/20).
15. PAS 1883: Hierarchical Taxonomy for Specifying an Operational Design Domain (ODD): https://www.bsigroup.com/en-GB/CAV/pas-1883/ (link as of 28/10/20).

COMMITTED TO
IMPROVING THE STATE
OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744
contact@weforum.org
www.weforum.org