Using machine learning to develop

natural, human like vehicle control

HumanDrive: Autonomous Vehicle Project Safety Management

Richard Hillman, Transport Systems Catapult

Date: March 2019

HUMAN DRIVE

Executive Summary

HumanDrive is a collaborative R&D project, part funded by innovate UK, that is developing an autonomous system that uses artificial intelligence and deep learning to enable more 'humanistic' driving behaviour than that typically associated with Autonomous Vehicle (AV) prototypes, with the project culminating in a 'Grand Drive' of over 200 miles, incorporating a range of challenging road types.

This document describes the approach to managing safety within the HumanDrive project, with the overall objectives being to:

1. **Ensure an acceptable level of safety is achieved for those involved in the project;**
2. **Ensure an acceptable level of safety for members of the public in the vicinity of HumanDrive vehicles;**
3. **Provide an example of a safety case for a vehicle operating at higher levels of automation such that the knowledge gained can be exploited by other projects.**

The safety of the project is divided into two separate areas: operational safety and system safety.

Operational Safety considers how safe procedures will be used to deploy the system in its surrounding environment. It includes:

- Risk assessments and method statements for each type of test to describe how to conduct the tests in a safe manner;
- Assessment of the ability of the safety driver to intervene, bearing in mind driver qualification, driver awareness of the HumanDrive system's limits, and the quality of the Human-Machine Interface (e.g. whether it is clear to the driver what the current operating mode of the driver is, whether warnings are given clearly, whether the driver is able to override the system without having to fight against it);
- A review of the route to identify any particular hazards to take into consideration.

System Safety considers the Functional Safety and Safety of the Intended Function (SOTIF) of the vehicle itself, particularly focussed upon the automated driving system but also considering safety of the base vehicle. It includes:

- Verification that the vehicle meets a comprehensive set of safety requirements when subjected to specific test cases (the creation of these requirements representing a significant volume of work in itself);
- Validation to ascertain whether the vehicle performs safely during extended testing mileages when tested in increasingly challenging scenarios;
- Configuration control procedures to ensure that updates are introduced safely.

As an R&D project developing cutting-edge technology, it is expected that the system will not reach production vehicle levels of robustness, and will therefore occasionally take sub-optimal paths. Ultimately, the safety driver is the main method of mitigating against risk, and therefore the key to the safety of the project is ensuring that the vehicle is only operated autonomously in situations where the safety driver has sufficient time and situational awareness to intervene if required. For example, on narrow country roads, it is important to ensure that there is sufficient space for the safety driver to keep the vehicle in lane should the system make a steering error in the vicinity of oncoming traffic, nearby pedestrians etc.

Contents

1. Introduction

1.1 Project Description

The HumanDrive consortium, led by Nissan Technical Centre Europe (NTCE), is developing a prototype Autonomous Vehicle (AV) capable of achieving high levels of automation, with the aim of successfully demonstrating an autonomous journey from Cranfield to Sunderland (over 200 miles) in live traffic. This will include navigating country roads, motorways and dual carriageways. The project commenced in July 2017 and runs for a duration of 30 months, finishing at the end of December 2019.

One of the major innovative aspects of HumanDrive will be the development of an advanced control system designed to allow the AV to emulate a natural, human-like driving style. A key aspect of this is the integration of an artificial intelligence (AI) controller utilising artificial neural networks and deep learning for perception and decision-making.

The objective of the safety management work being undertaken in HumanDrive is to ensure that the project is conducted in an appropriately safe manner such that the risk level presented to both those involved in the project and the wider public are at an acceptable level, and all reasonably proportionate steps to identify and mitigate risks are taken. It is therefore expected that the project will be in line with industry best practice for safety, including compliance with all safety responsibilities set out in 'the Pathway to Driverless Cars: A Code of Practice for testing' (2015), compliance with all UK road traffic laws, and compliance with Highways England's GG104 risk assessment process. It should be noted that a Code of Practice update was introduced during the life of the project (Code of Practice: Automated Vehicle Trialling, 2019); this was not analysed to the same level detail as it was introduced when the safety case for the project was already at an advanced stage, but was reviewed to confirm that there were no new requirements that were at odds with the existing safety plan.

Provided that the overall safety of the project can be shown to be at least equivalent to the existing manually-driven vehicle fleet, it is not proportionate to apply production level standards to the functional safety of an R&D system, especially bearing in mind that many of these standards are not necessarily intended for high levels of autonomous driving. However, three standards in particular were reviewed and borne in mind as a benchmark:

- **ISO 26262 Road Vehicles – Functional Safety (2011)**: in particular, this influenced the approach to requirement management and to the Hazard Analysis and Risk Assessment (HARA).
- **ISO/PAS 21448 Road Vehicles- Safety of the Intended Functionality (2019)**: this was released midway through the project, but was reviewed once available and deemed to be broadly compatible with approach being taken in the HumanDrive project.
- **ROGS (2006)**: This is the regulation governing 'Railways and Other Guided Transport Systems' in the UK, which includes procedures for safe operation within a controlled environment; this was relevant to the approach taken to safe working practices within HumanDrive, particularly when testing on private tracks (an AV limited to a pre-determined route could be viewed as a guided transport system). Although less directly relevant than the other sources, this was still a useful benchmark, bearing in mind that automotive safety standards don't include operational procedures.

The Transport Systems Catapult (TSC), with support from Horiba MIRA, are providing advisory reports to NTCE. NTCE are ultimately responsible for the safety of HumanDrive vehicles within the project.

Cyber security is considered by a separate work package within the HumanDrive project, so does not directly form part of the safety case. However, as an R&D system in the early stages of development, it is anticipated that the system will on occasion produce suboptimal outputs, and the safety case must therefore prioritise ensuring that the driver is able to correct any mistake that the system could plausibly make. As such, it makes little difference to the safety of the project whether an error is caused by a fault (functional safety), a limitation of the designed system (SOTIF), or a cyber-attack, provided that there is sufficient confidence that suboptimal outputs will be robustly limited to levels that are controllable by the safety driver in all cases.

As the system will be integrated into a production Nissan Leaf vehicle, there is a high level of safety assurance regarding both the mechanical components on the base vehicle, which has been subjected to the European Community Whole Vehicle Type Approval (ECWVTA) process, requiring passing a range of tests relating to safety in crashes, likelihood of injuring pedestrians, anti-lock braking performance etc. It is therefore unnecessary to repeat any such testing, and oversight of these aspects will be limited to a visual inspection of the changes that have been made (e.g. confirming no hardware has been mounted where it could contact an occupant, pedestrian or airbag in an accident, ensuring all hardware is sufficiently secure). However, it should be borne in mind that any AV projects using a non-production vehicle would need to incorporate additional steps into their safety plan to address the risks that have been mitigated by ECWTA in the case of HumanDrive.

1.2 Scope of the Safety Work Package

A summary of the safety case scope is set out in table 1.

In scope	Out of scope
Operational safety (i.e. safe methods of testing the AV)	Full adherence to standards aimed at production vehicles (e.g. ISO 26262)
Functional safety of the autonomous system (i.e. ensuring that faults don’t present an unacceptable safety risk)	Workshop safety during modification, installation or maintenance work upon the base vehicles (this will be covered by existing Nissan procedures)
SOTIF of the autonomous system (i.e. ensuring that when performing as intended without faults, the design is suitably safe)	Office safety during desk-based engineering or administrative activities (covered by existing Health and Safety policies of project partners
Risk assessment of how trials are conducted	Analysis or mitigation of risks not associated with safety (e.g. financial risks, reputational risks, which are covered under the overall project management)
Constructing a safety argument to show how the individual evidence combines to prove that the project, as a whole, is acceptably safe	Prevention of hacking/ cyber-attacks (covered by a separate work package, although the safety work does consider ability of the safety driver to mitigate against suboptimal paths caused by hacking)

Table 1: summary of what is in and out of scope for HumanDrive safety management

2. Achieving 'Acceptable' Safety

2.1 Health and Safety Executive Guidance

The aim of this safety plan is to ensure that risks are kept at an acceptable level. The guidance from the UK Health and Safety Executive (HSE, 2001) is that a risk to human life of 1 in 1,000 per year for an employee, or 1 in 10,000 per year for a member of the public who have a risk imposed upon them 'in the wider interest of society', is the threshold at which the activity presents an 'unacceptable' risk, i.e. any risk at or above this level would be unacceptable. A risk to human life of 1 in 1,000,000 per year represents the threshold for risks to be 'broadly acceptable', and between these thresholds lies the 'tolerable risk region', where risks can be accepted as long as they are deemed to be 'ALARP' (As Low As Reasonably Practicable), meaning that the cost of further risk reduction would be 'grossly disproportionate' to the cost of the risk. An illustration of the different risk regions is shown in Figure 1.

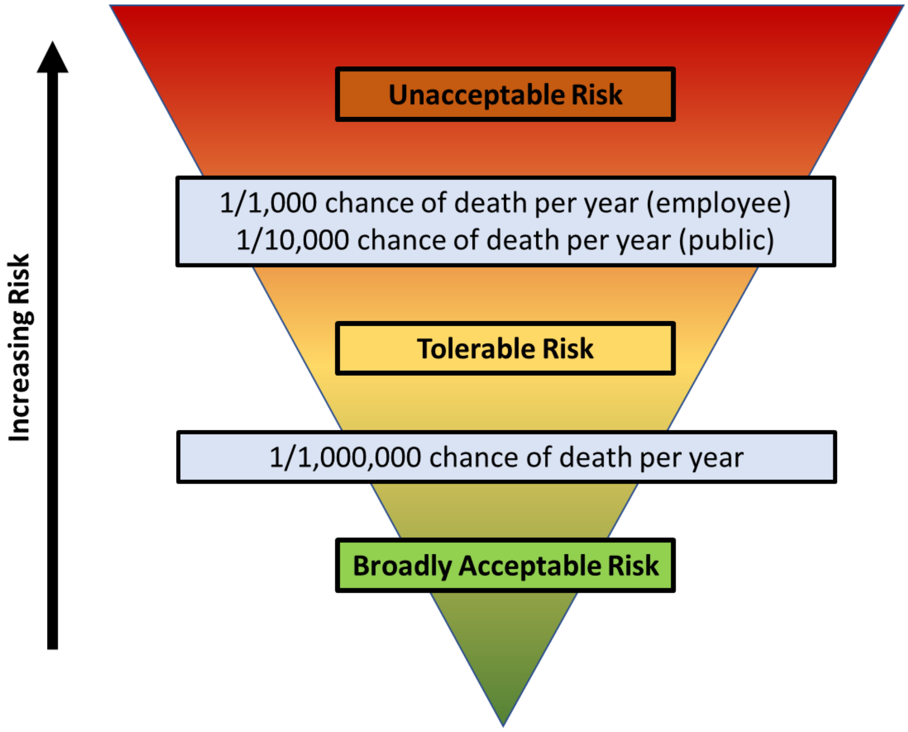

Increasing Risk
Unacceptable Risk
1/1,000 chance of death per year (employee)
1/10,000 chance of death per year (public)
Tolerable Risk
1/1,000,000 chance of death per year
Broadly Acceptable Risk

Figure 1 : Illustration of the risk 'regions' described by the UK Health and Safety Executive

The initial estimate of road traffic fatalities in the UK for 2017 is 1793 (DfT, 2018), versus a population of around 66 million (ONS, 2016). Working on the assumption that the number of people in the population who are not road users (i.e. not car drivers, passengers, pedestrians or cyclists etc.) at any point in the year is negligible, this gives a probability of death for each person in a given year of 1 in 38,000. This is towards the higher risk end of the tolerable risk region, especially bearing in mind that many of the fatalities are to people who have had the risk imposed on them (where the intolerable limit would be 1 in 10,000 rather than 1 in 1,000) and bearing in mind the extremely high number of people categorised as seriously injured (27,010 for the same period).

As shown in chapter 5.3, the overall safety of the project is expected to represent only a modest improvement in overall safety relative to the average manually-driven vehicle, and therefore the project is not expected to enter either the 'broadly acceptable' or the 'unacceptable' regions described above; it is therefore required that the project is conducted in line with ALARP principles.

2.2 Qualitative rather than Quantitative Approach

A Quantitative Risk Assessment approach (Dearden, 2016) was not deemed appropriate for the HumanDrive project for the following reasons:

- Software safety is difficult to model probabilistically as failures are *systematic* due to a flaw in the specification or coding, rather than being caused by *random* failure as would be expected for mechanical parts (e.g. Gaussian distribution of number of failures versus age of part);
- Similarly, the performance of humans interacting with the system is difficult (although by no means impossible) to predict quantitatively, meaning that a disproportionate level of research would be needed;
- Although it would be possible to identify failure rates of components, in practice this would be disproportionate given the R&D nature of the project, and would also be of limited value as limitations of system performance when no faults are present (as covered by SOTIF) are expected to be responsible for the majority of errors.

HumanDrive therefore used qualitative risk assessment tools (e.g. Risk Matrix, Failure Modes and Effects Analysis) to prioritise risks.

2.3 GAMAB

An alternative to the UK concept of tolerable risk and ALARP described above is an approach derived from French safety practice known as GAMAB (Globalement Au Moins Aussi Bon, which translates as 'globally at least as good'), sometimes referred to as GAME (Globalement Au Moins Équivalent, or 'globally at least equivalent'), which works on the basis that the overall risk presented by a new system is acceptable if it can be shown to be as low or lower than the overall risk presented by the previous system(s). This approach has recently gained wider exposure as it is referenced in ISO/PAS 21448 Road Vehicles- Safety of the Intended Functionality (2019), which is targeted at Advanced Driver Assistance Systems (ADAS) in production vehicles, and is therefore partially relevant to HumanDrive, though not directly applicable.

There are some logical flaws with this approach being used in isolation: the safety would be deemed acceptable if a very unsafe system was replaced by moderately unsafe one, but would not be deemed acceptable if an extremely safe system was replaced by one that is only marginally less safe. As such, this approach could be seen as rewarding low past performance in safety, whilst penalising high past performance.

It was therefore decided that a comparison to past systems as a justification of acceptable safety would only be used if there is evidence of widespread public acceptance of previous risk (for example, it may be inferred from widespread car use that the public view the risks of road travel as being justified by the rewards, although individual vehicle types or driving behaviours presenting particular risks may not have widespread public acceptance) and there is confidence that the old and new systems are directly comparable (e.g. an evolution of an existing system). In particular, this approach is applied in the review of how HumanDrive affects the list of hazards that Highways England have identified, as described in chapter 5.3.

3. Description of Trial Activities

Table 2 shows the level of safety analysis required for each test location and activity. HumanDrive defines static testing as testing where there are no moving objects in the environment (i.e. the HumanDrive vehicle does move itself, but it only has to navigate the road infrastructure, and doesn't have to react to other road users), and dynamic testing as testing where there are (or could be) moving objects that the vehicle has to react to.

		Test location		
		MUEAVI Test Track	**Horiba MIRA Proving Ground**	**Public Road**
Activity	**Manual driving: Data gathering**	Minimal Safety Analysis Risk comparable to normal driving	Minimal Safety Analysis Risk comparable to normal driving. Use MIRA safety procedures	Minimal Safety Analysis Risk comparable to normal driving
	Automated: Static obstacle testing	Moderate Safety Analysis Relatively immature AV, but in controlled environment	N/A Not in project plan	N/A Only possible by closing roads – not in scope
	Automated: Dynamic obstacle testing	Moderate Safety Analysis Relatively immature AV, but in controlled environment	Moderate Safety Analysis Relatively immature AV, but in controlled environment. Use MIRA safety procedures	Detailed Safety Analysis Partially mature AV operated in environment with limited controls
	Automated: Grand drive	N/A Test track not in Grand Drive route	N/A Test track not in grand drive route	Detailed Safety Analysis Partially mature AV operated in environment with limited controls

Table 2: Matrix of Test Types and Locations indicating the level of safety analysis required in proportion to the risk

The level of detail involved in the safety analysis should be proportional to the expected risk: lower risk activities are marked as minimal safety analysis, since the risk is similar to manual driving, but when the vehicle is controlled by a relatively immature automated driving system operating in a controlled environment, the detail required in the safety analysis increases, and when that system is deployed on public roads the detail of the safety analysis increases further still.

4. Safety Argument

4.1 Top Level Safety Argument

The safety argument consists of two aspects, the system safety (i.e. confirmation that the autonomous system performs in an acceptably safe way) and the operational safety (i.e. confirmation that the system is deployed in an acceptably safe way). In practice, the balance between system safety and operational safety is controlled by the Operational Design Domain (ODD), which is the system boundary defined by the road types, weather conditions, locations or traffic scenarios etc. that the system is designed to work within.

For example, if responding to a horse and rider is within the ODD, then this would be covered by system safety, and it would be necessary to define what constitutes acceptable behaviour around a horse and rider and ensure that the system complies. However, if a horse and rider is not within the ODD, then this would be controlled by operational safety, with measures put in place to remove the risk (e.g. ensuring it isn't possible for a horse and rider to access a test track, or the safety driver taking manual control of the vehicle as soon as a horse and rider are encountered on public roads, before the system has a chance to respond).

Many items of evidence are required to demonstrate the overall project safety, but ultimately, they can be viewed as building blocks of four main 'pillars' upon which the safety case is built:

- **Verification** – safety requirements are created and the vehicle is tested to confirm it meets them;
- **Validation** – the safety performance is monitored as mileage is accumulated;
- **Safe Working Practices** – method statements and risk assessments for each test;
- **Safety Driver Intervention** – it is ensured that the driver is able to override when appropriate.

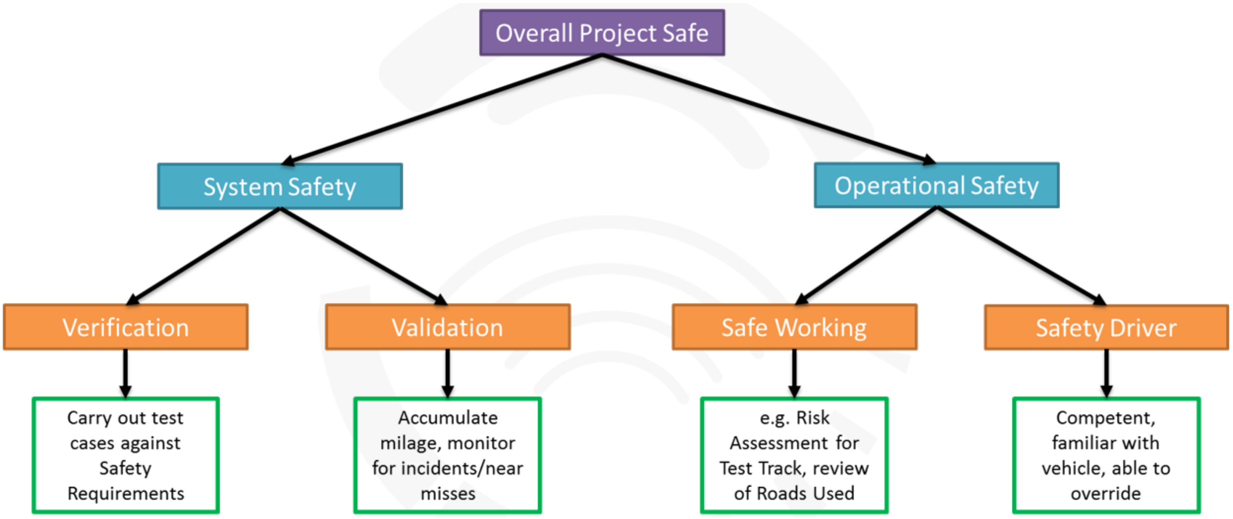

Overall Project Safe
System Safety
Operational Safety
Verification
Validation
Safe Working
Safety Driver
Carry out test cases against Safety Requirements
Accumulate milage, monitor for incidents/near misses
e.g. Risk Assessment for Test Track, review of Roads Used
Competent, familiar with vehicle, able to override

Figure 2: Illustration summarising the Safety Argument

This top-level safety argument is shown in Figure 2, which is a simplified version of the Goal Structuring Notation (GSN) model of the safety case constructed for HumanDrive, and only loosely applies the GSN syntax in order to provide an easily accessible illustration for those unfamiliar with this method of graphical modelling. GSN proved to be useful for structuring the safety argument and organising data collection; for a compressive guide, see the GSN Community Standard (2011). More information about what scenarios would need to be covered to form a safety case for a highly-automated production vehicle, including extensive use of GSN models, can be found in the TSC report 'Taxonomy of Scenarios for Automated Driving' (TSC, 2017).

Because of the complexity of the HumanDrive safety argument, separate sub-models of each the four 'pillars' in Figure 2 were created to capture the detail; this 'nesting' improved readability relative to attempting to convey the entire safety argument in one diagram. The safety arguments making up the four separate pillars are described in the following sections.

4.2 Verification Safety Argument

Verification refers to testing the vehicle against a specific set of test cases. The test cases for safety verification are derived from the safety requirements, and must ensure complete coverage of all safety requirements (it is permissible for a single test case to cover multiple requirements, or for multiple test cases to be needed to cover a single requirement).

These requirements needed a significant time investment to ensure that they are suitably comprehensive, and they were derived from the following sources:

- A Failure Modes and Effects Analysis (FMEA) of the HumanDrive Functional Architecture (chapter 5.1);
- A Hazard Analysis and Risk Assessment (HARA) for the system when deployed on the test route (chapter 5.2);
- A review of Highways England's list of hazards (chapter 5.3);
- An item-by-item review of relevant regulations, codes and standards (chapter 5.4).

The advantage of verification relative to validation is it allows test resources to be targeted at specific test cases likely to uncover significant safety compromises, e.g. by reducing the number of test cases that effectively challenge the system in the same way (Equivalence Partitioning), focusing tests on the system limits (boundary testing) and ensuring test coverage of each aspect of the system is proportionate to the risk presented.

Note that for a project that isn't based on a Type-Approved base vehicle, requirements and test cases would be needed for the primary and secondary safety of the mechanical parts and electronic/ electromechanical systems of the vehicle itself. However, this was not applicable to HumanDrive due to the use of a production Nissan Leaf as the base vehicle, and therefore oversight of the safety of the vehicle is limited to a workshop procedure to visually inspect and sign off modifications (e.g. ensuring equipment is securely attached, that nothing is mounted where it could contact and occupant, a pedestrian or an airbag, or ensuring that modifications do not invalidate the base vehicle safety case).

4.3 Validation Safety Argument

Validation refers to testing a system in a realistic environment over extended durations and monitoring any issues observed such as error states or hazards that were encountered, or performance that feels subjectively poor to the user.

The advantage of validation relative to verification is that although it is not able to focus on system limitations and faults that were foreseen during the requirement generation process ('known unknowns'), it is able to capture issues that were entirely unforeseen and not targeted by the requirements and test cases ('unknown unknowns'), and is therefore a vital filter to capture risks that would otherwise 'escape' from the drawing board to the road.

Figure 3 shows how evidence of safety will be built up throughout the project. The first step, simulation testing, was used to give basic assurance that the AI controller was able to produce reasonable control outputs prior to physical testing commencing – note that for a production vehicle aiming to achieve high levels of automation without requiring safety driver oversight, it is expected that far larger volumes of simulation would be required to gain adequate assurance that the vehicle is able to operate safely (TSC, 2018), with this simulation testing continuing in conjunction with physical testing. However, as HumanDrive vehicles will not operate without a safety driver, this level of simulation would be disproportionate given the scale of the project, and therefore the simulation was merely used as a 'shakedown'.

Next, the vehicle was tested in a static environment, i.e. with no moving objects present, to develop and validate the ability of the vehicle to navigate the roadway, including bends and a roundabout; this was carried out on Cranfield University's Multi User Environment for Autonomous Vehicle Innovation (MUEAVI) test track, with no other road users in the vicinity. The following stage is to introduce dynamic scenarios (i.e. the vehicle having to react to the movements of other road users), again in a controlled environment (at MUEAVI and at Horiba MIRA). The system is required to demonstrate satisfactory performance in these environments before testing progresses to public roads, where the environment will be uncontrolled.

The culmination of the project will be a 'Grand Drive' from Cranfield to Sunderland, but the entire route will have been covered by the public road testing beforehand, and as such, although the Grand Drive is the headline event, the major challenge for the safety case is having suitable safety evidence in place for the testing on public roads, as this is the point where the vehicle will first be exposed to the potential unpredictability of the real world; by the time the Grand Drive is undertaken, the vehicle will have already accumulated significant mileage on the route, meaning the risks are lower, although it must still be remembered that unexpected scenarios could arise that haven't previously been encountered by the vehicle.

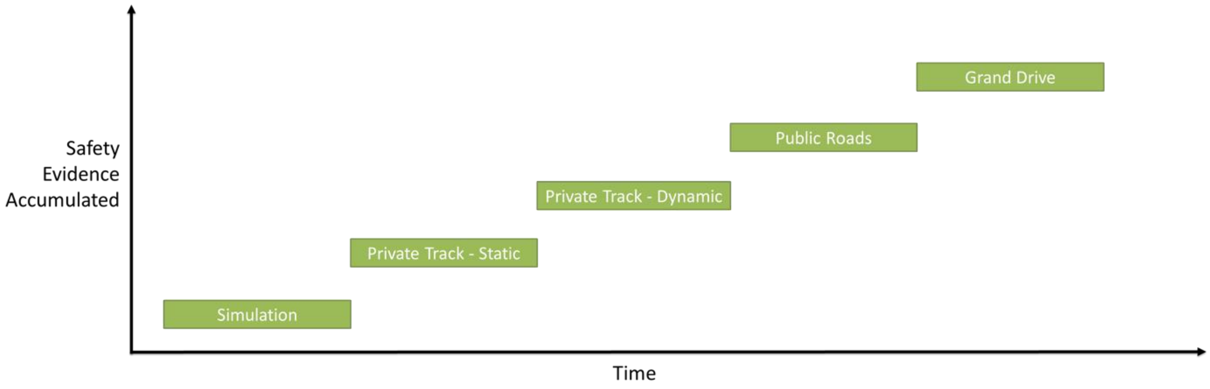


Figure 3: Illustration of how evidence of safety will be accumulated as the trials progress to more challenging scenarios

Parallel to this, an incident reporting spreadsheet will be maintained to keep an up to date log of issues encountered during any phase of testing. This will capture near misses as well as any actual incidents, as the former may be expected to be more likely and therefore a more useful source of data. The incident reporting process and format may be seen in Appendix 3.

For a production AV, extremely high test mileages would be required to gain adequate assurance that the vehicle provides an acceptable level of safety, as the number of miles travelled per incident needs to be extremely high if there is no safety driver present.

4.4 Safe Working Practices

Risk assessments are being created for every stage of testing; these are completed in line with standard practice, using a probability versus severity matrix to prioritise risks, and listing additional mitigation measures where appropriate. These risk assessments are incorporated into method statements, which are used to describe the scope of that particular stage of testing, what the roles and responsibilities are with respect to safety for individuals involved (e.g. leadership on the day, marshalling, first aid), and what steps must be carried out within the testing to ensure adequate safety.

Examples of such measures adopted for HumanDrive include the presence of an engineer in the AV in addition to the safety driver (to undertake non-driving tasks such monitoring the systems via computer screens and inputting data, so that the safety driver isn't distracted, and to monitor safety driver attentiveness), presence of marshals when testing at Cranfield University's MUEAVI test track to prevent public access, or checking the vehicle level of electric charge prior to commencing a journey to reduce the risk of the vehicle running out of charge mid-journey.

Operational safety requirements were generated during the review of regulations, codes and standards described in chapter 5.4; each requirement applicable to a given stage of testing was included in a table within the method statement for that stage, to ensure visibility to all personnel involved in the test.

For the on-road testing, particular attention was paid to identifying hazards that might be presented by the routes themselves, such as particularly narrow sections where the driver may not have sufficient time to make steering interventions before the vehicle leaves its lane, areas where visibility is limited etc. The approach taken was to undertake a desk-based review of the route with particular attention paid to hazards such as narrow lanes, tight bends, obstacles immediately next to the running lane, slip roads, etc.

The most challenging section of the Grand Drive involves travelling along country roads between Cranfield and the M1, where the roads are narrow with many bends, crests and potholes to contend with. This section was therefore subjected to a more detailed review, with a survey-quality lidar scan being commissioned and the data processed using ArcGIS software to extract dimensions from the scan such as road width (recorded at 5 metre intervals along the route), proximity of obstacles etc. Areas with a particularly narrow width were then further investigated verify the correctness of the measurements and to record details within a report to be used by consortium partners, as shown in Figure 4.

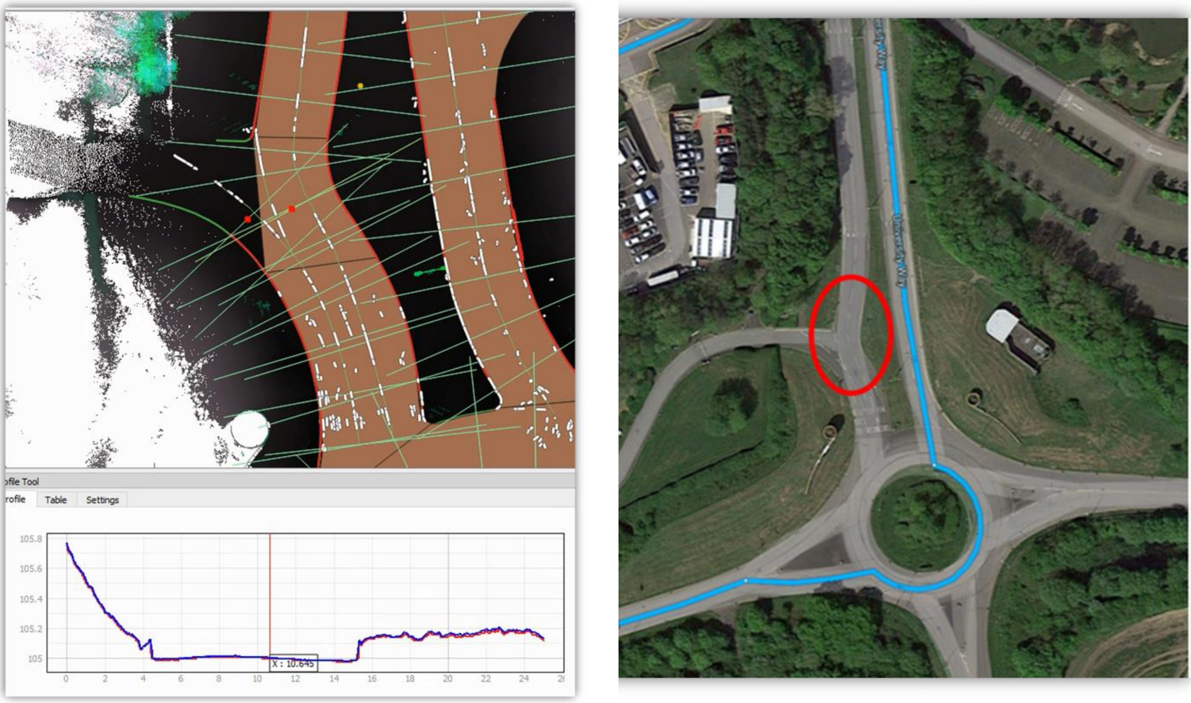


Figure 4: Analysis of the route using ArcGIS software to process the lidar scan (left), with cross sections where width was measured on the map view and one of the cross sections selected for review below, and the location marked on a map image (right)

In particular, the data on road widths proved extremely useful, as it allowed the qualitative impression of a section of road appearing to be narrow to be converted into quantitative data on the available space (as illustrated in Figure 5) for comparison by NTCE to the data on how much space the safety driver requires to intervene (collected on a test track by injecting the maximum possible steering error in a number of scenarios). Any sections of road that are too narrow for the safety driver to be able to reliably intervene can then be mitigated against by the safety driver taking control pre-emptively at this point, effectively removing it from the ODD, although in some cases this would only be necessary if there is other traffic present at the time.

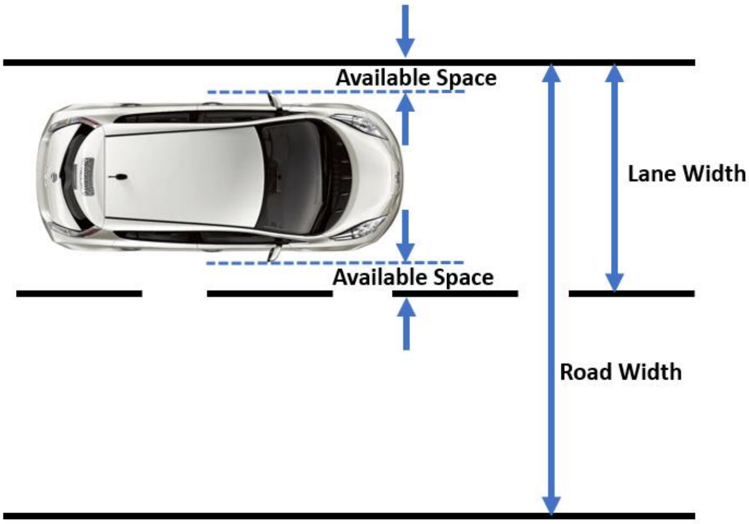


Figure 5: Illustration of what is meant by 'available space', compared to road width and lane width

This analysis works on the assumption that the HumanDrive vehicle and any other nearby vehicles are travelling approximately in the centre of the lane. However, if there is an oncoming or overtaking vehicle encroaching into the HumanDrive vehicle's lane (e.g. a wide load, a lane-splitting motorcyclist or a car cutting a corner), the safety driver should take manual control immediately, as the above assumption becomes invalid and safety can no longer be assured. In the case where the HumanDrive vehicle or other vehicles deviate significantly from the lane centre whilst still staying within lane (e.g. corner cutting), the safety driver must exercise judgement as to whether the available space has been reduced below the level required to correct steering errors, and take manual control where appropriate.

4.5 Ability of the Safety Driver to intervene

As an R&D project, it is reasonable to expect that the HumanDrive vehicle will make sub-optimal calculations in its path or speed; the presence of a safety driver who is ready and able to intervene to correct any such mistakes must therefore be the crux of the safety case. The following evidence of the ability of the safety driver to intervene was required for HumanDrive:

- All safety drivers must be NTCE employees who hold the highest driving qualification available in the company, covering on-the-limit handling to ensure the vehicle can be corrected in emergencies. They must also hold Nissan qualifications for driving to conduct an engineering evaluation and driving for business (the latter qualifications being specific to driving on UK public roads, ensuring familiarity with the relevant rules and regulations);
- All safety drivers must be familiar with the Operational Design Domain (ODD) of the HumanDrive system. For example, if responding safely to the presence of a horse and rider is not within the ODD, the safety driver needs to know that they would be required to intervene as soon as the vehicle encounters this scenario;
- All safety drivers must gain extensive familiarity with the HumanDrive vehicle through training on private tracks before being responsible for the vehicle on public roads. This includes reacting to deliberately injected faults to ensure that the driver is able to intervene appropriately within the available space (as identified in section 4.4).

It should be noted that the assessment of the ability of the safety driver to intervene in response to an incorrect path or velocity, or when requested to do so by the system, involves a two-way interaction; how quickly a safety driver can correct a particular error is not just a function of the ability of the safety driver, but also a function of the quality of the Human-Machine Interface (HMI), e.g. how clear warnings requiring the driver to take over are, or how easy it is to input sufficient steering torque to override the torque input from the autonomous system.

The task of testing safety driver responses can be made significantly easier by applying limits to control requests that can be passed to the vehicle from the autonomous control system, such as the acceleration, brake pressure, steering torque or steering rate; any requests above the limit would result in the actuator applying this specified maximum, and no more. This ensures that manoeuvres are kept within reasonable bounds, thereby reducing the ability of system errors to cause uncontrollable scenarios, and is an approach that is in widespread use for production systems such as Lane Keep Assist (ISO 26262 Road Vehicles – Functional Safety, 2011), with the limit values typically varying as a function of speed (for example, high steering angular rates that are appropriate at parking speeds would be inappropriate at motorway speeds).

By limiting the possible errors that can be made at vehicle-level, it is possible to have a well-defined worst-case scenario in which to test safety driver responses. The fault injection testing should therefore use the worst-case for each type of fault, applied over the speed range, with acceptable performance in the worst-case demonstrating that faults below the worst-case can also be managed acceptably, without requiring further testing.

If the limiting is done by production level subsystems, these can be taken to have a high level of robustness relative to an R&D autonomous system providing inputs; however, if no production subsystem is available, a bespoke subsystem can be made robust with a thorough development process, including verification and validation targeted at the limiting functionality. Although this introduces a significant engineering overhead, it will be far more practicable than attempting to achieve the same level of robustness higher upstream within the autonomous system. The safety driver cannot be argued to be a suitable backup system unless it can be demonstrated that they can consistently intervene whenever required, and evidence of manoeuvres being limited to reasonable values forms a key component of demonstrating controllability.

It is also important to include a means for the safety driver to quickly deactivate the system with a single, easily-performed action (e.g. press of a button that is suitably prominent and conveniently located), so that they can take manual control of the vehicle in response to an error, a notification of a fault, or the vehicle going outside the ODD. This would typically include a ‘soft’ switch to cause the software to transition into a standby mode, but should also include a ‘hard’ switch that breaks the electrical connection to the autonomous system such that the vehicle cannot become stuck in autonomous mode due to a software error. The ability to deactivate the system quickly should be thoroughly verified through testing.

In recognition of the possibility that safety driver performance could reduce over time as a result of complacency (particularly if the system goes along time without adopting any sub-optimal paths) or fatigue (bearing in mind the challenge of maintaining concentration when monitoring a system as opposed to actively controlling it), there will always be an engineer on board the vehicle in addition to the safety driver, who must ensure the safety driver remains engaged and ready, and regular breaks and driver changes will be enforced.

4.6 The Importance of the Safety Driver

Figures 6 and 7 illustrates how critical the safety driver is to the safety of the project. Figure 6 shows a representation of a production AV, with the size of the arrows representing the frequency of hazardous occurrences. Production AVs, once available, will have to be developed to rigorous safety standards, and will therefore produce a very low number of errors (depicted by a narrow arrow coming from ‘system error’). For a hazardous event to occur, a hazardous environment will also need to be present, e.g. an oncoming vehicle in the area that the AV is steering towards; this is represented as a medium arrow, as sometimes physical hazards will be present, but sometimes they will not. Furthermore, the driver may manage to correct some errors, even though they are not specifically trained to intervene or required to monitor the driving task. As the system errors are low, not all occur in hazardous environments, and some may be filtered out by the safety driver, hazardous events will therefore be rare (very narrow arrow)

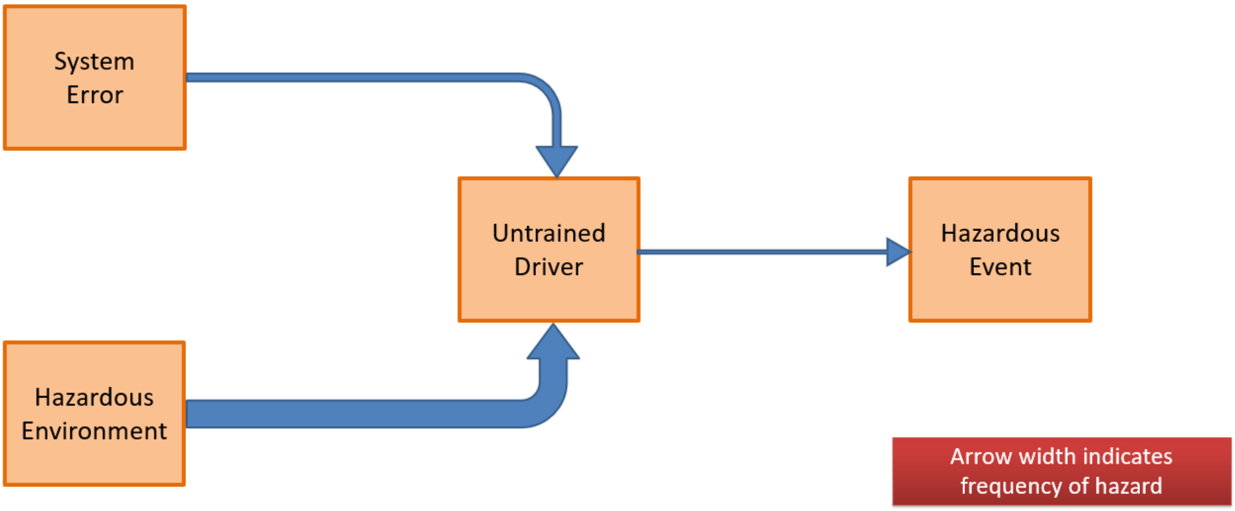


Figure 6: Schematic illustration of the frequency of hazards for a production autonomous system

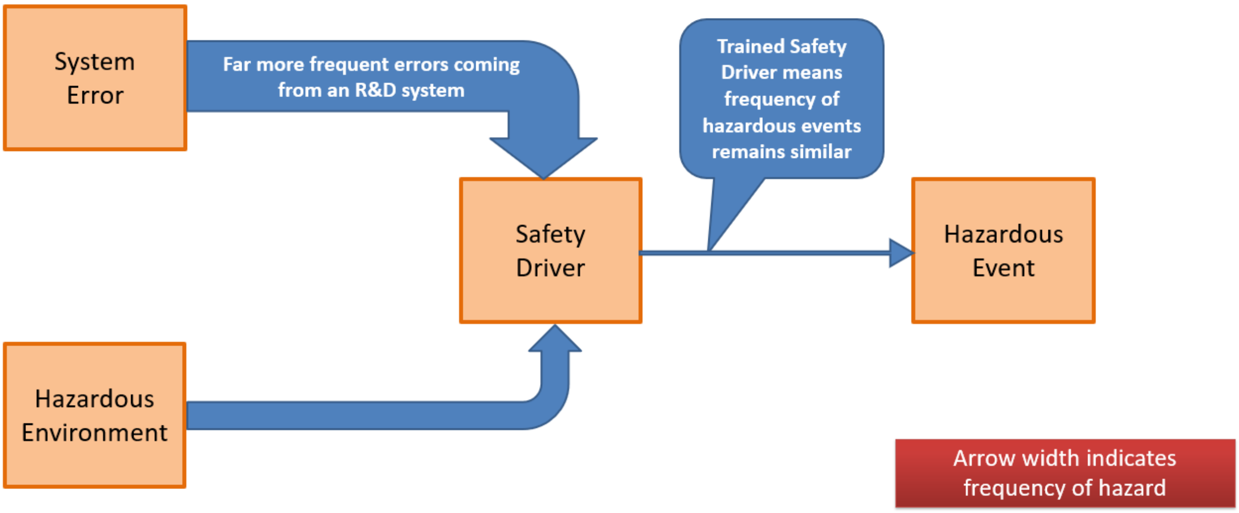

System Error
Far more frequent errors coming from an R&D system
Trained Safety Driver means frequency of hazardous events remains similar
Safety Driver
Hazardous Event
Hazardous Environment
Arrow width indicates frequency of hazard

Figure 7: Schematic illustration of the frequency of hazards for an R&D/ Prototype autonomous system

For an R&D autonomous system (Figure 7), or indeed for initial prototypes of a production AV before it is fully developed, the number of errors may be expected to be significantly larger (this worst-case should be assumed until proven otherwise). This is represented as a wide arrow. The frequency of hazardous environments is assumed to remain the same for public road testing (although note that this can be reduced significantly where needed in the early stages of testing, e.g. by using a private track). As a result of the increased likelihood of a system error, the probability of the safety driver being required to prevent an incident is significantly higher than for a production vehicle. The only way to mitigate this is to have a highly-trained safety driver who is very familiar with the vehicle characteristics and remains attentive at all times; the safety driver therefore forms the backup safety 'system' that ensures the arrow leading to a 'hazardous event' remains narrow.

5. Safety Documentation

5.1 FMEA

Failure Modes and Effects Analysis (FMEA) is a widely-used technique within safety engineering. The architecture of a system is evaluated, considering each possible fault that could occur within the system and identifying how that fault will affect the rest of the system, including any failures or other undesirable behaviours that the fault could cause the system as a whole to display. This is referred to as a 'bottom-up' or 'inductive' approach, as it starts with the fault (at a low level) and works up to the effect on the high-level behaviour of the system, as opposed to 'top-down' or 'deductive' analysis (e.g. Fault Tree Analysis), which starts at the high-level failure and works downward to identify the fault(s) that form the root cause.

In the case of HumanDrive, a functional architecture diagram was developed, identifying the subsystems that make up the overall system and the sub-subsystems that make up each subsystem. A workshop was then held with the project consortium, where participants were asked to identify what possible failures could occur in each sub-subsystem, what the local effect would be on the subsystem in which it resides, and whether this would result in any undesired vehicle-level behaviours, with the output from this, plus follow up work to close off open points raised during the workshop, allowing the FMEA to be completed. The analysis was limited to single point failures (i.e. failure modes that only require a single fault to occur) other than in the case where faults can remain latent (i.e. not detected by the system or any personnel); as testing will cease as soon as a fault is identified, multiple point failures where faults do not remain latent (therefore requiring two independent faults to occur almost simultaneously) can be considered highly unlikely.

It was decided not to apply severity and likelihood scores to each failure (a pseudo-quantitative approach commonly referred to as Failure Modes, Effects and Criticality Analysis, or FMECA) due to the difficulty in estimating reasonable scores; each failure typically causes a low chance of a high severity accident, a high chance of a low severity accident, plus an infinite number of variations on the spectrum between these extremes, making selection of representative values difficult. Furthermore, it was unclear how such scores would be of practical benefit, as it was decided early in the project that it would be unfeasible and disproportionate to apply ASIL (Automotive Safety Integrity Level) ratings to components and functions in the manner of ISO26262, given the R&D nature of the project.

Safety goals (i.e. high-level requirements for what should be done to mitigate hazards) were logged for each possible fault that was identified, and these were incorporated into the project requirements management spreadsheet, with corresponding test cases and acceptance criteria then being added. Furthermore, the list of possible vehicle-level hazards (i.e. undesired behaviour such as false-positive braking, sudden loss of autonomous control etc.) was collated as an input to the HARA (section 5.2).

Although it is not possible to share the HumanDrive FMEA, due to the commercially-sensitive nature of the information it contains, a template version of the document, with some generic examples to illustrate the typical data that would be entered, has been included in Appendix 1.

It is worth noting that regardless of the analysis done in the FMEA, and any resulting work to improve robustness against failure or ensure that the system fails in a safe way, it must still be expected that sub-optimal outputs will be produced in an R&D vehicle, and therefore operational safety measures such as the use of a safety driver or a controlled environment form the crux of the safety case. However, the process of generating the functional architecture diagram and the FMEA proved highly productive in terms of improving the understanding between the stakeholders of how the system as a whole functions. Furthermore, although the safety driver is able to prevent safety-related incidents, nonetheless regular errors by the system are undesirable if a convincing demonstration of the technology is to be presented, and efforts must therefore be made to evolve system safety if the ultimate aim of fully-autonomous vehicles is to be realised.

5.2 HARA

The hazards identified by the FMEA, together with further hazards identified through prior experience and knowledge of the system limits, were fed into a HARA (Hazard Analysis and Risk Assessment) to analyse exposure to risk from the perspective of system safety (a risk assessment from an operational safety perspective was created for each trial as described in chapter 4.4). Three separate risk assessments were provided for the list of same hazards, depending on whether the vehicle is deployed on a private track, local roads or Highways England's Strategic Road Network (SRN).

The probability scoring is based on both the probability of the hazard (e.g. incorrect steering input) occurring **AND** the probability that other traffic and obstacles around the vehicle poses a threat at the moment that the hazard is triggered such that it is reasonably possible for an incident to result. This is multiplied by the score for severity of the resulting incident, and the controllability (a low score indicating that the driver has a good chance of being able to prevent or mitigate the incident, i.e. the scoring reflects the difficulty of intervening successfully), to obtain an overall risk score. A schematic of the approach is shown in figure 8. Thresholds were then introduced to define the boundaries between different risk categories, as illustrated in figure 9, with the boundaries being set by experimentation to arrive at thresholds that produce reasonable results.

It should be noted that the severity scores are based on the most serious reasonably-foreseeable outcome, and therefore the numbers awarded may appear conservative to those used to scoring systems that score the severity based on the most likely outcome. The rationale for the approach adopted is that, given the protection provided by modern passive safety provisions such as survival cells and airbags, it is entirely plausible that if, in a manual vehicle, a driver takes their hands off the wheel, depresses the accelerator and lets the vehicle continue until it hits something, the most likely outcome may still not be a fatality, particularly in the case of the HumanDrive Grand Drive route which primarily consists of roads where the population of vulnerable road users (cyclists, pedestrians etc.) tends to be relatively sparse. It is therefore not feasible to base the scores on the most likely outcome as it would mean it was impossible, even with the most dangerous failures of the system, to reach the top severity level. This approach is adapted from and broadly consistent with ISO26262.

In line with Health and Safety Executive (HSE) Guidance, risks were rated as unacceptable, tolerable or acceptable. However, the HSE approach advises that where risks are tolerable, they should be demonstrated to be ALARP (As Low As Reasonably Practicable), assessed with techniques such as comparing the cost of fatalities to the cost per life saved of mitigation measures. For an R&D project, such an analysis isn't proportionate, so it was deemed that a better approach to judging whether 'reasonably practicable' steps have been taken was to separate tolerable risks into an upper (amber) and lower (yellow) band, with a higher burden of mitigation in the upper band.

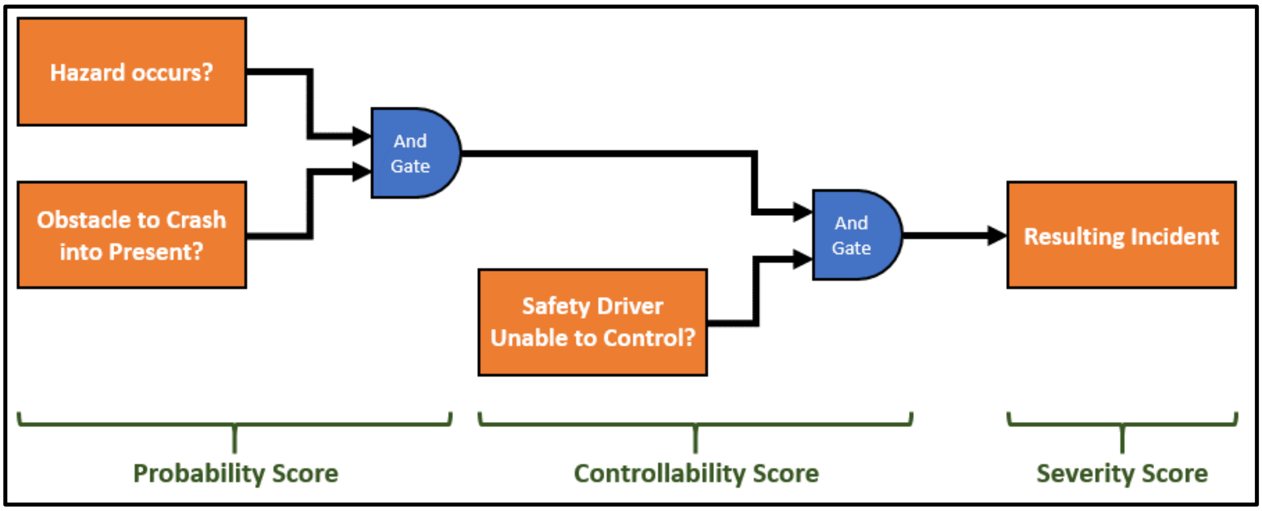
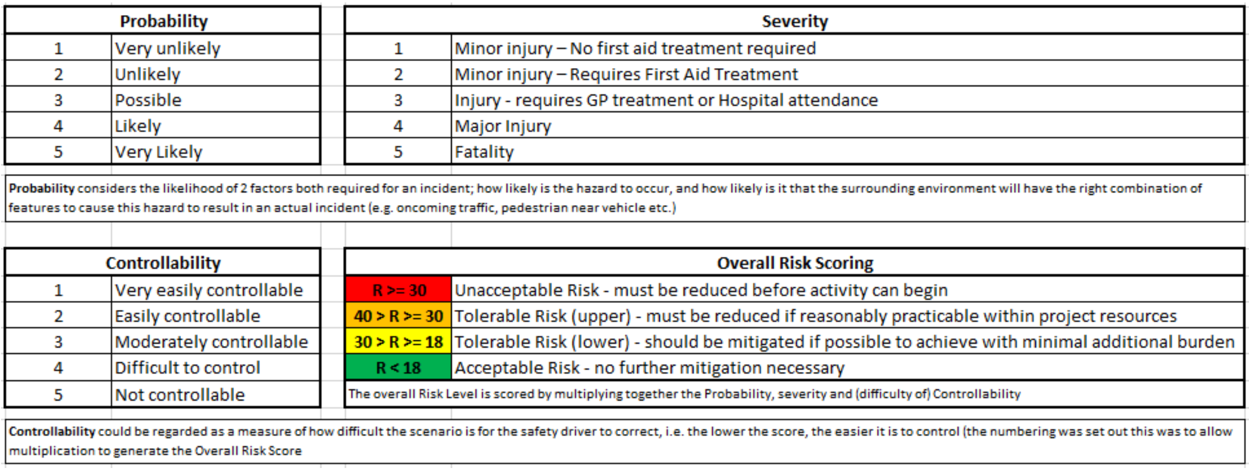


Figure 8: Schematic illustration of how the Probability, Severity and Controllability contribute to the overall risk within the HARA methodology



Probability	
1	Very unlikely
2	Unlikely
3	Possible
4	Likely
5	Very Likely

Severity	
1	Minor injury – No first aid treatment required
2	Minor injury – Requires First Aid Treatment
3	Injury - requires GP treatment or Hospital attendance
4	Major Injury
5	Fatality

Probability considers the likelihood of 2 factors both required for an incident; how likely is the hazard to occur, and how likely is it that the surrounding environment will have the right combination of features to cause this hazard to result in an actual incident (e.g. oncoming traffic, pedestrian near vehicle etc.)

Controllability	
1	Very easily controllable
2	Easily controllable
3	Moderately controllable
4	Difficult to control
5	Not controllable

Overall Risk Scoring	
🟥 R >= 30	Unacceptable Risk - must be reduced before activity can begin
🟧 40 > R >= 30	Tolerable Risk (upper) - must be reduced if reasonably practicable within project resources
🟨 30 > R >= 18	Tolerable Risk (lower) - should be mitigated if possible to achieve with minimal additional burden
🟩 R < 18	Acceptable Risk - no further mitigation necessary
The overall Risk Level is scored by multiplying together the Probability, severity and (difficulty of) Controllability	

Controllability could be regarded as a measure of how difficult the scenario is for the safety driver to correct, i.e. the lower the score, the easier it is to control (the numbering was set out this was to allow multiplication to generate the Overall Risk Score

Figure 9: Scoring system used for the HARA

The scoring system proved useful as a means to show an approximate representation of risk and as a trigger for discussion, although care should be taken not to divert too much focus towards the scoring, as the scores and the scoring system are subjective rough estimates; the most important output of the HARA is the consideration and documentation the hazards and of suitable mitigation measures.

A generic example of a HARA using the HumanDrive template is included in Appendix 2, for the benefit of other projects wishing to re-use the approach developed for HumanDrive.

5.3 Review of Highways England’s List of Hazards

As part of HumanDrive’s work with Highways England under their GG104 risk assessment process, a review was undertaken to establish the relative risk presented by the project when compared with the risk presented by regular manually-driven traffic. This utilised a hazard log that Highways England had previously created as part of the transition from traditional ‘Dual 3-lane Motorways’ to ‘All Lane Running’ Smart Motorways, with this list being prioritised into eighteen ‘Top Level Hazards’ (those identified by

Highways England as the most serious when considering likelihood and severity) and eighty-three remaining hazards that were not identified as being top-level.

The top-level and non-top-level hazards were reviewed, with particular focus placed upon the top-level hazards as Highways England have identified that these hazards constitute the vast majority of the overall risk as assessed for existing traffic. It was deemed worthwhile to review the lower level hazards as some of these hazards may become more significant when AVs are involved, but the analysis of these lower level hazards was less detailed than for the top-level hazards, bearing in mind the need to use resources efficiently and proportionately.

The change in the level of risk presented by each hazard as a result of the presence of the HumanDrive vehicle was assessed qualitatively as low, medium or high. By applying a score of 1, 2 or 3 respectively to these categories (with risk reductions given a negative sign and risk increases positive), it was possible to sum the individual scores to provide a rough 'quasi-qualitative' estimate of the overall change in risk. A moderate overall reduction to the risk was indicated by this (both for the top-level and non-top-level hazards), despite some individual risks increasing slightly. This should only be taken to be broadly indicative, rather than a precise measure; for example, a '+1' increase in one of the most prevalent or severe risks may, in the real world, be more of a concern than a '+3' increase in a more minor risk.

The overall approach taken could be seen as an example of the French-derived GAMAB approach to risk assessment, as described in section 2.3. This was judged appropriate as the initial safety baseline has already been established by Highways England, and the scenarios are similar enough when the HumanDrive vehicles are involved to allow a direct comparison to be made, thereby ensuring that safety is not compromised by the trial relative to what is currently accepted by authorities and the general public.

5.4 Regulations, Codes and Standards Review

In order to ensure that HumanDrive is compliant with the applicable regulations, codes and standards, a review was conducted of the following:

- Pathway to Driverless Cars: a Code of Practice for Testing (2015)
- The Highway Code (2015)
- The Road Vehicles (Construction and Use) Regulations 1986 (1986)
- The Road Traffic Act 1988 (1988)

Each relevant clause was copied into a separate line of a spreadsheet, with a separate tab being used for each regulation or code. Separate columns were provided for system safety and operational safety, with each clause resulting in safety goals (i.e. high-level safety requirements) in either column as appropriate.

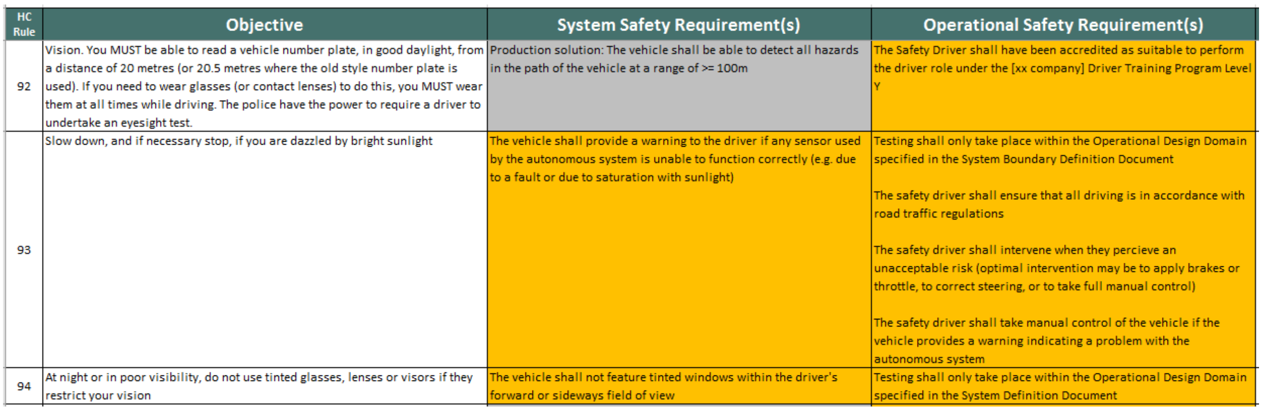
For example, if a vehicle is able to operate the indicators autonomously, there would be system requirements relating to each highway code clause that references indicators, both to ensure that the system is compliant with the Highway Code as a bare minimum, but also because the Highway Code proved to be an effective prompt for generating requirements, some fairly tangential to the original clause, that are important to capture. The only operational safety goal in this example would be a generic one for the safety driver to monitor the vehicle and correct any errors.

However, if the indicators are not operated autonomously, then the column for system requirements would be greyed-out for all Highway Code clauses related to indicator use, as the system wouldn’t be required to do anything in this regard, but there would be an operational requirement for the safety driver to activate and cancel the indicators manually. Where applicable, 'production solutions' were included in the greyed-out cells to indicate what sort of requirements may be needed in order to deploy the vehicle commercially, without a safety driver present.

Once all clauses had been addressed, the system safety requirements and operational safety requirements were each compiled into separate lists, and these were added to the requirement management document that was used for sharing and tracking of requirements between project partners. It was found to be important to keep the safety goals at a relatively high level of abstraction, so that they can be repeated to cover multiple clauses, as slight detail changes in the wording restricts the ability to reuse that same wording, resulting in the compiled list of requirements becoming unmanageably long with many near-duplicate requirements.

A particular effort was made to keep operational safety requirements that relate to how the safety driver should respond to situations at a high level; whilst it is possible to go through detailed system requirements when engineering a vehicle, there has to be a recognition that it is impractical for a safety driver to maintain an equivalent list of operational safety requirements in their head and process them real-time whilst controlling the vehicle, and it is therefore essential that the safety driver is allowed to use their judgement, experience and common-sense, rather than attempting to codify their behaviour with unrealistic requirements.

It is not possible to share the HumanDrive document as it contains commercially sensitive information, but a template version was created for sharing, populated with some generic examples, and an excerpt from this is shown in Figure 10.



HC Rule	Objective	System Safety Requirement(s)	Operational Safety Requirement(s)
92	Vision. You MUST be able to read a vehicle number plate, in good daylight, from a distance of 20 metres (or 20.5 metres where the old style number plate is used). If you need to wear glasses (or contact lenses) to do this, you MUST wear them at all times while driving. The police have the power to require a driver to undertake an eyesight test.	Production solution: The vehicle shall be able to detect all hazards in the path of the vehicle at a range of >= 100m	The Safety Driver shall have been accredited as suitable to perform the driver role under the [xx company] Driver Training Program Level Y
93	Slow down, and if necessary stop, if you are dazzled by bright sunlight	The vehicle shall provide a warning to the driver if any sensor used by the autonomous system is unable to function correctly (e.g. due to a fault or due to saturation with sunlight)	Testing shall only take place within the Operational Design Domain specified in the System Boundary Definition Document The safety driver shall ensure that all driving is in accordance with road traffic regulations The safety driver shall intervene when they percieve an unacceptable risk (optimal intervention may be to apply brakes or throttle, to correct steering, or to take full manual control) The safety driver shall take manual control of the vehicle if the vehicle provides a warning indicating a problem with the autonomous system
94	At night or in poor visibility, do not use tinted glasses, lenses or visors if they restrict your vision	The vehicle shall not feature tinted windows within the driver's forward or sideways field of view	Testing shall only take place within the Operational Design Domain specified in the System Definition Document

Figure 10: Some generic examples of how the Regs, Codes and Standards Review document template could be filled in, with system and operational safety requirements as appropriate for each Highway Code rule

Overall, this process was found to be a very effective way of generating requirements, ensuring coverage of a wide range of scenarios; the Highway Code captures many different permutations that can be encountered when driving, derived from years of experience of the operation of manually-driven vehicles on the UK's roads, and whilst it must be remembered that there are some differences when operating an AV, nonetheless the exercise of working through the clauses resulted in many requirements that otherwise may not have been conceived, thereby allowing potential issues to be logged and addressed early in the development cycle.

As may be imagined, the process did require a significant investment of person-hours, and having reviewed the Highway Code and the Code of Practice first, it was then found there were significantly diminishing returns when reviewing the Road Traffic Act and Construction & Use regulations, both because they are less scenario-based and because many of the safety requirements duplicated those already generated. It is therefore advised that organisations wishing to utilise this template tailor the selection of regulations, codes and standards for review to ensure it is appropriate and proportionate to the scope of the project; there should be no obligation to review the same list of documents as HumanDrive.

Standards such as ISO/PAS 21448 Road Vehicles – Safety of the Intended Functionality (2019) and ISO 26262 Road Vehicles – Functional Safety (2011) were used as a general reference for good practice, but it was assessed to not be proportionate to go through them item by item, as they are intended for production vehicles and therefore many clauses are not appropriate for an R&D project.

5.5 Requirements Management

Safety requirements for HumanDrive were derived from the following sources:

1. Regulations, Codes and Standards Review (separate lists for system safety requirements and operational safety requirements)
2. Safety goals generated by the FMEA and the HARA
3. Requirements generated from the GG104 Hazards Review as part of Highways England's risk management process

Requirements were shared between consortium partners using a spreadsheet, which included the following columns:

- *Requirement ID number (unique reference for tracking)*
- *Requirement type (shall, should or may)*
- *Requirement Text (main body of requirement)*
- *Requirement Cascaded to (matrix with an 'x' for each project partner that the requirement affects)*
- *Method of Verification (describes how it should be tested, including acceptance criteria)*
- *Sign-Off Gateway (i.e. when does it need to be signed off)*
- *Sign-Off Evidence (what evidence is required for sign-off)*
- *Stakeholder Agreement (to record whether all stakeholders have agreed the requirement yet)*
- *Compliance Status (green if signed off, amber if on course, red if issues need to be addressed)*
- *Change Control (increment number if requirement is updated to track changes)*
- *Origin/ Priority (is the requirement for regulations, safety, adhering to standards or a design decision?)*
- *Linked Rqt. (used to link to upstream or downstream requirements for traceability)*
- *Derived From/ Source (link to relevant documents, e.g. reports)*
- *Rationale/ Comments / Assumptions (column to capture free-form comments and explanations etc.)*

All requirement wordings were expected to be broadly in line with standard systems engineering practice, such as being clear, concise, atomic (i.e. only capturing a single point within each requirement, such that it cannot be further split) and testable, although experience has shown that it's more practical to avoid being too dogmatic and allow a certain level of flexibility in the way requirements are explained to make them readable; overly-rigid adherence to strict requirement syntax rules can result in requirements being very convoluted to read and cause a large administration overhead to fine-tune wording. The guiding principle used here, therefore, is that if all stakeholders to the requirement can reasonably be expected to understand what the wording means, the wording is acceptable.

6. Event Data Recording

A key objective of the HumanDrive safety work, in accordance with Pathway to Driverless Cars: a Code of Practice for Testing (2015), was to ensure that there is suitable event data recording such that any incident that occurs during testing can be investigated thoroughly.

TSC detailed what data need to be collected by the vehicle as part of the Event Data Recorder (EDR) functionality, and at what minimum frequency; these requirements were agreed with NTCE, who will collect the data at all times that the vehicle is driving autonomously. However, interpreting the data is not a trivial task, taking time and resource to ensure all outputs are understood and interpreted correctly.

This means that if the analysis process hadn't previously been tested using real data, there could be an unacceptable time delay between an incident occurring and the analysis being possible. Furthermore, there would be a risk that it would be realised upon commencing the analysis that some of the requested data were being recorded incorrectly, not recorded at all, or that the requirements failed to capture all the necessary information; in any of these cases, it would be possible that analysis would be significantly limited, or even rendered impossible, by the lack of data.

A trial run was therefore carried out, processing a sample of data from initial HumanDrive Trials at Cranfield University's MUEAVI test track, to ensure that it will be possible to provide a suitable analysis should an incident occur in the future. No incidents occurred within this phase of testing, meaning that the data was collected with the vehicle functioning as intended, but this was sufficient to test the data analysis process and ensure analysis of the vehicle's movements is possible.

A 'dashboard view' was created using Unity (2019), a cross platform games engine, to allow the various parameters to be visualised whilst also showing an image of the car in a virtual environment. Some example views can be seen in Figure 11 and Figure 12, where the dashboard is shown on the left and the corresponding view of the vehicle in the virtual world at that same moment in time is shown on the right.

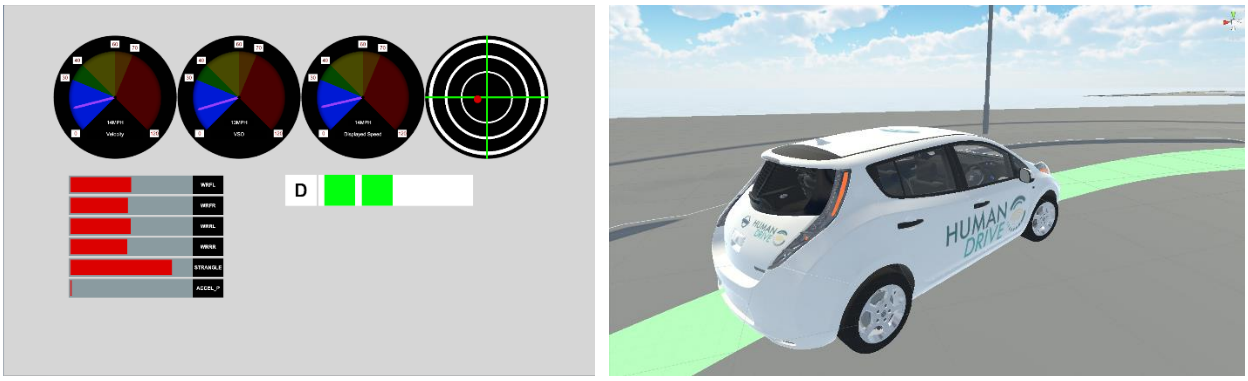


Figure 11: Negotiating the roundabout at Cranfield University's MUEAVI test track; steering right, indicated speed 14mph

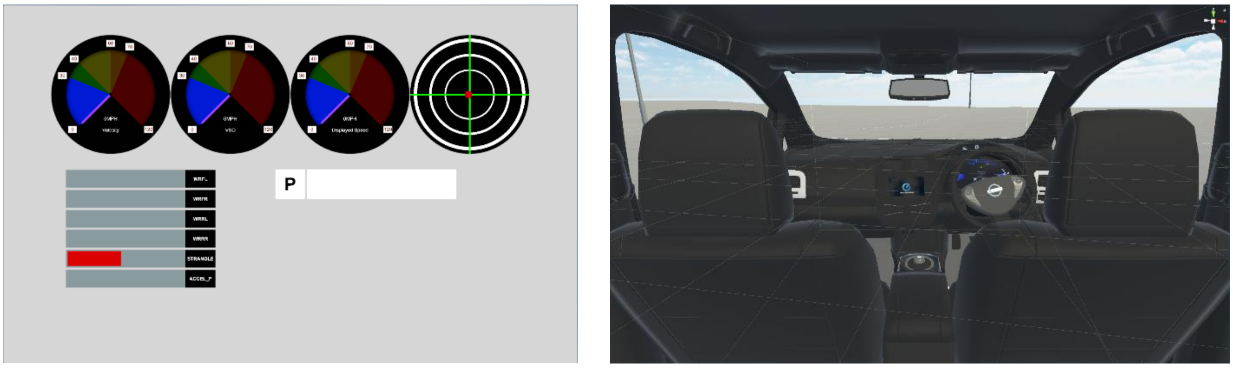


Figure 12: Vehicle stationary (speed and acceleration are zero), with wheel turned slightly left. Interior view of virtual vehicle shown to illustrate how the viewing position can be adjusted as required

This is an ongoing piece of work, as the analysis so far has only included reconstruction on an environment where there are no moving objects; as testing progresses to more challenging environments, further development of the process will take place to ensure all the relevant detail, such as the position and speed of other road users, is able to be reconstructed. However, the work undertaken so far has been successful in validating the completeness of the EDR requirements and the process for interpreting them with respect to scenarios with no moving objects present.

An incident reporting spreadsheet was created for personnel involved in trials to make a record of any incidents that may occur, including a brief description and information about the time and date; this data will be used to identify what data from the EDR should be subjected to further analysis. A process was documented detailing the reporting requirements for different levels of incident - see Appendix 3 for more information.

7. Conclusion

This paper has outlined measures designed to ensure that the HumanDrive AV trials are acceptably safe, and other projects are welcome to reuse or develop the processes and documents as appropriate. However, care must be taken to ensure that the safety measures in place are appropriate for each particular trial, and as such, the approach used in HumanDrive may not be appropriate for all trials, and it remains the duty of those conducting the trials to ensure that they carry out a thorough safety analysis to ensure testing remains in line with all regulations, standards and codes that are applicable.

It must also be noted that although there has been significant analysis of the safety of the HumanDrive system safety, nonetheless the key to the safety case is the operational safety (e.g. the ability of the safety driver to override, use of a controlled environment). For vehicles to be deployed commercially, it will be necessary to reach a level of system safety where little or no mitigation via operational safety is required over and above the rules and procedures used for existing manual vehicles. As such, while it is hoped that the HumanDrive system safety work is a step towards the development of a safety methodology that will allow commercial deployment of level 4 and level 5 AVs, it should not be seen as an indication of what a suitable safety case for such vehicles might look like.

The HumanDrive project remains ongoing, with on-road testing and the Grand Drive due to be completed later within 2019, along with the associated verification and validation evidence that will be collected. As such, the consortium will release more information as the project progresses, and interested parties can monitor progress via the project website (HumanDrive, 2019).

8. References

Code of Practice: Automated Vehicle Trialling (2019) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/776511/code-of-practice-automated-vehicle-trialling.pdf

Dearden, H. (2016) *Functional Safety in Practice,* SIS Suite.

DfT (2019) *Reported Road Casualties in Great Britain@ quarterly provisional estimates year ending September 2017,* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/681593/quarterly-estimates-july-to-september-2017.pdf

GSN Community Standard (2011) *GSN Community Standard Version 1*, http://www.goalstructuringnotation.info/documents/GSN_Standard.pdf

HSE (2001) *Reducing Risks, Protecting People. HSE's decision-making process*, http://www.hse.gov.uk/risk/theory/r2p2.pdf

HumanDrive (2019) *HumanDrive Project Website,* www.HumanDrive.co.uk

ISO/PAS 21448 Road Vehicles – Safety of the Intended Functionality (2019) https://www.iso.org/standard/70939.html

ISO 26262 Road Vehicles – Functional Safety (2011) https://www.iso.org/standard/43464.html

ONS (2016) *Population estimates for UK, England and Wales, Scotland and Northern Ireland: mid-2016,* https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/latest

Pathway to Driverless Cars: a Code of Practice for Testing (2015), https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/446316/pathway-driverless-cars.pdf

ROGS (2006) *The Railways and Other Guided Transport Systems (Safety) Regulations 2006*, https://orr.gov.uk/__data/assets/pdf_file/0018/2547/rogs-2006-consolidated-with-amendments.pdf

The Highway Code (2015), https://www.gov.uk/guidance/the-highway-code

The Road Traffic Act 1988 (1988), https://www.legislation.gov.uk/ukpga/1988/52/contents

The Road Vehicles (Construction and Use) Regulations 1986 (1986), http://www.legislation.gov.uk/uksi/1986/1078/introduction/made

TSC (2017) *Taxonomy of Scenarios for Automated Driving,* https://s3-eu-west-1.amazonaws.com/media.ts.catapult/wp-content/uploads/2017/04/25114137/ATS34-Taxonomy-of-Scenarios-for-Automated-Driving.pdf

TSC (2018) *AV Simulation Testing Report*, https://s3-eu-west-1.amazonaws.com/media.ts.catapult/wp-content/uploads/2018/03/23113301/00299_AV-Simulation-Testing-Report.pdf

Unity (2019) Unity (2019) https://unity.com/solutions/automotive-transportation?_ga=2.243859395.79586156.1552305307-232898215.1552305307

Appendix 1: FMEA Generic Example

The following is a generic example of an FMEA (Failure Modes and Effects Analysis) for an autonomous driving system, using the template developed for the HumanDrive project. It examines the possible failures in ‘Sub-System A’, provided by consortium member ‘Partner X’, with sub-sub-systems ‘A1’ and ‘A2’ (which make up Subsystem A) examined, with failure modes and how the failure will propagate around the system recorded along with control measures and ‘safety goals’ (i.e. high-level requirements related to safety). The first row (ref 1) covers failure of the entire sub-system A due to a hardware fault (in the example, it is assumed that A1 and A2 cannot suffer hardware failures independently of each other, meaning that all of A would fail in unison.

This analysis would be continued to cover all the possible failure modes for all the sub-sub-systems within every sub-system of the overall system’s Functional Architecture. Note that two columns have been omitted for clarity; one of these is for review comments, the other for version control (to indicate which version number the document was on when the row was last updated).

System Description				Failure Mode	Possible Failure Causes	Failure Effect / Safety Impact			Potential Outcome								Detection Method	Existing Controls Risk Elimination or Mitigation Measures	Safety Goal
Ref	Owner	Sub-System	Sub-sub-system			Local Level	System Level	Vehicle Level	Loss of AD Control	Unintended Braking	Unintended Accel	Unintended Steering	Lack of Braking	Lack of Accel	Lack of Steering	Driver take Control			
1	Partner X	Subsystem A	N/A - hardware failure of individual sub-subsystems within subsystem x not possible, housed on same chip	No data output - No situational awareness	Hardware failure of entire Subsystem	No message to Subsystem B	Vehicle has no path to follow Autonomous control ceases, driver to take manual control	AD system disconnects and reverts to manual driving mode Safety driver has to take control of vehicle as quickly as possible, with no prior notice	x								Visual and audible alarm when AD system disconnected	Highly trained safety driver Safety driver alert at all time and ready to take over at any point in time	• Safety driver to be provided with audio and visual alert when the vehicle transitions from AD to MD mode due to a fault • Safety driver to take manual control where they feel the error margin is insufficient to allow them to detect and correct an error in time • Safety driver able to transition instantaneously from AD to MD mode with a single button press when not satisfied with the safety of the current situation

2	Partner X	Subsystem A	Sub-Subsystem A1	Malfunction - no output or no interpretable output	Computational error etc.	No viable input to Subsystem B Diagnostics in Subsystem B recognise problem	Vehicle unable to maintain autonomous control Autonomous control ceases, driver to take manual control	AD system disconnects and reverts to manual driving mode Safety driver has to take control of vehicle as quickly as possible, with no prior notice	x								Visual and audible alarm when AD system disconnected	Highly trained safety driver Safety driver alert at all time and ready to take over at any point in time	• Safety driver to be provided with audio and visual alert when the vehicle transitions from AD to MD mode due to a fault • Safety driver to take manual control where they feel the error margin is insufficient to allow them to detect and correct an error in time • Safety driver able to transition instantaneously from AD to MD mode with a single button press when not satisfied with the safety of the current situation • The vehicle shall display the current operating mode (AD or MD) to the safety driver
3				Misrepresentation of the environment - error NOT detectable (i.e. plausible output)	Computational error etc.	Erroneous signal to Subsystem B	Diagnostic system unaware of potential issue, meaning error will be passed to the low-level controller	Vehicle remains in AD mode and attempts to adopt improper trajectory or speed		x	x	x	x	x	x		Safety driver identifies that vehicle has deviated from a reasonable trajectory/ velocity	Safety driver able to make manual control inputs to override the autonomous system	• Detection of driver intervention shall be ensured such that the autonomous system does not fight against manual interventions by the safety driver • Safety driver to take manual control where they feel the error margin is insufficient to allow them to detect and correct an error in time • Safety driver able to transition instantaneously from AD to MD mode with a single button press when not satisfied with the safety of the current situation • Steering actuation beyond specified limit shall be prevented
4				Misrepresentation of the environment - error detectable (i.e. implausible signal)	Computational error etc.	Erroneous signal to Subsystem B	Diagnostic system detects issue, error signal given to Subsystem C. Vehicle unable to maintain autonomous control Autonomous control ceases, driver to take manual control	AD system disconnects and reverts to manual driving mode Safety driver has to take control of vehicle as quickly as possible, with no prior notice	x								Visual and audible alarm when AD system disconnected	Highly trained safety driver Safety driver alert at all time and ready to take over at any point in time	• Safety driver to be provided with audio and visual alert when the vehicle transitions from AD to MD mode due to a fault • Safety driver to take manual control where they feel the error margin is insufficient to allow them to detect and correct an error in time • Safety Driver able to transition instantaneously from AD to MD mode with a single button press when not satisfied with the safety of the current situation • The vehicle shall display the current operating mode (AD or MD) to the safety driver

5	Partner X	Subsystem A	Sub-Subsystem A2	Malfunction - no output or no interpretable output	computation error, etc.	Erroneous signal sent to subsystem B	Diagnostic system detects error Parallel control system from Subsystem D able to replace functionality of A2	Vehicle remains in MD mode Safety driver to take control as a precaution, since there is no longer redundancy in system								x	Audio and visual alert given to the driver	Diagnostic system able to detect invalid path request Safety Driver able to make manual control inputs to override the autonomous system	• Detection of driver intervention shall be ensured such that the autonomous system does not fight against manual interventions by the safety driver • The vehicle shall alert the safety driver when control switches from A2 to system D • The vehicle shall display the current operating mode (AD or MD) to the safety driver • Ensure sufficiently smooth transition between paths output by systems A2 and D
6				Misrepresentation of the environment - error NOT detectable (i.e. plausible output)	Computational error etc.	etc.....													
7				Misrepresentation of the environment - error detectable (i.e. implausible output)	Computational error etc.	etc.....													

Appendix 2: HARA Generic Example

The following is an example of a HARA (Hazard Analysis and Risk Assessment) using the template that was created for HumanDrive. This is populated with generic examples to illustrate typical data that would be entered, and other testing organisations are encouraged to make use of the template if it is of assistance to their project. Details of the scoring system can be found in section 5.2. The full version of the document contains separate tabs in the spreadsheet for each testing environment (in the case of HumanDrive, a tab each for testing on private tracks, local roads and the Strategic Road Network), with each tab containing the same list of hazards, but with mitigation measures and scoring adapted to the particular type of testing. Three columns have been omitted from the excerpt below for clarity; these are for review comments, acceptance of risk (yes or no) and recording the version number of the document when that row was last updated.

Please note that the Risk Assessment scores entered are selected at random as a generic example, and therefore should not be taken to be indicative of those for HumanDrive

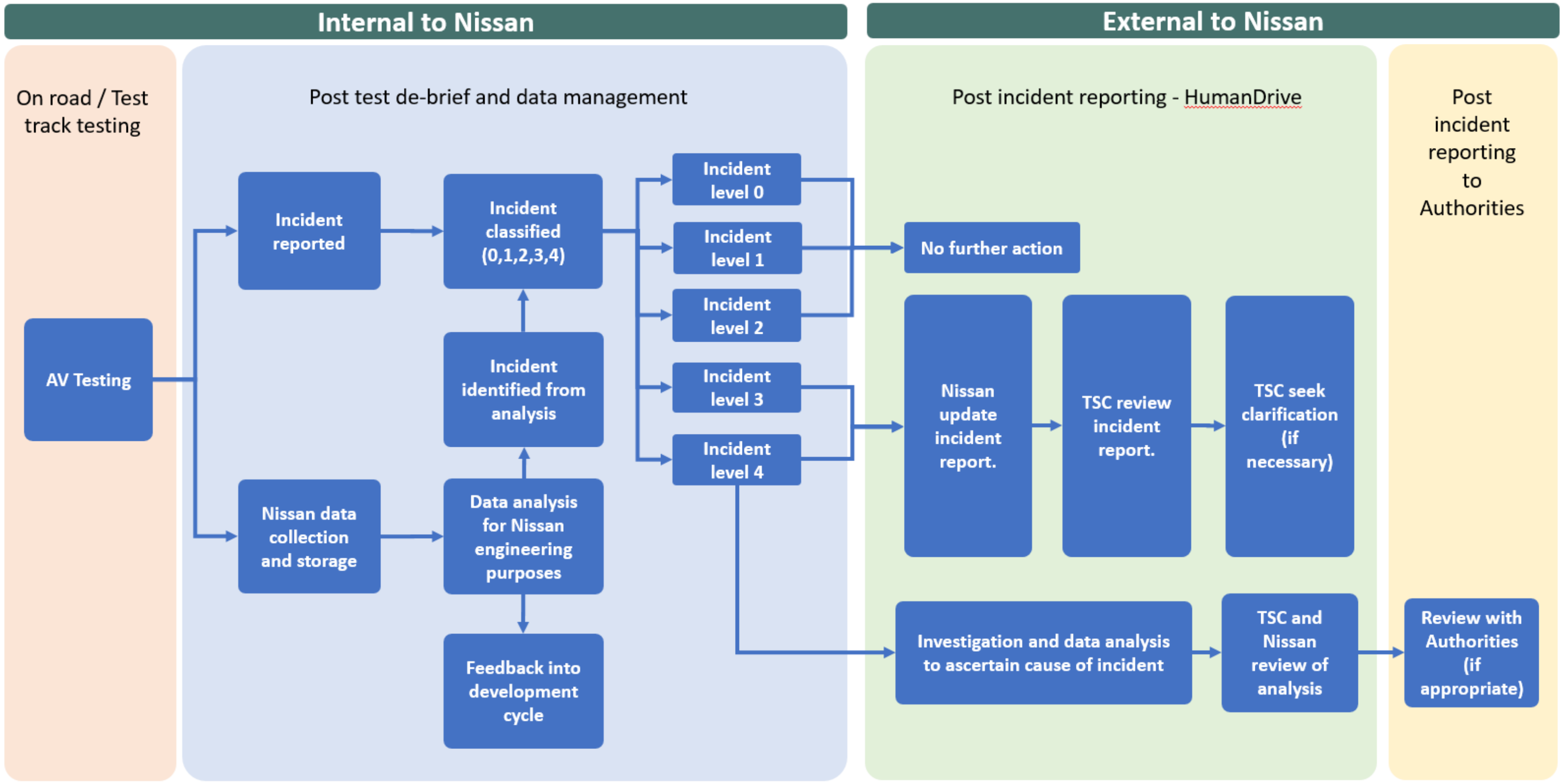
Hazard Analysis					Risk Assessment				Risk Mitigation					Rationale/ Notes
Hazard ID no.	Hazard Description	Existing Mitigation Measures	Hazard Type (Functional Safety, SOTIF, Operational etc.)	Hazard Target (Employees/ public/ both)	P Probab-ility	S Severity	C Contro-llability	R Risk Level	Additional Mitigation Measures	P Probab-ility	S Severity	C Contro-llability	R Risk Level	
T1	**Sudden Loss of Autonomous Control**	* Safety driver ready and able to take control * Controlled environment to limit possibility of collision within safety driver's reaction time * Safety driver to take over pre-emptively if scenario outside ODD is perceived * etc…….	Functional Safety SOTIF (e.g. sensor blinded by sun)	Both	3	3	3	**50**	* Route review to identify where available gap is narrower than time needed for driver to correct vehicle * Driver to take control pre-emptively where there is insufficient space * Driver to take over where assumptions of route review are invalidated, e.g. other vehicle encroaching into lane, HD vehicle travelling close to lane edge * etc......	2	2	2	**20**	*Although loss of autonomous control is very likely to happen at some point, probability is only a 4 as a hazardous traffic scenario (e.g. oncoming vehicle, nearby obstacle etc.) as well as the failure would be needed for an incident to occur * Good controllability as test track ensures plenty of room for driver to make corrections

T2	**Incorrect steering**	* Safety driver ready and able to take control * Controlled environment to limit possibility of collision within safety driver's reaction time * Safety driver to take over pre-emptively if scenario outside ODD is perceived * Behaviour of other road users can be controlled to ensure they don't encroach upon the safety driver's reaction space * etc......	Functional Safety SOTIF (e.g. sensor blinded by sun)	Both	3	3	3	**50**	* Route review to identify where available gap is narrower than time needed for driver to correct vehicle * Driver to take control pre-emptively where there is insufficient space * Driver to take over where assumptions of route review are invalidated, e.g. other vehicle encroaching into lane, HD vehicle travelling close to lane edge * etc.....	2	2	2	**20**	* Hazard includes too much or too little steering - effect similar * Probability also takes into account that hazard would only result in accident if oncoming vehicle or other obstacle is present and near enough for collision to be plausible
T3	**Inappropriate acceleration**	* Safety driver ready and able to take control * Controlled environment to limit possibility of collision within safety driver's reaction time * Safety driver to take over pre-emptively if scenario outside ODD is perceived * Behaviour of other road users can be controlled to ensure they don't encroach upon the safety driver's reaction space * etc.....	Functional Safety SOTIF (e.g. sensor blinded by sun)	Both	3	3	3	**50**	* Vehicle to give prior notice to driver when about to pull out at junction, roundabout etc., so safety driver has advance warning, and hence time to react * etc.....	2	2	2	**20**	
T4	**Inappropriate lack of acceleration**	* Safety driver ready and able to take control * Controlled environment to limit possibility of collision within safety driver's reaction time * Safety driver to take over pre-emptively if scenario outside ODD is perceived * etc.....	Functional Safety SOTIF (e.g. sensor blinded by sun)	Both	3	3	3	**50**	None	2	2	2	**20**	
T5	**Inappropriate braking**	* etc.....						**0**					**0**	
T6	**Inappropriate lack of braking**							**0**					**0**	
T7	**Mechanical breakdown of vehicle**							**0**					**0**	
T8	**Vehicle catches fire**							**0**					**0**	
T9	etc....							**0**					**0**	

Appendix 3: Incident Reporting

The following excepts are drawn from the HumanDrive Incident Reporting Spreadsheet, which includes information on the process to determine what level of analysis different categories of incident should be subjected to. Other projects are welcome to reuse or adapt the approach used by HumanDrive.

Process for Reporting


Internal to Nissan
External to Nissan
On road / Test track testing
Post test de-brief and data management
Post incident reporting - HumanDrive
Post incident reporting to Authorities
AV Testing
Incident reported
Incident classified (0,1,2,3,4)
Incident identified from analysis
Nissan data collection and storage
Data analysis for Nissan engineering purposes
Feedback into development cycle
Incident level 0
Incident level 1
Incident level 2
Incident level 3
Incident level 4
No further action
Nissan update incident report.
TSC review incident report.
TSC seek clarification (if necessary)
Investigation and data analysis to ascertain cause of incident
TSC and Nissan review of analysis
Review with Authorities (if appropriate)

Incident Level Descriptors

Incident level	Title	Description
0	**Minor deviation**	Suboptimal AV lateral and/ or longitudinal control, not conforming to engineering intent but not posing a safety threat. Engineering tuning and optimisation required. May or may not have resulted in intervention from the safety driver
1	**Vehicle Protection Mechanism engaged.**	System attempts to output a trajectory/ speed/ acceleration that lies outside acceptable parameters
2	**Event Outside Operational Design Domain of System**	The safety driver intervenes due to the vehicle encountering a situation it is not designed to deal with, i.e. the vehicle has performed as intended but intervention is still needed for safety. Examples include: - Weather conditions changing unfavourably - Other drivers performing dangerous/ illegal manoeuvres that the HumanDrive vehicle cannot compensate for - Encountering a type of road user that the vehicle is not designed to react to
3	**Near Miss**	The safety driver intervenes to prevent a serious incident (i.e. if the safety driver hadn't intervened, a level 4 incident would have occurred), vehicle comes close to serious incident without the safety driver intervening (e.g. not enough time to react), or significant violation of road traffic laws. Examples could include: "Driver corrected steering due to vehicle cutting across white centre line on apex of corner where there was oncoming traffic that posed a hazard" "Driver took control due to system failure warning. Intervened before vehicle departed lane" "Vehicle failed to slow for 30mph zone, safety driver applied brakes" "Sudden steering input put vehicle close to parked car, safety driver unable to react in time, narrowly avoided contact"
4	**Serious Incident**	Resulting in: - Damage to AV or other road vehicles or roadside furniture - Injuries to one or more parties May or may not be due to system fault - full investigation required either way

Spreadsheet for Recording Incidents

ID No.	Date	Time	Vehicle ID	Incident Description 1 or 2 sentences describing nature and severity	Location e.g. City Street/ Highway/ Coutry Road/ Motorway	Conditions e.g. Sunny/ raining/ cloudy/ night/ wet surface/ fog etc.	Incident Duration	Test Type e.g. static obstacles, private roads, dynamic obstacles, private/ public road, Grand Drive	Incident Definition Use code on 'Incident Definitions' tab
Example	09/12/2017	14:15	NS16 TSC	Vehicle drifted over centre markings as oncoming traffic approached. Safety driver corrected path to ensure safe passing distance	Country Road	Sunny, wet surface	1-2 seconds	Dynamic testing on public road	2
1									
2									
3									
4									
5									
6									

URL: HumanDrive.co.uk
Twitter: @HumanDriveCav
Facebook: HumanDriveCav