Delft University of Technology

MSc Transport, Infrastructure and Logistics
Master Thesis

A dynamic assessment framework for the safe performance of Sidewalk Autonomous Delivery Robots on public sidewalks

Focus area: the Netherlands

Sebastiaan Beekes (4485327)

Coordinators TU Delft
J.A. Annema
A. van Binsbergen
L.A. Tavasszy

Coordinator The Future Mobility Network
T. Klein

To be defended publicly: August 26, 2022 at 14:00

Final version: July 27, 2022

Preface

This publication is written for the completion of the MSc programme in Transport, Infrastructure and Logistics and marks the end of my time as a student at Delft University of Technology. A place where I have grown into who I am today, where I have had valuable experiences and where I have made friends for the rest of my life.

In recent years, our consumer behaviour has changed enormously, putting our delivery system under pressure. Where I used to go into town with my mother when I needed something, Nowadays, the online alternative is increasingly attractive for many people. The consequences of this are visible, and we as a society feel them every day. With my research on delivery robots, as a sustainable alternative for current delivery methods, I want to contribute to the development of a more sustainable society.

When I started my research back in February, the 'learning by doing approach' prevalent at The Future Mobility Network did not always do justice to the principle of a structured scientific research. After a few months, when my research took shape and also in a scientific context it turned out that the most can be learned by most structurally doing, these two extremes came together very well. Thank you Tim for you supervision during this thesis project.

I would like to thank all the academics at Delft University of Technology for their contribution to my research. Thanks to their enthusiasm, expertise and network, they encouraged me to push my thesis to the next level. Bringing together different scientific disciplines has led to broad and structured research into the development, implementation and acceptance of delivery robots. I am curious to see where this research field will mature to in a few years time. Special thanks go to Lóri Tavasszy, Jan Anne Annema and Arjan van Binsbergen as the graduation committee of this research.

I would also like to thank my sister, parents and girlfriend for their unwavering support during my graduation project. The way I have worked towards this goal may not have always been the most social for those around me, but without you by my side, I could not have done this. Rozemarijn, Geert-Jan, Annette and Marlein, I think of you every day.

Graduation can sometimes feel as if you are on your own, a stark contrast to the MSc courses where you are encouraged to work in teams as much as possible. I would like to thank the 'Fokker group' for their support during my graduation process. I would especially like to thank Annedoor, who went through the graduation process simultaneously, who always helped me to structure my thoughts, both on and off my project. I forgot whose turn it is to get us a coffee, but the next one is on me!

To know what you know and what you do not know, that is true knowledge.

Sebastiaan Beekes
Delft, July 27, 2022

Summary

Delivery robots are a promising innovation to mitigate the negative externalities associated with the last mile logistics problem. However their potential is known, so far, no delivery robots can be tested for their performance on the public sidewalk in the Netherlands. In fact: little is known about the actual risks involved in the deployment of delivery robots, there are not yet any good procedures to evaluate the safe performance of delivery robots and there is a lack of objective and complete data from other geographical areas that substantiates that delivery robots can also be safely deployed on the sidewalk in the Netherlands. The aim of the present research is to design a dynamic assessment framework that contributes to the safe and sustainable testing of delivery robots on public sidewalks in the Netherlands. This framework is to be used by stakeholders to collaboratively determine the authorised operating area for delivery robots based on actual data and to define socially desirable system boundaries over time. The Design Science Research (DSR) approach is used to develop a problem solving design suggestion based on the knowledge gained from preceding research steps.

The Operational Design Domain (ODD), which is the set of operating conditions under which a given driving automation system, or feature thereof, is specifically designed to function, used to be classified as binary. In this research it is hypothesised that the ODD actually has different degrees of difficulty and that this difficulty change over time for a specific location due to the dynamic nature of the urban environment. This hypothesis is explored by first applying the 6-Layer model to objectively and completely map the ODD of delivery robots. The 6-Layer model is a structured approach to map an environment as completely as possible on the basis of six distinct categories, including the road network and traffic guidance objects, roadside structures, temporary modifications of the former, dynamic objects, environmental conditions and digital information. Subsequently, it is determined from literature what risks theoretically originate from the complete overview of location characteristics that could be present within the ODD. Based on theoretical evidence it is found that there indeed exist different degrees of difficulty within the ODD, as is also validated with experts. It is outlined how those location characteristics can be objectively measured, such that the foundation has been laid on which geographical locations can be assessed for relative ODD difficulty. To illustrate: it might be more difficult for a delivery robot to drive on an inclined and narrow dusty sidewalk while it is raining than it is difficult for the delivery robot to drive on a flat wide asphalt sidewalk in broad daylight. Because the angle of inclination, road surface friction coefficient, sidewalk width, amount of rain and light intensity of the previous examples can be objectively measured, both locations can be compared in terms of relative ODD difficulty.

The design suggestion proposed in this research is a digitized solution, in line with the outcome of the DSR approach. Differences in ODD difficulty can be used to gradually allow delivery robots on public sidewalks in the Netherlands from a risk minimisation perspective. Delivery robots will first be tested in the known easiest conditions and will be allowed to undergo progressively more difficult operating conditions when the easier conditions are proven to be safe to drive by delivery robots. To do so, the actual location circumstances are monitored and stored in a database along the performance of a delivery robot. The exact location circumstances, in line with the division in the 6-Layer model, are divided into spatial, dynamic objects and environmental conditions. The ODD difficulty of all geographic locations can be assessed along those three condition axes. Based on the data present in the database, the exact ODD difficulty of a network can be displayed at link level on a map. By clicking on a specific link in the network, an explanation is given of the elements in the operating environment that have the greatest impact on the ultimate ODD difficulty level for a delivery robot. This representation gives stakeholders insight into why a location is or is not passable for a delivery robot. In this research it is decided to display the dynamic assessment framework on a link level because routes consists of several adjacent links and because to date there are still many uncertainties which need to be addressed before an even more detailed representation is justified. A possible incident reporting function has been elaborated because this is mandatory in leading countries on autonomous driving innovations. Incident reporting also contributes to system transparency. In addition to making the system understandable to human stakeholders, the dynamic assessment framework can be linked to the navigational component of delivery robots. By means of geofencing, a delivery robot can be denied access to locations with difficult ODD levels if it has not yet proven its ability to perform safely in such circumstances.

Because delivery robots as innovation and the dynamic assessment framework proposed in this research are far from being mature, it is argued that the dynamic assessment framework should be applied within a living lab. Businesses, researchers, authorities and citizens work together in a living lab to evaluate, learn and develop delivery robots and the assessment framework in a real-life context according to what is learned. Over time developments can be made to the delivery robots, in the way ODD levels are classified, in the included dynamic assessment framework functionalities and to the delivery robot performance metrics proposed in this research. The living lab approach ensures that it is not only investigated whether delivery robots perform technically

safe, but also works towards the most socially acceptable integration of delivery robots into the existing traffic system.

From the division in spatial, dynamic objects and environmental conditions it is concluded that it is currently most difficult to determine ODD difficulty associated with dynamic objects. This difficulty changes more frequently due to the position, rotation and dynamics of the objects on the sidewalk surface. Additional difficulties are that, to objectively determine the difficulty, all obstacles on the sidewalk must be mapped in real-time and no accepted methods and metrics exist yet to measure these obstacles. In this research the 'Sidewalk Minimum Clear Width Accounting for Static Obstructions', the 'Average Busyness by Vulnerable Road Users' and 'total occupied sidewalk area' are proposed to measure the ODD difficulty resulting from dynamic objects. Static LiDAR systems seem to be a promising solution in contributing to measure these variables, but certain systems are currently too expensive to install in an entire city network. A more cost-effective approach is to use counters. To determine relative ODD difficulty resulting from the spatial component, a sidewalk network has to be assessed on its characteristics per sidewalk section. To do so, possible existing Geographic Information System data can be used. The spatial assessment can be periodically revised to keep the dynamic assessment framework up to date. To keep track of the ODD difficulty resulting from environmental characteristics existing real-time data sources can be used.

Two geographic locations classified on ODD difficulty by their spatial, dynamic objects and environmental characteristics can be graphically displayed in the *xyz*-space. The distance between the two datapoints can be calculated with the three-dimensional Euclidian distance. If the two points in the three dimensional space are close enough to each other, one could argue that if a delivery robot manages to drive safely at the first location, it will most likely do so as well at the second location. This makes knowledge gained in one geographical location translatable to other locations and solves the identified lack of objective and complete data on current delivery robot performances.

During the expert validation of the dynamic assessment framework two different perspectives came up. The first perspective is that the location characteristics used to determine ODD difficulty in the current proposal of this thesis are currently too generic and abstract to be useful to benefit delivery robot tests on the public sidewalk. The second perspective is that the proposed system is better than anything that currently exists regarding delivery robot tests on the public sidewalk and that with the living lab approach elements of the dynamic assessment framework that prove to be too generic or abstract can be improved over time. Theoretically, the proposed framework is what the performance assessment of delivery robots needs. Empirical evidence is the missing link between the theoretically established ODD difficulty levels and the envisaged dynamic assessment framework. Further development of the proposed framework should be driven by factors that actually impact the performance of delivery robots.

Development of the proposed dynamic assessment framework may contribute to safely testing and integrating delivery robots into the Dutch traffic system and society over time. The proposed dynamic assessment framework provides a means to learn about socially acceptable delivery robot performance in its objective and dynamic context, while at the same time bring society closer to a world where autonomous delivery alleviates the externalities of the last mile logistics problem.

Contents

List of Figures

List of Tables

Abbreviations

ABC	Achieved Behavioral Competency
AD	Aggressive Driving
ADR	Autonomous Delivery Robot
ADS	Automated Driving System
ADSA	ADS Active
AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
APK	General Periodic Test
AV	Autonomous Vehicle
CBR	Centraal Bureau Rijvaardigheidsbewijzen
CFI	Corporate Finance Institute
CI	Collision Incident
DSR	Design Science Research
EU	European Union
FC	Functional Constraint
FMEA	Failure Modes and Effects Analysis
FO	Functional Objective
FoV	Field of View
GIS	Geographical Information System
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HARA	Hazard Analysis and Risk Assessment
HTCDER	Human Traffic Control Detection Error Rate
HTCVR	Human Traffic Control Violation Rate
ISAD	Infrastructure Support Levels for Automated Driving
ISO	International Standards Organization
KNMI	Royal Netherlands Meteorological Institute
LiDAR	Light Detection and Ranging
LSAD	Low-Speed Automated Driving
MSDCE	Minimum Safe Distance Calculation Error
MSDF	Minimum Safe Distance Factor
MSDV	Minimum Safe Distance Violation
MTTC	Modified Time-to-Collision
NFC	Non-Functional Constraint
NFO	Non-Functional Objective
NHTSA	National Highway Traffic Safety Administration
ODD	Operational Design Domain
OES	Operating Envelope Specification
PDD	Personal Delivery Device
PRA	Proper Response Action
RADAR	Radio Detection and Ranging
RRV	Rules-of-the-Road Violation
SADR	Sidewalk Autonomous Delivery Robot
SAE	Society of Automotive Engineers
SBAS	Satellite-Based Augmentation Systems
SJR	SCImago Journal Rank
SOTIF	Safety of the Intended Function
SWOV	Institute for Road Safety Research
TTC	Time-to-Collision
UK	United Kingdom
US	United States
V2I	Vehicle to Infrastructure
V2N	Vehicle to Network
V2P	Vehicle to Pedestrian
V2V	Vehicle to Vehicle
V2X	Vehicle to Everything

VRU	Vulnerable Road User
WADGPS	Wide Area Differential GPS

1 Introduction

A theoretically suitable solution to mitigate negative externalities of the last mile problem is the deployment of Sidewalk Autonomous Delivery Robots (SADRs) (Vleeshouwer, Rotterdam, & Verbraeck, 2017; Jennings & Figliozzi, 2019; Boysen, Schwerdfeger, & Weidinger, 2018). In order for this solution to be suitable in practice, it must be proven that this technology is safe, but proving that SADR technology is safe enough to operate on the public sidewalk is a harsh question to tackle. Standard measures of safety and performance that are currently used for motorized vehicles in traffic, or for automated vehicles in restricted areas, cannot be directly used to argue that the autonomous driving technology is safe. For example, the number of kilometres driven without any intervention by a teleoperator can be an incentive for a supplier to drive many (self reported) kilometres on quiet, very well-surfaced infrastructure, to drive very slowly or not to intervene in unsafe situations. The number of safe kilometres driven only becomes useful if it can be put into perspective in which environment and under which conditions the kilometres were driven. The quality and surface material of the sidewalk may influence the extent to which an SADR is able to move safely on the sidewalk, but the presence of a group of pedestrians or perhaps the specific weather conditions might also have an influence. The question then arises what level of detail of these conditions is appropriate to make sound assessments? In addition, the set-up of the test method of autonomous driving technology can be an incentive for undesirable behaviour of the supplier, such as "the so-called gaming of tests, which means that the manufacturer optimizes the system's performance in the predefined test cases" (Ponn, Gnandt, & Diermeyer, 2019, p. 1). A good example is the Dieselgate scandal, where 11 million cars from the Volkswagen Group were found to be equipped with software that masked the emission of harmful substances (Velzen, 2021). This example confirms that there is natural distrust between the different parties in the approval field of new vehicles or vehicle techniques. The regulator needs to develop good and critical test methods to test the safety of SADRs and cannot rely too much on the safety evidence provided by the SADR supplier. Current safety assessment methodologies for automated driving systems have started to focus on the automated driving systems of passenger vehicles, little is known about SADRs. The tricky part of Operational Design Domain (ODD) determination within Automated Driving Systems (ADS) in passenger cars is that there is a boundary situation where an ADS must give control of the vehicle back to the driver. The driver may not be aware of this or may not have a full overview of the situation, which can lead to unsafe situations. In addition, these kinds of situations often occur at high speeds where the vehicle is surrounded with other road traffic. This situation is different for SADRs. When an SADR encounters a border situation which cannot be handled with certainty, the help of a teleoperator is called in. This probably does not have to happen within a second because an SADR is able to take up a safe position on the sidewalk. In addition, an SADR drives at a maximum speed of 6 kilometres per hour. Combined with the 40 kilogram weight of the SADR of focus in this study (see Appendix E), this means that the kinetic energy of a moving SADR is significantly lower than that of a moving autonomous car. Furthermore this could mean that the environment in which an SADR finds itself changes much less quickly than the environment in which ADS in a passenger vehicle finds itself. An SADR could therefore better be able to avoid unsafe situations. Current research and safety assessments are mainly focused on the assessment of autonomous passenger vehicles, while the probable lower safety risks associated with autonomous driving mean that an SADR can be used to gain experience with self-driving systems on a more controlled scale. In addition, the ODD of a delivery robot is limited to only the urban context and therefore more manageable, allowing for a faster development of a methodology for classifying the ODD into a usable standard, before it is translated into the complete ODDs of ADS.

A sound methodology to prove that SADR technology is safe given the dynamic environments SADRs operate in is currently non existent. Knowledge gained in practical tests can hardly be translated to other use cases. This research investigates how a dynamic assessment framework for the safe performance of SADRs on public sidewalks in the Netherlands should be designed and what it should look like. The focus of this research is on the test phase of SADRs. Conducting safe tests does not necessarily lead to the permanent admission of SADRs to all public sidewalks.

1.1 Knowledge gaps

Several knowledge gaps can be identified within this research. First, the lack of a scientifically accepted method to define the ODD. subsequently applied to the case of SADRs. Second, the relation between infrastructural and environmental characteristics and the driving ability of SADRs has never been explicitly assessed, nor quantified. Three, there is no structured and objective method by which delivery robots can be safely tested on the public sidewalk. Testing their performance on the public sidewalk is of interest because the conditions in that test environment best mimic the conditions that delivery robots will encounter in daily life. Passing the test phase could eventually lead to admission to the public sidewalk.

1.2 Empirical research: case company

This research was conducted in cooperation with the Future Mobility Network, a knowledge and consultancy company in the Netherlands focused on the mobility of the future. One of the Future Mobility Networks projects is focused on SADRs. The interest of the Future Mobility Network in this research is that an accepted dynamic testing framework for SADRs on public sidewalks in the Netherlands can contribute to an increase in the number of pilot projects. Both the geographical scope and the type of SADR as subject of this study have been chosen in consultation with the Future Mobility Network. The research is documented in such a way that the research can be carried out for other types of delivery robots or different operating areas using the same methodology.

1.3 Design Science Research approach

The research will follow the design science research (DSR) approach to contribute to the presented knowledge gaps. This approach studies and creates artifacts that are innovative and solve real-world problems (Merwe, Gerber, & Smuts, 2017). For this research specifically the design science research cycle approach of Vaishnavi & Kuechler (2004) is adopted, because the research effort tends to be problem-solving focused in the approach as opposed to questions or problems that are answered through explanation. The artifact to be designed using this approach is a dynamic assessment framework to safely test SADRs on the public sidewalk in the Netherlands. The approach starts by taking the context of the problem and the problem awareness, which suits the scope of this research because SADR safety assessment has many different ramifications that need to be taken into account. By identifying the challenges and requirements for an applicable safety assessment framework, the boundaries and demands for the design suggestion are outlined (Johannesson & Perjons, 2021). The DSR approach helps in the structured development of such an artefact. As displayed in Figure 1 in this research process steps one to four will be executed. The outcome of this research can be used as the basis for the remaining process steps to develop a fully functioning dynamic SADR safety assessment system. To provide structure in the report the mapping between the proposed structure of a research report and the DSR process model of Vaishnavi & Kuechler (2004) established by Merwe, Gerber, & Smuts (2017) is used. This will be further explained in Section 2.

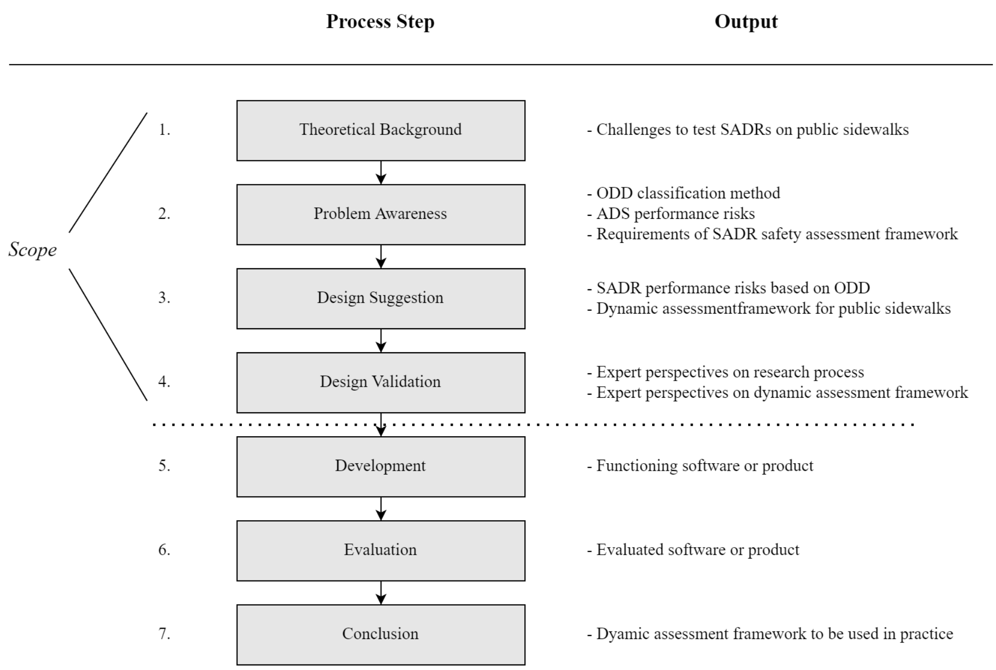


Figure 1: Design Science Research cycle adapted from Vaishnavi & Kuechler (2004)

1.4 Research questions

To guide and structure this research, the following research goal, corresponding sub-questions and methods will be used throughout the research:

To design a dynamic assessment framework for the safe performance of Sidewalk Autonomous Delivery Robots on public sidewalks in the Netherlands	
Sub-question	**Method(s)**
1. What challenges and needs exist in the Netherlands to test SADRs on the public sidewalk?	Desk research
2. What state of the art methods can best be used to objectively identify and classify the different static and dynamic elements in the Operational Design Domain of Automated Driving Systems?	Literature research and desk research
3. What are the risk factors associated with driving SADRs on public sidewalks?	Literature research, desk research and informal interviews
4. What are the requirements for a dynamic assessment framework for the safe performance of SADRs on public sidewalks in the Netherlands?	Literature research and desk research
5. What dynamic assessment framework can be designed to assess if an SADR is able to operate safely at different ODD-levels?	Creative methods
6. What expert perspectives exist regarding the establishment of different levels of difficulty within the Operational Design Domain of SADRs?	Semi-structured (face-to-face) expert interviews
7. What expert perspectives exist regarding the applicability of the proposed dynamic assessment framework design suggestion?	Semi-structured (face-to-face) expert interviews

The main research goal is as follows:

To design a dynamic assessment framework for the safe performance of Sidewalk Autonomous Delivery Robots on public sidewalks in the Netherlands.

The dynamic nature of the to be designed assessment framework is on the one hand focused on the dynamics of location characteristics and traffic participants on and next to the sidewalk. On the other hand a dynamic assessment framework must take into account that both the innovation of SADRs and the requirements for safe performance of this innovation evolve over time. The main research goal can be reached by successively answering the sub-questions presented. These sub-questions are divided over the four process steps by Vaishnavi & Kuechler (2004) that fall within the scope of this research: the theoretical background step, the problem awareness step, the design suggestion step, and the design validation step. This will be further elaborated on in Section 2.

2 Research methodology

In this section the research flow framework is presented, according to which the study is structured. Based on the process steps of the DSR cycle, the research questions and the corresponding research methods will be discussed. The research flow framework is extended with an overview of the research positioning, from which follows that the key concepts of this research and their relationships are gradually explored.

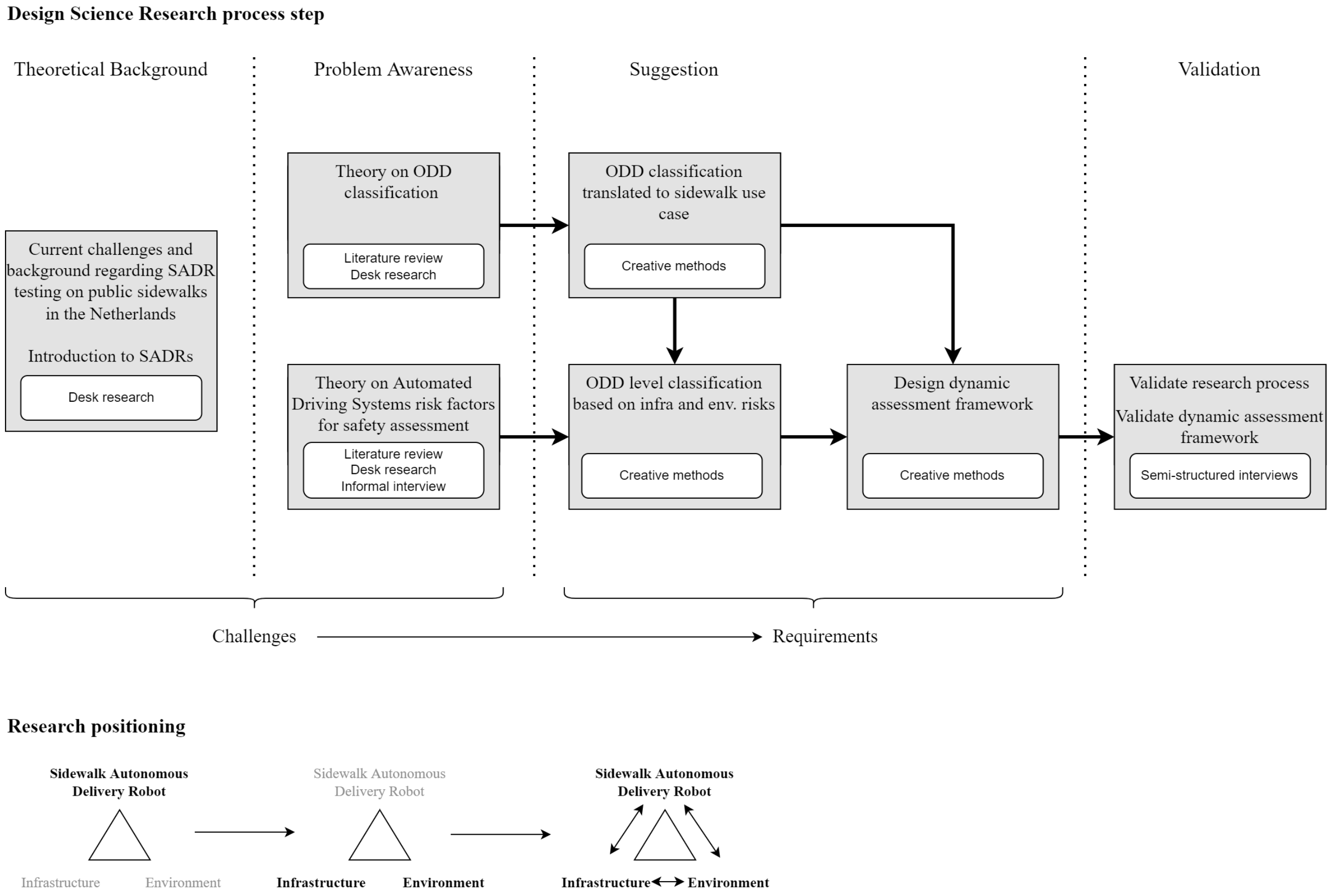


Figure 2: Research flow framework

2.1 Theoretical background

The purpose of this process step is to address the complexities and currently existing barriers to test SADRs on the public sidewalk. Background information on current driver safety, vehicle safety and Autonomous Driving System safety assessments can form a basis for the future safety testing procedure for autonomous delivery robots given the overlap present. To overcome the knowledge gap that there is no sound safety assessment methodology for SADRs, it is important that the current challenges why this is so are addressed in this research so that they can be taken into account in the design suggestion. In addition, it is important to explicitly describe the SADR innovation and the (administrative) playing field in which the SADR is situated, so that it is clear to researchers on what known (scientific) knowledge the research is based. In this way, no assumptions or prior knowledge are assumed by the readers. The first sub-question in this research is:

1. *What challenges and needs exist in the Netherlands to test SADRs on the public sidewalk?*

This sub-question will be answered by means of desk research on policy documents, use cases, (news) articles and technical documents. Because of the nature of this research, there is not yet sufficient scientific knowledge, which rules out carrying out a literature review.

2.2 Problem awareness

In the problem awareness step of the research the practical problem is investigated in more detail and is being worked towards the design suggestion (Merwe et al., 2017), by analysing relevant literature, Two knowledge areas to be further analysed in this study are the objective classification of the Operational Design Domain and the risk factors associated with driving SADRs on public sidewalks. The corresponding second and third sub-question are:

2. *What state of the art methods can best be used to objectively identify and classify the different static and dynamic elements in the Operational Design Domain of Automated Driving Systems?*

3. *What are the risk factors associated with driving SADRs on public sidewalks?*

Both sub-questions will be answered by means of a literature study, if necessary expanded with informal interviews and desk research to add required knowledge based on grey publications. The exact steps followed in the literature studies are described in detail in the associated sections. Because the scientific literature on SADR risk factors is non-existent, it has been chosen to focus the literature review on ADS risk factors and later translate these risks into SADRs, as shown in Figure 2.

2.3 Design suggestion

As shown in Figure 2, the existing challenges to assess the safe performance of SADRs on the public sidewalk identified in the theoretical background and problem awareness process steps form the basis for the requirements that will be used to guide the actual development of the design suggestion. Based on the insights gained from the first three sub-questions an answer can be formulated to the fourth sub-question:

4. *What are the requirements for a dynamic assessment framework for the safe performance of SADRs on public sidewalks in the Netherlands?*

Defining the requirements for a testing framework is a crucial part of this research. Establishing the requirements can be described a transformation of the problem into criteria for the design suggestion (Johannesson & Perjons, 2021). Requirements are not only determined for functionality, it will also help to focus on working towards the design suggestion and provide structure in doing so. The knowledge and insights gained in the problem awareness step of a research also contribute to the design suggestion step (Merwe et al., 2017). The requirements to be met by a testing method for SADRs on public sidewalks is a prerequisite for the actual design suggestion development that will be guided by the fifth sub-question:

5. *What dynamic assessment framework can be designed to assess if an SADR is able to operate safely at different ODD-levels?*

The approach to answer this question is twofold. The first step to answer this question focuses on a method to determine different ODD levels. This method subsequently forms the basis to design the dynamic assessment framework for SADRs. Because to the authors knowledge no such assessment framework exists, creative methods are used to design the assessment framework.

2.4 Design validation

To conclude the research, it is important that the proposed design solution can actually be used for what it is intended to do, which is to provide an accepted objective method for parties to safely test SADR performance on public sidewalks in the Netherlands. First, it is important that the proposed design meets the drawn up design requirements. Second, the conducted research and resulting proposed design suggestion can be assessed by interviewing experienced academics and practitioners in the professional field. The sixth and seventh sub-questions to assess the validity of this research therefor are:

6. *What expert perspectives exist regarding the establishment of different levels of difficulty within the Operational Design Domain of SADRs?*

7. *What expert perspectives exist regarding the applicability of the proposed dynamic assessment framework design suggestion?*

By means of semi-structured (face-to-face) interviews the different views on the theoretical basis formed in this research and the proposed dynamic assessment framework built on this theoretical basis can be revealed. Semi-structured interviews suit this research, because the research area is complex and little about it is known (Saks & Allsop, 2012), because the open nature of the research questions can lead to an in-depth conversation pursued from the given answer (Newcomer, Hatry, & Wholey, 2015) and because the flexibility of the semi-structured interview approach can lead to new insights (Saks & Allsop, 2012; Newcomer et al., 2015), During the semi-structured interview, the research process is first validated, so that it is clear to the interviewed experts on what basis the dynamic assessment framework has been designed.

2.5 Research positioning

As shown in the bottom of Figure 2, the research is gradually expanded around the key concepts: the Sidewalk Autonomous Delivery Robot, the infrastructure, and the environment. In the Theoretical Background process step, the delivery robot is introduced as an innovation. In the Problem Awareness step, the infrastructure and the environment in which a delivery robot operates are discussed. In the Design Suggestion step, the interactions between the Sidewalk Autonomous Delivery Robot performance, the infrastructure, and the environment are brought together and discussed.

3 Theoretical background

For as long as anyone knows, motorized vehicles have always had a driver. A vehicle is inspected, authorised and driven by a driver who is certified to drive. These two authorisations, that of the vehicle and that of the driver, are completely separate. With the advent of self-driving vehicles, and vehicles where there is not even room for a driver, this system becomes inapplicable. In this section the contextual background of this thesis project is outlined. It is explained why the current processes and systems for vehicle inspection are not applicable to test SADRs and why new procedures cannot be set up just like that. The current knowledge about delivery robots with a teleoperator, levels and operation of autonomous driving are of interest to understand the research positioning and boundaries of the study. The knowledge in this section will be used to answer the first sub-question:

1. *What challenges and needs exist in the Netherlands to test SADRs on the public sidewalk?*

3.1 Current safety assessments

An overview of the current and future situation of vehicle and driver assessments is given in Figure 3. In this section the relevant processes and developments regarding vehicle approval and testing and driver assessment in the Netherlands will be discussed, so that the playing field in which a Sidewalk Autonomous Delivery Robot (SADR) finds itself becomes clear.

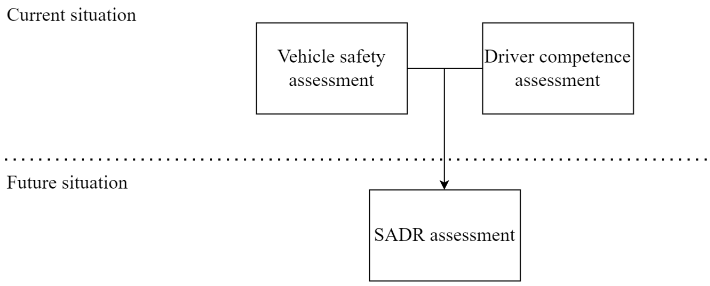


Figure 3: Context current and future situation of vehicle and driver assessments

3.1.1 Driver assessment in the Netherlands

Obtaining your driving licence in the Netherlands is a straightforward process and similar for each type of motorized vehicle. If you are of legal age and in good health to take driving lessons, you can do so at a certified driving school. When you have passed your theory exam, and your driving instructor judges that your level of driving is good enough, you get your driving licence by passing the practical exam. This practical exam is conducted by the Centraal Bureau Rijvaardigheidsbewijzen (CBR) (CBR, n.d.). The possession of a driving license is conditional. This means that you remain in possession of the driving licence as long as you exhibit the desired behaviour on the road. If you are frequently caught by the police committing (serious) traffic violations, your driving licence can eventually be confiscated and the public prosecutor determines for how long you are to lose your driving licence. You are not allowed to drive any motor vehicle during this period (Rijksoverheid, n.d.). It can be observed that in the Netherlands no interim assessment of drivers' competence to drive exists, for instance on a random basis. In addition must be noted that the current system makes it impossible for the police to observe and take action on every traffic violation at all times.

Although the process looks simple and clear, grey areas in the safety assessments of drivers can be distinguished. The question is: "When can someones ability to drive be classified safe?" Despite the fact that one may assume that both a driving instructor and an examiner from the CBR are good at their jobs and give an honest assessment of ones ability to drive safely, safety is an enormously difficult concept to measure. Especially in the dynamic environment of public roads. Also, this assessment is done at the beginning of a driver's career and never revised. At the end of this section, this topic will be further discussed regarding the safety assessment of driving SADRs.

3.1.2 Vehicle approval in the Netherlands

In the process of admitting new vehicles to the Dutch public roads a distinction is made between three types of vehicle type approval:

1. European type approval
2. National small series type approval
3. Individual approval

According to the Vehicle Certification Agency (2022) "Vehicle Type Approval is the confirmation that production samples of a type of vehicle, vehicle system, component or separate technical unit will meet specified performance standards". The specific type approval process is also known under the synonym homologation. A vehicle that has obtained European vehicle type is valid for registration in all European member states and member states of the European free trade association and Norway, Iceland, Liechtenstein and Switzerland (RDW, 2020). This standard has been developed to ensure that, once approved, manufacturers are able to trade approved vehicles freely in the aforementioned regions. An overview of the current vehicle categories distinguished is presented in Table 1.

Table 1: Overview of motorized vehicle categories, adapted from RDW (n.d.-c)

Vehicle	Category	Description
Passenger cars, buses and coaches	M	Motor Vehicles with at least four wheels designed and constructed for the carriage of passengers
Commercial vehicles	N	Motor Vehicles with at least four wheels designed and constructed for the carriage of goods
Trailers	O	Trailers (including semi-trailers)
2 and 3 wheeled vehicles	L	Mopeds, motorcycles, motortricycles and quadricycles
Wheeled tractors	T	Wheeled tractors
Track-laying tractors	C	Track-laying tractors
Agricultural trailers	R	Agricultural trailers
Interchangeable towed machinery	S	Interchangeable towed machinery
Mobile machinery	U	Motor Vehicles designed and constructed for activities other than the carriage of passengers or goods

For the vehicle categories from Table 1, the European regulation describes in detail which requirements they have to comply with and how these requirements are to be tested. According to the work of Dutch Safety Board (2019), these requirements may not be deviated from at national level, nor may member states impose additional requirements on top of the European requirements. The regulation covers a fairly homogeneous vehicle category. The vehicles within a category differ little in design, functionality and method of operation, as a result of which unambiguous requirements can apply to the vehicle category. Conversely, the specific technical requirements help to keep the vehicle category homogeneous. When regulating vehicles on public roads, the vehicle category is the starting point. The vehicle manufacturer or importer presents a vehicle for assessment in a certain vehicle category. It is then assessed whether the vehicle belongs in this category, after which the vehicle is tested against the requirements that correspond to this category. A visual representation of this process for European vehicle type approval is given in Figure 4.

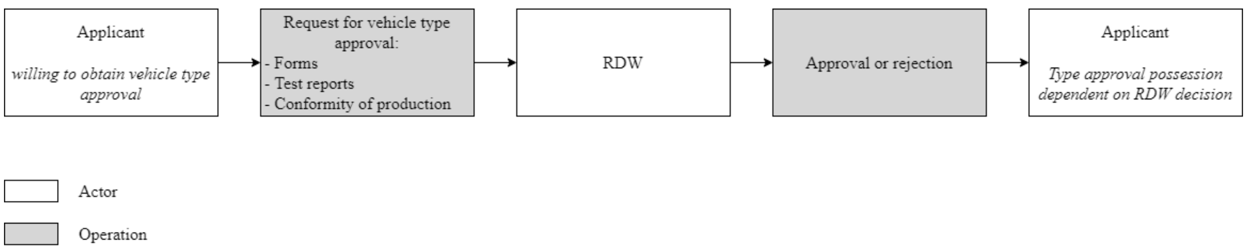


Figure 4: Type approval application process for EU approved vehicle type categories

In order to be authorised to be operated on public roads on the basis of a European vehicle category, a vehicle has to fulfil two different types of technical requirements: admission requirements and permanent requirements. Admission requirements are requirements that the vehicle must meet after its manufacture. The admission requirements that are tested by the RDW in this context, and on which approval or rejection depends, focus on the technical requirements related to safety. Specifically, the following test categories are distinguished (RDW, n.d.-a), whereby the specific requirements for the tests depend on the vehicle categories from Table 1:

- Noise measurements
- Tests of braking devices
- Steering gear test
- Application of forces to a vehicle structure (section)
- Tyre grip on wet road surface
- Tyre rolling resistance
- Determination of field of vision
- Functionality of (electronic) vehicle systems
- Presence and installation of vehicle parts on a vehicle (structure)

Permanent requirements are requirements for the use phase and are requirements that the vehicle must meet at all times during the usage phase. Permanent requirements are assessed during the yearly General Periodic Test (APK), which is legally required in Europe. During the inspection, a certified inspector assesses a vehicle for its road safety, impact on environment and administrative registration (ANWB, n.d.-b). In contrast to the driver's assessment in Section 3.1.1 it is striking that drivers do not have something similar to an APK.

The second type of vehicle type approval is national small series type approval. Countries may decide at national level to approve deviating vehicle configurations produced in small series, which apply outside the set European type approval, nevertheless. These vehicles are also subject to strict testing by RDW with regard to safety aspects. The third type of type approval, individual type approval, can be granted to a single vehicle that does not have type approval when tests show that the vehicle complies with the set requirements of an acknowledged vehicle category. This type approval will not be considered further in this research, because SADRs will not be approved to participate in road traffic on an individual basis.

For vehicles of which the layout, dimensions or weight do not comply with the rules in the Road Traffic Act, the RDW has the opportunity to issue a ZZ registration plate (RDW, n.d.-d). A vehicle with a ZZ registration plate is only allowed on the road if additionally an exemption has been obtained from the RDW or the road authority. It is not clear what technical or functional requirements are involved to obtain such an exemption.

It can be concluded that vehicles that correspond within one of the categories in Table 1 can be admitted to public roads after vehicle type specific tests. However, for vehicles that do not fall within a specific category, such as SADRs, there are no established requirements and test methods, so that there is no direct possibility to be admitted to road traffic.

3.2 Establishment of a new vehicle category in the Netherlands

This section describes the last time a that a new vehicle category was introduced in the Netherlands, and what the consequences were. It is rare that a new vehicle category is established. Since SADRs do not currently fall into any vehicle category, it is to be expected that in a few years' time, when delivery robots can be safely admitted to public sidewalks, a similar procedure will have to be gone through again. The information presented in this section is sourced from the Dutch Safety Board (2019), and it is indicated otherwise. This was chosen because this publication precisely and completely describes the creation of a new vehicle category and the problems arising from it, which is the purpose of this section.

3.2.1 The special moped as new vehicle category

The Segway could not be admitted as a moped because it did not fit into the category of two-wheeled motor vehicles, category L from Table 1. There was also no legal possibility for the Segway to be exempted as a vehicle type in the Netherlands (Dutch Safety Board, 2019, p. 30). Under political pressure of the House of Representatives, the Minister of Transport and Water Management was willing to admit the Segway to public roads. "The European Commission offered the Netherlands the option of creating a national vehicle category for the Segway. The Minister decided to broaden this option by introducing a new vehicle category of 'special mopeds' [*bijzondere bromfiets*], which would allow for other innovative vehicles to be admitted to the roads, in addition to the Segway, in the future" (Dutch Safety Board, 2019, p. 34).

3.2.2 Admission requirements for the special moped

There were no implementing rules defining the requirements for admittance and the admission process when the legislation regulating the special moped category took effect in 2011. In contrast to the existing European categories, the national category for special mopeds was more flexible and had less restrictions on the types of vehicles that could and could not be accepted in the admittence process. In Table 2 the differences in admission procedures for existing vehicle categories is set out against the admission procedure for special mopeds.

Table 2: Existing vehicle category admission compared to the special moped vehicle category

Existing vehicle category admission	Special moped vehicle category admission
Strict definitions	Open character
Admission requirements and permanent requirements	Permanent requirements
Permission granted by executive organisation that conducts safety tests	Permission granted at administrative level
Permission granted based on meeting technical standards	Permission granted based on non-binding vehicle technology recommendations and road safety recommendations

According to European vehicle categories, a vehicle must meet both entry requirements and ongoing requirements in order to be authorised to travel on public roads. “Requirements for admission are standards that vehicles must meet when they leave the factory. Permanent requirements are requirements for the use phase and are requirements that vehicles must meet at all times” (Dutch Safety Board, 2019, p. 40). The Minister directed the RDW to solely evaluate the special moped vehicle category in accordance with the permanent requirements. Insight into the existence of flaws in the design and production stages of this vehicle category was therefore limited (Dutch Safety Board, 2019, p. 42). The Institute for Road Safety Research (SWOV) conducted the road safety study of special mopeds. This road safety study, as used for the first generation special mopeds, was limited in scope and only based on the Segway specific study, used to shape the vehicle category of the special moped. This, while the special moped vehicle category indicated above had an open character, which meant that vehicles other than the Segway could also be admitted in this way.

A vehicle will be accepted to a category of vehicles when it satisfies all technical requirements in the European admission process. The vehicle technology study (by RDW) and road safety study (by SWOV) had the status of recommendations when special mopeds were authorized in the Netherlands (Dutch Safety Board, 2019, p. 38). The road safety study was not even required to be executed. The recommendations of both organisations were not binding, giving the Ministry the opportunity to deviate from them. In the European procedure, the decision to grant permission is taken by the implementing organisation that conducts the safety tests, as opposed to the special moped vehicle category where this decision was made at the administrative level. In deciding so, this deviated from the standards in force in the European admission procedure.

Because the vehicle category was so flexible and did not have such strict requirements as other existing vehicle categories, from 2012, the number of applications for admission of vehicles that could not meet the requirements of the special category of mopeds increased. This increasing number of applications caused the Ministry a lot of unnecessary burden. “The Ministry tried to reduce this number of applications for these types of vehicles by adjusting the procedure, tightening the requirements and increasing the costs for the applicant” (Dutch Safety Board, 2019, p. 43). Although the requirements for admission to the admission procedure in this way were tightened at the front end, the substantive tests remained simple and the derogation from European approval remained: technical testing according to visual inspection and the decision by the Minister based on non-binding advice from safety bodies.

3.2.3 The Stint as special moped

The Stint was admitted in the Netherlands under the aforementioned regulations. “The Stint is a vehicle that was especially designed for use by childcare organisations to transport children” (Dutch Safety Board, 2019, p. 48). The Stint’s manufacturer did not have any prior knowledge of admittance procedures when it created the stint, hence the Stint was not intended to suit a particular vehicle category. This deviated from the typical path of vehicle development by doing so. The Ministry still chose to accept the Stint to the special moped procedure even though it did not fulfill the size requirements and was meant for the transportation of children, for whom there were no specific requirements in the category of special mopeds. Although both the RDW and the SWOV issued a negative recommendation about the admission of the Stint, the Ministry decided that the Stint was sufficiently safe to participate in traffic (Dutch Safety Board, 2019, p. 49).

On 20 September 2018 a Stint drove off on a railway crossing and drove on while the crossing barriers were already closed (Dutch Safety Board, 2019, p. 6). A passing train collided with the Stint. The impact was fatal for four of the five children on board, and fifth child and the driver were seriously injured. After this collision, the Dutch Safety Board looked into how light-motorized vehicles and special mopeds are admitted to the road. According to their research: “in the past, road traffic safety was primarily seen as a possible barrier to innovation" (Dutch Safety Board, 2019, p. 9). The history of the unique moped category demonstrates how the House of Representatives requested that the former Minister of Transportation and Water Management came up with a quick and simple solution to admit innovations without a suited vehicle category. In the political decision-making process, safety was made subordinate to the desire for innovation in road traffic. Following the aforementioned investigation with the Stint, on 2 May 2019 the Minister of Infrastructure and Water Management (*the Ministry has since been renamed*) tightened the rules for allowing a special moped on the road, the so-called admission framework (Rijksoverheid, n.d.).

3.2.4 Current situation

Based on the investigation conducted following the accident with the Stint, the Dutch Safety Board concluded that "insufficient attention is paid to safety when admitting light-motorised vehicles that do not fall within the European admission procedures" (Dutch Safety Board, 2019, p. 61) and that "there is currently insufficient insight into the safety of light-motorized vehicles and the effect of these vehicles on traffic safety. An integral approach to the safety of light-motorized vehicles is missing within the road safety strategy. As a result, an important instrument for monitoring and improving traffic safety is insufficiently used" (Dutch Safety Board, 2019, p. 62).

The Stint accident and the attention it generated also put the admission and inspection of new vehicles in a new perspective. Testing and admission of vehicles within existing vehicle categories is carried out as intended. For the possible testing and admission of vehicles within the category of special mopeds, it is now first strictly checked whether a vehicle actually belongs in that category before the more stringent procedure is started. Vehicles that cannot be classified in a recognised international category due to their characteristics do not currently find a good opportunity to be admitted to the public roads in the Netherlands. The actors in the admission field are reticent, in order to prevent fatal accidents under their responsibility.

3.3 Sidewalk Autonomous Delivery Robot

An innovative vehicle that currently cannot be classified in any European vehicle type category is the SADR. "SADRs [, Autonomous Delivery Robots (ADRs) and Personal Delivery Devices (PDDs)] are pedestrian sized robots that deliver items to customers without the intervention of a delivery person" and mainly drive on sidewalks (Jennings & Figliozzi, 2019, p. 317). SADRs are equipped with cameras, sensors and the Global Positioning System (GPS) to map their surroundings and move around safely (Jennings & Figliozzi, 2019; Ghaffarzadeh, 2019). An extensive explanation on the functioning of ADSs is given in Section 3.6.2. When an SADR has arrived at its destination, a customer is informed via a smartphone application and able to unlock and open the cargo bay lid to receive the shipment (Boysen, Schwerdfeger, & Weidinger, 2018). Currently, SADRs are not fully operated autonomously yet. Suppliers also have remote teleoperator centres where remote human operators take control when delivery robots encounter situations they cannot handle with high confidence (Ghaffarzadeh, 2019). According to SADR technology provider Starship Technologies (n.d.) their robots operate autonomously for 99%. As mentioned earlier in this report, self-reported statistics must be placed in the context of the operating conditions that applied to the delivery robots during their movement. It cannot be said with certainty whether 99% is a very good or a disappointing score. The maximum speed of SADRs is generally limited to 6 km/h, which roughly equals the average walking speed of a pedestrian. This speed limit is there to increase safety, give robots more thinking time, and give remote teleoperators the opportunity to intervene (Ghaffarzadeh, 2019). The research positioning of this section is visible in Figure 5. The SADR of study in this research is visible in Figure 6, the technical specifications can be found in Appendix E.

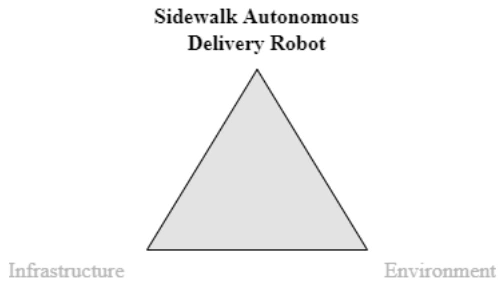


Figure 5: Research positioning SADR

Figure 6: Example of Cartken SADR, Rosie. Photo by Jonathan van Rijn.

Besides the fact that delivery robots differ from established vehicle types in their technical characteristics, a delivery robot is self-driving. Self-driving vehicles are still very new innovations. In addition, there are not yet any good methods of testing that indicate or proof whether a self-driving vehicle is safe or not. The current approaches for safety testing of autonomous vehicles (AVs) are first to conduct simulations, then perform closed circuit tests, and finally on-road tests (Shetty, Tavafoghi, Kurzhanskiy, Poolla, & Varaiya, 2021). Each approach has its advantages and limitations. Simulated testing, for example, can only be an approximation of real-world driving conditions. According to Shetty et al. (2021, p. 712) “small differences in the simulated and real-world environments can translate to large errors in the interferences drawn about safety performance". Closed-circuit tests partially overcome this problem, but lacks the total conditions that can be encountered at the environments in which ADS systems will be deployed. That is why on-road tests are the most instructive.

Most research in the field of autonomous driving focuses on autonomous driving passenger cars. These are vehicles that drive on public roads at considerable speed, which induces certain safety risks. In addition, this form of autonomous driving has not yet been approved. A delivery robot differs from an autonomous driving passenger vehicle in its dimensions, speed and operating domain. A delivery robot operates on the sidewalk instead of on public roads, and will therefore have different interactions with different road users than an autonomous vehicle has. In addition, the sidewalk is a much less standardised road surface than public roads, connecting roads or motorways. There is no standard material, standard width, standard markings, et cetera. The space on the sidewalk is also used inappropriately: in some places, it is cluttered with bicycles, shared mopeds and scooters; when the weather is nice, people set up picnic tables outside; and the sidewalk sometimes serves as a storage area for building materials. Section 4 will deal with this more in depth.

Starship Technologies is among the largest providers of SADRs and offers zero-emission autonomous delivery in many parts of the EU, UK and the US in cities, academic campuses and industrial campuses (Starship Technologies, 2022). On a daily basis Starship robots complete numerous deliveries in a row 100% autonomously, without the intervention of a teleoperator, involving over 100,000 road crossings every day. To date, Starship states its delivery robots have completed more than 2.5 million commercial deliveries, during which more than 5 million kilometres have been travelled. It must again be noted that Starship does not offer any insights in where exactly these kilometres were driven and under what conditions. Also the cost of a Starship delivery is now said to be lower than the human equivalent, however there is no (independent) calculation that proves this claim. Starship is able to make Level 4 deliveries everywhere it operates since 2018. Level 4 refers to the six levels of automated driving established by the Society of Automotive Engineers (SAE), ranging from no driving automation (Level 0) to full driving automation (Level 5) (SAE International, 2021a). This will be dealt with in more detail in Section 3.6.1.

Although in some countries delivery robots are making their appearance more often, this development is lagging behind in the Netherlands. This can be linked to the earlier identified issues with vehicle type approval. Dutch authorities are are reluctant to test illegitimate vehicles on public roads. Regarding SADRs in the Netherlands three different experiments can be distinguished trough time. An experiment by Starship Technologies with Domino’s Pizza Enterprise Limited on the public sidewalk in the area of Rotterdam Hoogvliet should take place in 2018. A SWOV advisory report was drawn up for this practical trial (Petegem, Nes, Boele, & Eenink, 2018), however because no official permission had been obtained the experiment did not proceed. In 2020, a practical trial took place with a robot from Airlift Systems on the enclosed grounds of Breda University of Applied Sciences (Cartens, 2020). Currently, March 2022, there is an ongoing practical trial with a robot from Cartken on the enclosed grounds of the Rotterdam Erasmus University (Erasmus University Rotterdam, 2021). Because the latter two tests were both on closed sites, no SWOV or RDW assessments were neccessary and are therefore not available. It is currently not permitted to test SADRs on the public sidewalk in the Netherlands.

3.4 Experimental Act

As described at the end of Section 3.3 autonomous driving systems are only allowed on non-public roads, such as enclosed test grounds. To give an insight into the current way in which testing of self-driving systems takes place in public, this section explains how autonomous passenger vehicles are allowed on public roads in the Netherlands.

According to the RDW (2019) since July 1, 2019, the Experimental Act has existed in the Netherlands. This law makes it possible to carry out tests with self-driving vehicles without a driver in the car, who can sit remotely in a control room. Under strict conditions, self-driving vehicles may be tested on public roads. Applications for driving without a driver are received by the RDW. The RDW informs the Ministry of Infrastructure and Water Management about the application and then enters the process with the Ministry. The Ministry examines which laws and regulations need to be exempted from the Road Traffic Act. The RDW focuses on the technical requirements of the vehicle to ensure that road safety is not compromised. The advice of the police, the road authority and SWOV is also requested. RDW advises the Ministry about the granting of the licence.

The foregoing seems hopeful for testing autonomous vehicles. However, in practice, few projects are carried out on public roads. In addition, the geographical locations where experiments were allowed to take place are often limited in terms of the operating domain, so that little knowledge can be transmitted to other operating domains. This inhibits the extent to which knowledge can be gained about autonomous driving. One of the difficulties identified on the basis of the Experimental Act is that “a risk analysis in relation to environmental factors and the route of the Operational Domain” should be carried out (Regeling vergunningverlening experimenten zelfrijdende auto, nr. IENW/BSK-2019/134685, 2019). To date, there is no accepted standard methodology to conduct this risk analysis and to describe the Operating Domain, let alone for SADRs. As a result, such analyses are often described in qualitative terms and, in addition, it is often unclear to applicants on what basis their application is evaluated.

Since 2021, the Dutch legislation has been further adjusted, allowing autonomous passenger vehicles to drive on public roads. The law now allows vehicles to drive autonomously on motorways with separate carriageways up to a maximum speed of 60 kilometres per hour. In practice, this means that driving in a traffic jam can be taken over by the vehicle (Bright, 2021). Other forms of autonomous driving on public roads are hereby explicitly forbidden. It must be noted that most of the (grey) publications on the admission and testing of new vehicles in the Netherlands focuses on autonomous passenger vehicles to be operated on public roads. There are no proper publications that address both issues for delivery robots that are meant to drive on the public sidewalk in the Netherlands.

3.5 Stakeholders

Based on the previous sections the actors and their responsibilities or interests towards admission of SADRs as a new vehicle on public sidewalks in the Netherlands can be set out.

European Commission
The European Commission can write legislative proposals about adjustments or expansion of the existing categories for vehicle type approval after internal agreement. These are then submitted for approval to the European Parliament and the Council, who together determine the EU legislation (European Commission, n.d.). When, at European level, delivery robots are incorporated into legislation and are permitted, the Netherlands must adopt these rules. The Commission could impose admission and permanent requirements to SADRs.

Minister of Infrastructure and Water Management
The Minister of Infrastructure and Water Management has the authority to decide on adjustments and appointments at the national level concerning vehicle type approval, as underwritten with the case of the special moped vehicle category. The same authority can be used in shaping approval procedures for SADRs.

RDW
The RDW carries out its tasks on behalf of the Ministry of Infrastructure and Water Management. The RDW's tasks mainly include: the admission of vehicles and their components, supervision and enforcement of vehicles and components, the registration of vehicles and components, the provision of information and the issue of documents (RDW, n.d.-b). As described earlier, the RDW has a role in advising on technical vehicle safety. In relation to SADRs, the RDW will have to issue an advice or, in the future, test whether SADRs comply with the technical safety requirements set for a delivery robot. Since the accident with the Stint it is likely that the RDW will be hesitant towards admitting new and unproven vehicles to public sidewalks. Next to the admission procedure, the RDW has a major role in checking the permanent requirements of vehicles. If permanent requirements for SADRs are drawn up, the RDW will have to monitor them, which means an increase in the amount of work for the inspectors.

SWOV
SWOV is the national scientific institute for road safety research (SWOV, n.d.). Their goal is to use knowledge from scientific research to contribute to safer road traffic. They do this by helping policymakers and other road traffic professionals to answer questions regarding road safety. As described earlier, the Ministry of Infrastructure and Water Management can request an optional safety report from SWOV to assess the safety of an innovation, for example an SADR, as part of an admission procedure.

CROW
CROW is a transportation, infrastructure, and public space technology platform (CROW, n.d.-a). It is a non-profit corporation in which the government and businesses collaborate on the design, construction, and management of roads and other traffic and transportation facilities. CROW focuses on disseminating knowledge goods to all target groups and is active in research and regulation in the Netherlands, especially aimed at municipalities and provinces. With the Dutch Roads taskforce for self-driving cars, CROW is active in exploring all facets of autonomous driving and connected vehicles (CROW, n.d.-b).

Police / enforcement
The police are in charge of enforcing road safety and fulfil various tasks in relation to this. As described earlier, the police play an important role in enforcing safe driving behaviour by drivers and the observance of traffic rules by all road users. More specifically in relation to SADRs, the police also enforce on wrong-way parking on the sidewalk, on people littering the sidewalk, on people committing vandalism, etcetera. However, the police can never enforce on all places at the same time. This is why it can happen that a sidewalk that is supposed to be freely passable for an SADR is not passable or is passable only to a limited extent. An additional problem is that there is currently a shortage of police personnel in the Netherlands (Politie, 2022). This makes it difficult to perform all the different tasks at all locations at all times, let alone when the deployment of SADRs adds more tasks. Another unknown aspect is that it is currently unclear how a traffic violation by a delivery robot can be dealt with.

SADR manufacturer
The manufacturer of a delivery robot is, on the one hand, responsible for the technical vehicle safety of the SADR, and on the other hand, the manufacturer will be responsible for the autonomous driving system contained in the SADR. Manufacturers will do everything they can to see that delivery robots are soon admitted to public roads, additionally manufacturers will want to impose as few rules as possible to drive on the public sidewalk. On the other hand, manufacturers are currently reluctant to start the Dutch approval procedure because the costs are now for the applicant and can easily excess 100,000 euros.

Road authority
"The road authority is responsible for the construction and maintenance of the road and road equipment (e.g. speed bumps, refuge islands, lighting columns, bus locks). In most cases, this is a government body, such as the national government, the Province, the municipality or the Water Board." (ANWB, n.d.-a). SADRs drive on the sidewalk, which is managed by the road authority. However, for sidewalks no harmonized regulation or regulation at all exists. It may happen that infrastructure deteriorates, making it harder for a delivery robot to move around. In this case, there will be an interaction between the operator of the delivery robots and the road authority. There is a potential conflict of interest when a municipality is simultaneously the road authority. A municipality may want to improve its image by profiling itself as progressive in innovation. As a result, it can be decided too quickly and without proper evaluation of the associated safety risks to allow SADRs to enter a road section while this is not desirable given the characteristics of the road or characteristics of the operating

environment.

Pedestrians
Pedestrians in this context refers to all participants in sidewalk travel. Children, people in wheelchairs, pedestrians and so on. This is the group that will have the most interaction with SADRs because they use the same road surface. The biggest safety risks for road users also lie in the interaction between people and the delivery robot. In addition, this is the group that often causes the improper use of the sidewalk.

Public road users
Public road users in this context refers to all participants in public road traffic. At some point during a delivery, an SADR will have to cross a road to move from sidewalk section to sidewalk section. This is similar to a pedestrian who sometimes has to cross roads to continue his or her journey. While crossing, an SADR will encounter different road traffic than the average sidewalk situation. This includes cyclists, motorists, public transport, etc. On this stretch of public road, the differences in speed between different road users can be significant, and an unpredictable manoeuvre by an SADR poses possible safety risks.

3.6 Automated driving

In this research, the industry-wide classification standard for the various levels of driving automation by the SAE International will be cited several times in order to be able to state the level of self-driving technology and to compare different studies. Therefore, in this section an overview will be presented of the six levels of driving automation as drawn up by the SAE International (2021a) and the technology behind the functioning of autonomous driving systems will be explained.

3.6.1 SAE levels of driving automation

The six levels of driving automation include the following (SAE International, 2021a):

Level 0: No driving automation
Level 1: Driver assistence
Level 2: Partial driving automation
Level 3: Conditional driving automation
Level 4: High driving automation
Level 5: Full driving automation

The graphic established by the SAE International to visually clarify the included features per driving automation level is presented in Figure 7 and distinguishes between what the human in the driver's seat has to do regarding the driving tasks and what the automated driving features do.
The levels have been established from the point of view that human-controlled vehicles are gradually evolving into fully autonomous ones. Sidewalk Autonomous Delivery Robots mainly appear in SAE levels three to five. The translation of what a human in the driver's seat has to do relates to what a teleoperator in a control room has to do. Delivery robots have not been created from a point where they were first manually controlled and have gradually become more autonomous, but developed to drive as autonomously as possible. The presence of a teleoperator is mostly there for safety reasons. The ultimate objective is to be able to drive SADRs without the interference of a teleoperator.

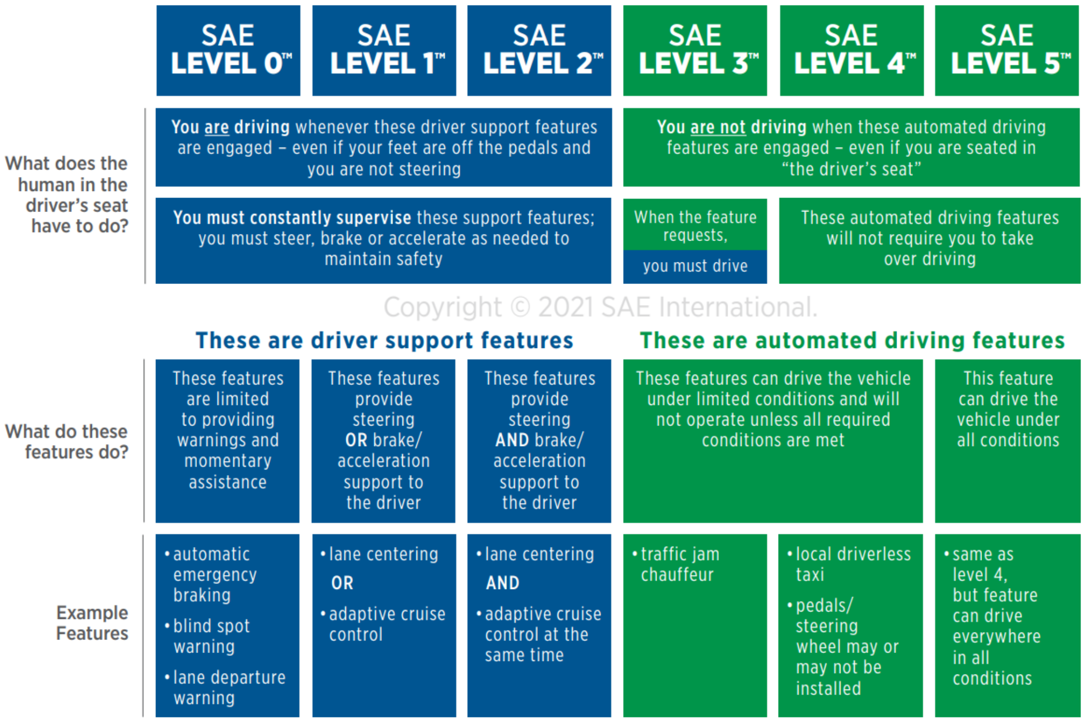


Figure 7: Visualization of SAE levels of driving automation

3.6.2 Functioning of Automated Driving Systems

An SADR, and ADS in general, is made up of several hardware and software components that, together with an operating system, ensure that the driving task can be performed. In this section, the functioning of this operating system will be explained.

In more detail: an ADS relies on "sensors, actuators, complex algorithms, machine learning systems and powerful processors to execute software" (Synopsys, 2022). The different functions these components have to fulfil are position determination, object detection, path/trajectory determination, path/trajectory execution of the vehicle and communication.

Exact position determination of a robot can be achieved in several ways, such as via the GPS, Satellite-Based Augmentation Systems (SBAS), Wide Area Differential GPS (WADGPS) and Global Navigation Satellite System (GNSS). In order to verify this determined position, often digital maps are used that are compared to reality on the basis of current sensor data or by using external sensors in the operating environment to verify landmarks (Kocsis, Zöllner, & Mogan, 2022). In odometry, data from sensors, for example that measure the rotation of the wheels, is used to determine the change in location of a robot in a unit of time. Kuutti et al. (2018) conducted a research on state-of-the-art position determination techniques and their potentials for autonomous vehicles.

To detect possible objects in the operating domain of vehicles Radio Detection And Ranging (RADAR), Light Detection And Ranging (LiDAR) and cameras with image processing software are used. Lidar is a method to scan the environment with lasers. Lidar systems pulses beams of infrared laser light and use the laser light reflection and speed of light to calculate distances to objects in the operating domain (Udacity, 2021). Radar works in a similar way by transmitting and receiving radio waves. In comparison to radar, Lidar is more often unreliable at nighttime or in inclement weather, but in general more accurate in detecting nearby objects with high precision (Manjunath, Liu, Henriques, & Engstle, 2018; Udacity, 2021). By using cameras and image processing software objects can not only be sensed, but also visually identified and classified into different possible categories of objects. This categories can subsequently be used to to predict behaviour, movement and other characteristics of the object nearby. The current operating system of the SADR of focus in this research (see Appendix E) uses the robot information about the velocity to predict where the robot would be in 0.2 seconds. This information is used together with the information from the depth cameras to check if a collision is going to occur. If that is the case an emergency stop is triggered, where the robot stands completely still in between 10 and 30 centimetres.

Based on the exact location of the robot, objects in the vicinity of the robot and the target location of the robot, an optimal path is determined. This path is then communicated to the robot, which follows the proposed path as accurately as possible by means of controlling the actuators. In the study of Kocsis et al. (2022), a separate system is used to determine the trajectory of the delivery robot. The aforementioned would not be possible without several forms of communication. The sensor data observed by the sensors must be communicated to a central point in order to reach a decision based on supplemented external information from databases or real-time data. Examples are the digital maps mentioned above or the external sensors to verify landmarks. An SADR must also be able to communicate with the application or portal used by the customer. Another important part of communication for an SADR is the necessary communication with the remote teleoperator in the control room. When a delivery robot needs remote assistance, it is important that the teleoperator has all the accurate information it needs to make an informed decision. This includes all observed information using the lidar and radar, current camera images and ambient noise. In the opposite direction it is important that the decision made by the teleoperator is communicated to the robot in a short time. The robot should respond immediately to prevent that the environment has changed to an unsafe situation, causing undesired situations. The last form of digital communication is communication with other smart devices and sensors in the environment, for example through vehicle to everything (V2X) communication. A more detailed explanation can be found in Section 6.2.6. A non-digital way of communication is the communication to other road users by using light and sound signals.

As mentioned, automated driving system functions rely on complex algorithms and machine learning systems. A fully autonomous driving system, SAE level 5, must be able to handle all driving tasks and deal with all possible traffic situations. This means that the controlling and predictive model behind this functioning is extensive because it must be able to make a safe driving decision at all times. The disadvantage of machine learning algorithms is that it is a black-box model and little to no transparency in the formation of a decision is known (TNO, n.d.). Small predictive black-box models based on a limited number of input parameters can be manageable, but more sophisticated black-box models, as needed for ADS, are mathematically complex, and consequently are quite often not understandable by the end-users of the model or outsiders such as regulators (Martin, Winkler, Grubmüller, & Watzenig, 2019). Based on such models, it is impossible to determine how an ADS will react in specific circumstances and it is also impossible to check the quality of the programmed code in advance. Regulators are therefor reluctant to allow such black-box driving systems on public roads. Companies behind ADSs do currently not provide insight into their complex algorithms for competitive reasons.

3.7 Consequences of a remote teleoperator

As mentioned in Section 3.3 SADRs are not fully operated autonomously yet. Suppliers have remote teleoperator centres where remote human operators take control when delivery robots encounter situations they cannot handle with high confidence. The second application of the remote teleoperator centres is remote assistence, where a teleoperator grants an SADR permission to cross an intersection. Although at first glance this may seem safe for the sake of safety, it has several consequences. Teleoperation can as well be a viable solution to expand the ODD of level 4 and level 5 autonomous vehicles (AVs) (Thorn, Kimmel, Chaka, & Hamilton, 2018).

A remote teleoperator is significantly different from a safety operator present near the vehicle. If control is transferred to the remote operator, he or she must make a safe decision based on the information available at that time. This includes all observations of the robot supplemented with information from external databases and sources. The teleoperator gives the control instructions from a distance. From a safety perspective, it is important that the time between sending and receiving all data to the operator and sending, receiving and executing the control instructions by the robot is as short as possible, so that there are no changes to the environment and the traffic situation remains the same. This is significantly different from a safety operator who is in the vicinity of the robot and must make a safe decision based on his own observations and can control a vehicle in real-time. The latter can better be compared with the current situation of autonomous passenger vehicles with a safety operator behind the wheel. If the ADS cannot handle the driving task, the driving task is returned to the driver, who actually holds a driving licence to drive a car.

The example of Section 3.3 offers many follow-up questions. Starship Technologies states that they are able to drive autonomously for 99%, which indicates SAE level 4 or 5 (Figure 7). This means that 1% is not driven autonomously, which indicates control to a teleoperator and a regression to SAE level 3. The first question is whether the SAE classification is dynamic? If this is not the case, is SAE level 4 or 5 used to describe Starships delivery robots because that is the average or median of the performance, or is SAE level 3 used because that is the lower limit of the current system? In addition, there exists no such thing as a remote teleoperator licence, which raises the question of how SADR driving safety can be assessed if the remote operator is in control of the

SADR? Should this be treated as a separate case or as inextricably linked to an autonomous driving SADR? In this case, should an SADR still be judged on the extent to which it can drive safely or is the question actually: how well is a remote teleoperator able to safely drive an SADR? And what are the consequences of this in the event of an accident? The existence of these theoretical questions show that there are many details that need to be considered before SADRs will be allowed to enter the public sidewalk. This research focuses on the design of a dynamic assessment framework for SADRs. Through controlled testing and practical experience with delivery robots, the theoretical questions listed can be examined in a real-life context.

3.8 Conclusion

The goal of Section 3 is to provide an overview of the context of this research and provide an answer to the first sub-question:

1. *What challenges and needs exist in the Netherlands to test SADRs on the public sidewalk?*

Currently, SADRs cannot be categorised into an existing vehicle category, which means that processes and benchmarks around the safety assessment of SADRs do not exist and SADR performance cannot be tested on the public sidewalk. Additionally there is no accepted complete and objective method to describe the Operating Domain for SADRs. On the one hand, this lack means that the risks associated with the deployment of SADRs cannot completely and objectively mapped out. On the other hand, this lack means that it is not clear under what conditions safe kilometres are driven in other countries, and that no conclusions can be drawn from those safe kilometres for the Dutch sidewalk case. Lastly, SADRs are controlled by complex black-box models that regulators have no insight into and SADR providers do not want to give insight into from a competitive perspective. There is as yet no good way to assess such black-box models as safe, or to assess SADRs on safe performance when a human teleoperator is in control of a delivery robot.

The most obvious solution for a non-existent vehicle category, the creation of a new vehicle category, is a time-consuming process that takes several years. The last time this process was undertaken in the Netherlands critical errors have been made, resulting in new vehicles being admitted too easily without passing the required safety tests. The wrongful admission of the Stint in this way, and the subsequent accident in which young children died, have ensured that the admission of vehicles can no longer take place in the Netherlands without a suitable vehicle category. Knowledge about self-driving systems is nowadays mainly gained on public roads with autonomous (passenger) vehicles, the niche of the delivery robot is not taken into account. There is currently no independent, accepted method to map the risks associated with the deployment of SADRs or ADSs. The hypothesis is that various static and dynamic environmental characteristics affect SADR performance. However, these risks have never been mapped for delivery robots. Because a complete and objective methodology to map these risks is missing, different parties in the admission field foresee different safety risks regarding SADRs operating on the public sidewalk, which is challenging to overcome when field trials are set up.

4 Objective and complete Operational Design Domain description

In this section, an objective classification methodology for the Operational Design Domain (ODD) of Sidewalk Autonomous Delivery Robots (SADRs) will be discussed and a way of comparing different geographical areas in the Netherlands on the basis of infrastructural characteristics will be sought for. First, the literature on describing the ODD regarding Automated Driving Systems (ADS) on public roads will be reviewed according to the literature review methodology proposed by Snyder (2019), after which the most suitable theory will be motivated for the case of the SADR on the public sidewalk. In the broader overview of this study, this section focuses on the bottom two sides of the triangle in Figure 8: the structured categorisation of the infrastructure and environment that an SADR faces.

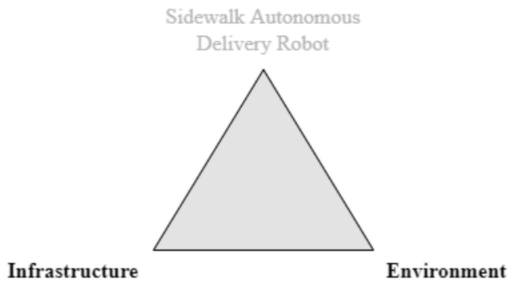


Figure 8: Research positioning ODD

4.1 Literature study methodology

According to Snyder (2019) relevant literature review is essential for all research disciplines and all research projects. By doing so an author describes previous research to map and assess the research area to motivate the aim of the study and to justify the research questions and hypotheses. However, for a literature review to become a proper research methodology, proper steps need to be followed and action taken to ensure the review is accurate, precise and trustworthy. This literature study follows the four phases of literature review, as proposed by Snyder. In phase one the scope of the review is designed; in phase two will be determined when literature can be considered relevant, where after in phase three the relevant literature will be searched and selected. In phase four the actual analysis and structured review will be presented.

4.1.1 Scope

The scope of this literature review is to identify different scientific methods to objectively define the Operational Design Domain of an SADR. By exploring all possibilities and determining the advantages and disadvantages of the different methods regarding the description of the ODD of ADS, the ODD of an SADR can subsequently be defined. For the purpose of this research, the ODD in general will be briefly introduced.

4.1.2 Literature Relevance

For this literature study primarily Google Scholar is used as search engine. Literature relevance, regardless of exact content, is classified as follows. A paper is considered relevant when it is published in a scientific journal and is peer-reviewed. A paper is considered most relevant when it is published in a top journal, based on SCImago Journal Ranking and Impact Factor. Papers of a more recent nature are assigned a higher relevance in view of the rapid technological developments in the automation sector. To assess papers on content relevance, the keywords in Table 3 are used. Note that the keyword concepts in Table 3 have been extended with the 'descriptive method' concept after having identified the first relevant publications. Grey publications are used to clarify matters if no scientific publication is at hand.

Table 3: ODD literature study concepts and keywords

Concept	Keyword(s)
Innovation	delivery robot automated driving system automated vehicles
Infrastructure or environment	operational design domain road sections factors infrastructure

Table 3 continued from previous page

	environment
Classification	classifying classify
Descriptive method	ontology operating envelope layer model

4.1.3 Literature search and selection

The search queries from Table 4 were used to find publications that match the topic. Based on the abstract, conclusions and recommendations, it was determined whether a paper would be a valuable addition to the literature list. If so, the publication was added to the list and the contribution of the paper was briefly noted, to keep the overview. The principle of 'backward snowballing' was applied to recent literature to find theories that current researchers are building on and to see what the developments have been in recent years. The list resulting from both steps was then thoroughly read and it was determined which papers would be a valuable addition in terms of content given the scope of the literature review. Because the purpose of this part of the literature study is to provide an overview of different ODD classification methodologies, papers that showed too much overlap with each other have been compared on SCImago Journal Ranking, number of citations and Impact Factor if possible. For subjects from the literature review for which not enough qualitative literature was found, it was decided to expand the search term in Google Scholar and to go through the above process again with the literature.

The final literature list for the benefit of this literature study consists of the following scientific publications: (Koopman & Fratrik, 2019), (Riedmaier, Ponn, Ludwig, Schick, & Diermeyer, 2020), (Griffor, Wollman, & Greer, 2021), (Geyer et al., 2014), (Czarnecki, 2018b,c,a), (Cho, 2020), (Scholtes et al., 2021), (Bagschik, Menzel, & Maurer, 2018), (Bock et al., 2018), (Schuldt, 2017) and (Khatun, Glaß, & Jung, 2021).

Table 4: Google Scholar search queries

Search query	Publications (#)	Search date
"delivery robot" AND "operational design domain"	11	16-3-2022
"classifying" AND "operational design domain" AND "factors" AND "infrastructure"	109	16-3-2022
"classify" AND "operational design domain" AND "automated driving system"	179	16-3-2022
classify AND "road sections" AND "operational design domain" AND "automated vehicles"	53	16-3-2022
"ontology" AND "classify" AND "infrastructure" AND "operational design domain"	49	17-3-2022
"operating envelope" AND "automated driving system"	29	17-3-2022
"layer model" AND "automated driving system" AND "operational design domain"	39	17-3-2022

4.2 The Operational Design Domain

Per SAE standard J3016, published in 2021, the Operational Design Domain for a driving automation system is defined as "operating conditions under which a given driving automation system, or feature thereof, is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics" (SAE International, 2021b, p. 32). The Operational Design Domain has been created to distinguish between the operating domains where the self-driving system is able to control the vehicle and where the driver should control the vehicle. This is a result of the development of self-driving technology, where self-driving systems were not able to drive safely in autonomous mode in specific operating conditions for a long time.

The ODD is also used in the evaluation of self-driving systems. A major aspect of overall system validation is ensuring that autonomous cars will function satisfactorily in their intended operational context (Koopman & Fratrik, 2019). The greatest challenge in safety assessment is that road traffic is an open parameter space in which an infinite number of different traffic situations can occur (Riedmaier et al., 2020), and that it is not viable to test all possible combinatorial road traffic situations, for example by simulation or on test sites, to classify an autonomous driving system as safe. According to Koopman & Fratrik (2019) to ensure that training and testing of autonomous vehicles is complete, this requires at least ensuring that all aspects of the ODD have been addressed. This can either be done by assuring safe system operation or by ensuring that the system is capable of detecting and mitigating deviations from the defined ODD (Koopman & Fratrik, 2019). First of all, it should be mentioned that the current knowledge of describing and classifying ODDs or derivation thereof is limited. There is not yet an objectively accepted methodology to describe or classify factors and elements that make up ODDs. Current publications or practical tests with self-propelled systems often describe the domain in a qualitative way, so that no relative comparison can be made with other domains and the knowledge gained cannot easily be translated to new geographical areas. By classifying the infrastructure and environment according to objective factors, a start can be made to structure the testing and learning of self-driving systems on public roads. Besides the fact that there are few publications on the ODD, there is no approach to classify the ODD on the sidewalk yet. Most knowledge of the Operational Design Domain for a driving automation system has been gained in the light of self-driving (passenger) vehicles. These logically operate on public roads, so nothing is known yet about the operating domain of the sidewalk, where SADRs will operate.

It can be assumed that various static and dynamic factors of the infrastructure and environment influence the extent to which an SADR is able to move safely on the sidewalk (Roh & Im, 2020). To illustrate: driving during the day on a tarmac surface without obstacles will be easier for a robot than driving at dusk on a cobblestone surface with many people moving around at the same time. First the current knowledge on static and dynamic factors that influence the ODD at the public road level will be assessed, to translate and extend these to the case of the public sidewalk. This section will lay the basis for a framework to classify the relative difficulty for SADRs to cope with certain infrastructure and environmental characteristics.

4.3 Current knowledge on infrastructural ODD factors

In this section five different methods to describe, categorize, classify the ODD of the public road for ADS will be discussed. Based on the pros and cons the most suitable method(s) will be used to guide the classification of the ODD for SADRs on the public sidewalk in the Netherlands in Section 6.2.

4.3.1 Operating Envelope Specification by Griffor et al. (2021)

"The concept of an Operating Envelope Specification (OES) is a structured description of environmental factors and elements an automated driving system may encounter during its operation and relates these to the Operational Design Domain of ADS-equipped vehicles" Griffor et al. (2021, p. 4). The operational design domain again relates to the conditions under which the vehicle is intended to function. To function properly, the ADS-equipped vehicle has to be aware of its current operating environment. Griffor et al. argues that there is a need for an abstraction that describes the operating conditions in a way that is measurable, relates these measurements to concerns about ADS operation and supports reasoning about operating conditions both off-board for assessment of vehicle behaviours and on-board for decision making. The OES builds on three types of information: OES Nominal, OES Actual and OES Reference.

OES Nominal includes roadway characteristics such as roadway components, their physical dimensions and transit paths. The associated relevant parameters and their nominal values are included in the nominal operating conditions. OES Actual contains real-time information on changes to OES Nominal and includes the changes in the nominal value of associated parameters. OES Reference is a compendium of operation condition names and parametrized definitions. Content includes guidance on inventory of roadway characteristics, geometry, angles, controls, design speeds/sight distances and markings. OES Reference can be defined as the vocabulary of the OES (Griffor et al., 2021).

In a case example Griffor et al. used SAE International (2020a) best practice to describe the ODD. The description consists of the following ODD categories: route network, sun angle, precipitation, operating speed, wind, lane width, road surface conditions, connectivity and rush hour. It is not mentioned why the authors chose for the specific metrics used to describe the ODD. By showing what the OES Reference looks like for a general intersection, the OES Nominal and OES Actual can be formulated to describe a specific intersection.

The benefit of the proposed OES methodology is that it provides guidance on describing the operating conditions under which a vehicle will operate. The methodology acknowledges that there can be temporary modifications to an existing infrastructure. By using the taxonomy of SAE International the industry standard is respected. What is not clear from this methodology is how to cope with the dynamic environment of the infrastructure and traffic. It is not prescribed how to define, measure or analyse this aspect of the urban environment. The ODD categories used to classify the infrastructure are high level and the boundaries of the functional operating area are therefor indistinct. It is not clear whether the OES is described on a road section level, road level or district level.

4.3.2 Ontology for a structured representation of the ODD by Geyer et al. (2014)

One of the first researchers to acknowledge that no consistent terminology for vehicle automation existed were Geyer et al. (2014). The authors proposed the following fundamental ontology to structure the ODD for automated vehicle guidance, presented in Figure 9.

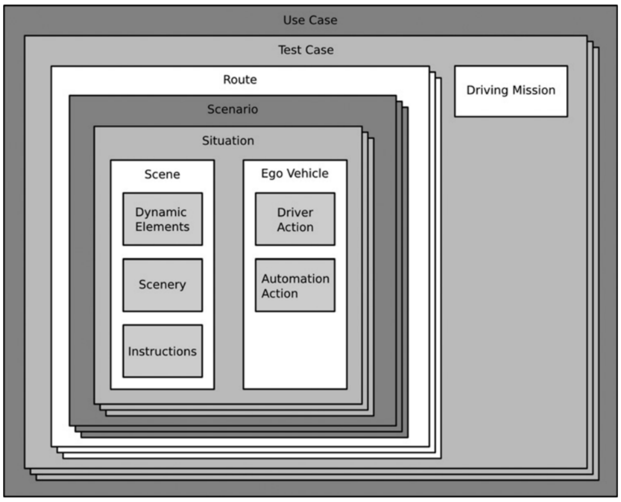


Figure 9: Fundamental ontology for assisted and automated vehicles guidance by Geyer et al. (2014).

The ontology is mainly built on the consecutive concepts of the ego-vehicle, the scenery, the scene, the situation, the scenario and the route (Geyer et al., 2014). The ego-vehicle consist of the vehicle itself with the driver and optional automation functions to drive the vehicle. The scenery is a structured collection of static elements and describes the static infrastructural elements. The scene is defined as a combination of the former described scenery, expanded with dynamic elements and optional driving instructions. The situation is a reflection of the ego-vehicle within a specific scene. More general a scenario is built of many different situations and a route consists of multiple scenarios. This ontology can be used to guide ODD classification in a structured manner, however it is not mentioned if the description or classification of ODDs should be qualitative or quantitative, as well that it is not mentioned in what level of detail the different elements in the ontology should be described.

4.3.3 Czarnecki's Operational World Model ontology for Automated Driving Systems (2018b)

Czarnecki has published one of the most, if not the most extensive works on an Operational World Model ontology for Automated Driving Systems. The ontology defines all elements that occur in road environments, including their attributes, relationships, and if present, behaviors. The focus of this research is on paved, structured urban and rural roads in North America. The publication overlaps with Geyer et al.'s research on the scene, situation and scenario. Divided into two publications, Czarnecki described the road structure (Part 1) and the road users, animals, other obstacles and environmental conditions (Part 2) of automated driving systems in detail (Czarnecki, 2018b,c). The road structure in Part 1 covers:

- Road structure
  - Road type and capacity
  - Road surface type and quality
  - Road geometry
  - Cross section design
  - Road traffic control devices
  - Pedestrian crossing facilities
  - Cycling facilities
  - Junctions
  - Railroad level crossings
  - Bridges
  - Tunnels
  - Driveways and driver access points
  - Temporary road structure

Part 2 includes the following elements:

- Road users
  - Road user classification
  - Road user behaviour

- Animals
- Other obstacles
- Environmental conditions
  - Atmospheric conditions
  - Lighting conditions
  - Road surface conditions

A detailed overview, covering all elements in the established ontology, can be found in Appendix D. The elements have been mapped in order to identify all relevant road environments in a simulation. The benefits of this ontology is that it is extensive and can be used to systematically to map all existing road situations and, based on the extensive categorisation, to explore in simulations the effect of road configurations on the performance of an ADS. A drawback of this ontology is that it is qualitatively possible to map the infrastructure and environment, but there is not yet an accepted method to measure it quantitatively in all areas in order to determine the impact on the performance of the ADS. In addition, the presented ontology is very detailed for the (North American) road case, but it does not provide any guidance on how to use the ontology for other ODDs, Lastly, the usability of this ontology has not yet been demonstrated by other scholars.

4.3.4 ODD determination based on relevant conditions (Cho, 2020)

Cho (2020) recognises that currently the automated driving technology is not yet capable of performing the fully automated driving task at SAE level 5 (Section 3.6.1). Therefore, he argues that because there are constant shifts of performing the driving task from the ADS to the human driver, the granularity of ODD conditions should be chosen so that it is clearly observable and distinguishable by the human operator in the vehicle. In this way, the human operator will be able to know with a higher degree of certainty when to perform the driving task. Relevancy of conditions in this case means the ability of a human operator to observe and distinguish conditions. The identified relevant conditions are presented in Table 5.

Table 5: Relevant ODD classification conditions by Cho (2020)

Condition Classes	Condition Variables
Road Type	Limited-Acces Highway
	Rural Highway
	Urban Arterials / Citystreet
Light Condition	Day
	Night
Weather Condition	Clear
	Precipitation
Prior Speed	Velocity (Continuous)

The advantage of this method is that it is easy to understand for a human driver in the vehicle, which is relevant given the current technical status of ADS. An additional disadvantage is the lack of detail of the method, which does not allow for nuance in the analysis between overlapping traffic situations. In addition, the method mainly describes binary condition variables, and the dynamic character of other road users is not included in this analysis. Due to the high level of abstraction, however, the method is translatable to other geographical areas.

4.3.5 6-Layer Model by Scholtes et al. (2021)

A detailed and structured way to approach ODDs is by using a layered model. Schuldt (2017) designed a 4-layered model to structure environments for automated driving systems. The model was first extended by a fifth layer (Bagschik et al., 2018), and then by a sixth layer by Bock et al. (2018). Bock et al. adapted the 6-Layer Model to describe motorway scenario's. Scholtes et al. (2021) built on this work to describe urban scenario's with the 6-Layer Model. The 6-Layer Model is used in the concept as a tool to structure influencing factors on autonomous driving systems, allowing for the formation of scenario equivalence classes and, as a result, the selection of appropriate test methods (Scholtes et al., 2021). The six layers present in the model are shown in Figure 10. To ensure that no misinterpretations arise from paraphrasing the explanations, it was decided to quote the explanations of the six layers as described by Scholtes et al. (2021):

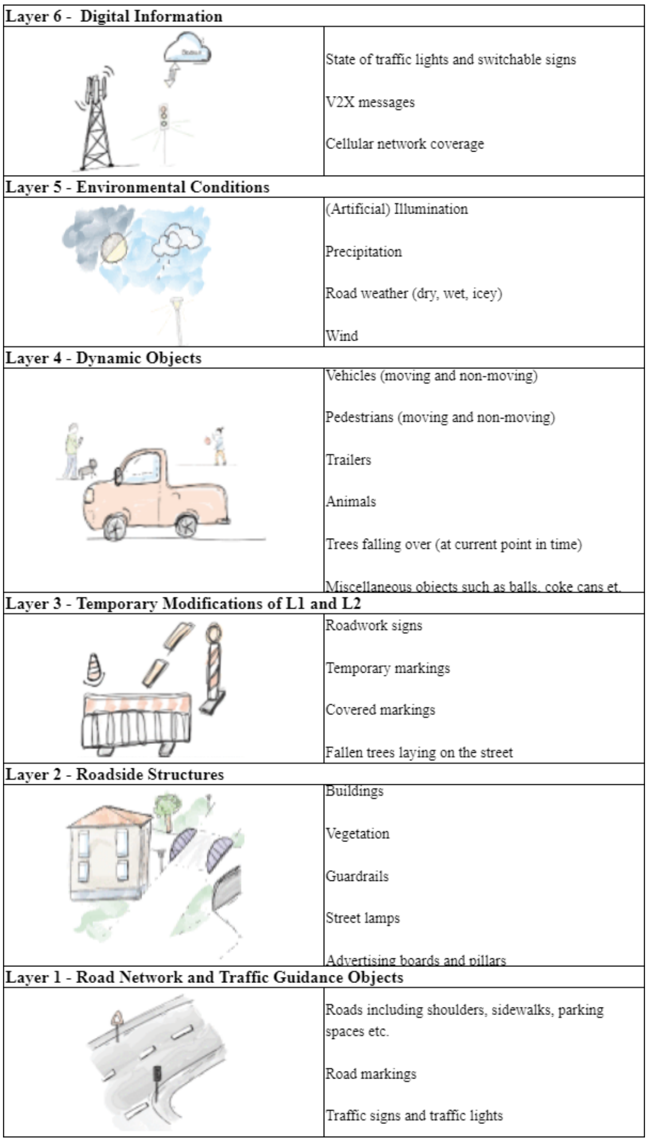


Figure 10: Six layered model for automated road driving systems by Schuldt (2017), Bagschik et al. (2018) and Bock et al. (2018). Figure by Scholtes et al. (2021)

“Layer 1 [of the 6-Layer Model] describes the road network together with all permanent objects required for traffic guidance. Given the road network and traffic guidance objects, Layer 1 summarizes where and how traffic participants can drive. Layer 2 addresses the roadside structures and contains all static objects that are usually placed alongside - and not onto - the road. Layer 3 is comprised of temporary modifications of elements of Layer 1 and Layer 2. Layer 4, ‘Dynamic Objects’, is the first layer that introduces a time-dependent description. It is roughly speaking the ‘traffic layer’ as it includes movable objects whose movements could evolve over time and are described by trajectories or maneuvers. Layer 5 contains environmental conditions. These consist of weather, atmospheric and lighting conditions. Layer 5 also includes road weather conditions. These are weather related modifications of the road surface like dry, wet or icy roads. Layer 6 is defined to focus on all kinds of information exchange, communication, and cooperation on basis of digital data only."

For clarification the 6-Layer Model can also be displayed with differences in spatial and temporal separation between the layers (Figure 11).

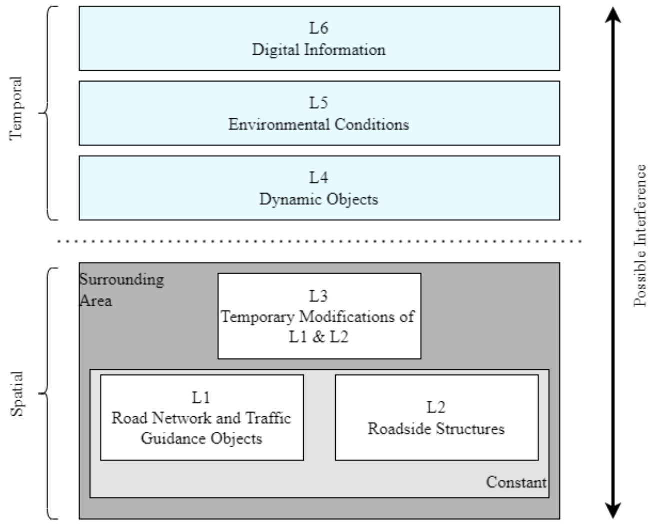


Figure 11: Overview of the six layers including spatial and temporal separation, adapted from Scholtes et al. (2021).

The usefulness of the Layer Model is demonstrated in the research of Khatun et al. (2021). In their research the authors used the first five layers of the 6-Layer Model to examine relevant scenarios that should be used to verify the safety of the highly automated driving function of autonomous vehicles. However the authors state they are aware of the sixth layer, it must be noted that the digital information layer is left out of the scope of their research.

The advantage of this 6-Layer Model approach is that it distinguishes between the static infrastructure (including the possibility of temporal adjustments) and the dynamic environment. The digital connectivity layer demonstrates a holistic approach. By not only discussing the different layers, but also giving guidelines on the correct classification of layers and illustrating this with examples from practice, it is clear how the 6-Layer-Model should be used. The disadvantage of the model is that there are no accepted metrics yet on how to define the entities within the layers and what level of detail should be included accordingly. The model can be applied to all geographic areas and ODDs and is therefore translatable.

4.4 Conclusion

As there is no accepted standard or methodology to map the sidewalk infrastructure for the deployment of an SADR, several methods are described to map the ODD of an Automated Driving System so that a similar approach can be taken for the ODD of a delivery robot. Based on the literature study an answer can be formulated to the second sub-question:

> *2. What state of the art methods can best be used to objectively identify and classify the different static and dynamic elements in the Operational Design Domain of Automated Driving Systems?*

The best methodology to objectively and completely map the ODD of ADS is currently the 6-Layer Model by Scholtes et al. (2021). For operating domains that show overlap with the North American road case, the 6-Layer model can be extended with the work of Czarnecki (2018b,c).

The methodology proposed by Cho is very clear and understandable for a vehicle driver. ODD determination for ADS should be understandable for human drivers because they ought to take control at any moment at the driving task, while for SADRs this is not directly the case. The robot can find itself a safe space on the sidewalk and ask for help of the teleoperator if the ODD is not viable for the robot. The trained teleoperator in the control room can be fully updated on monitored ODD characteristics before driving the SADR. The condition variables by Cho should not be ruled out in the sidewalk case, but some additional level of detail can be included for safety and is possible because SADRs will mostly operate at SAE level 4 and 5 (Section 3.6.1). The publications of Geyer et al. (2014) and Griffor et al. (2021) both succeed in providing structure in classifying elements of the ODD, but both lack in providing guidelines on how to classify elements within the proposed sub parts. The work of Scholtes et al. (2021) mitigates this problem by dealing with practical examples in its publication and establishing guidelines for classification. These guidelines are fairly generic, which makes them easy to translate to the case of the sidewalk for the SADR. In addition, Scholtes et al.'s work also takes into account several sub parts in the form of six layers, and is the only researcher to include the digital layer that results in a holistic overview. Due to the overlap with the purpose of Scholtes et al.' work and this research, nevertheless applied to a different use case, it is assumed that applying the 6-Layer Model within this research also fits this research. As of today no publications exist yet that apply the 6-Layer Model to the Operational Design Domain of a Sidewalk Autonomous Delivery Robot. The only difficulty foreseen in classifying the elements within 'Layer 2 - Roadside Structures' is that for the AV case these elements are actually next to the infrastructure that is being driven on, and for the SADR case these are elements that are on the infrastructure of the SADR: on the sidewalk. To describe the elements within the different layers in detail, a link can be made with the work of Czarnecki (2018b,c). However his publications focuses on elements on North American urban roads, severe overlap between elements exist with the ODD an SADR will operate in.

5 Safe performance risks of SADRs on public sidewalks

In the broader overview of this study, this section focuses on the interaction between the elements of the sides of the triangle in Figure 12 and the associated risks arising from these sides. The risks arising from the SADR itself that are not related to the infrastructure and environment will also be addressed.

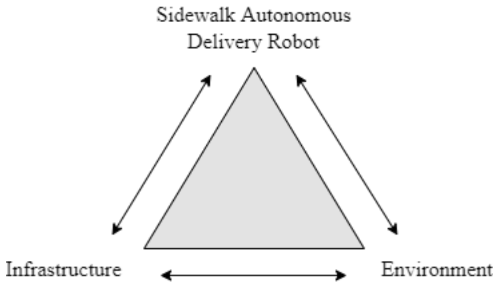


Figure 12: Research positioning risk factors

Because SADRs cannot currently be classified as one of the vehicles in the established vehicle type categories, as described in Section 3.1.2, there are no international standards for delivery robots established by the International Standards Organisation (ISO) or the SAE. Given the parallels between ADS and delivery robots, it is a realistic assumption that international safety standards for delivery robots will be derived from currently accepted standards for ADS. It is therefore of value to list the existing standards with their description, focus and specific safety assessment methodology (Table 6).

Table 6: International standards overview

International standard	Description	Year	Vehicle focus	Safety Assessments	Comment	Sources
ISO 26262 Functional Safety	automotive-specific international standard that focuses on safety components, including an automotive safety life cycle, an automotive specific risk-based approach for safety assurance, and requirements for validation and confirmation measures	2018	All road vehicles except mopeds	Hazard analysis and risk assessment (HARA)	Standard is to be used during the product development phase	(ISO, 2018)
ISO/PAS 21448 Safety of the Intended Functionality (SOTIF)	provides further guidance on applicable design, verification and validation measures to achieve the absence of unreasonable risk due to hazards resulting from functional insufficiencies of the intended functionality or by foreseeable misuse	2019	SAE level 1 and 2 vehicles	HARA, qualitative analyses and Process Failure Modes and Effects Analysis (FMEA)	Standard is to be used during the product design, verification and validation phases. To be replaced by ISO/FDIS 21448 (under development)	(ISO, 2019)
J3018^{TM} Guidelines for Safe On-Road Testing of SAE Level 3, 4, and 5 Prototype Automated Driving Systems	published by SAE International this document provides safety-relevant guidelines for testing automated vehicles (SAE level 3 / 5) in mixed traffic environments and on public roads	2020	SAE level 3 / 5 vehicles with in-vehicle fallback test drivers	TBD	Guidelines specifically exclude remote driving.	(SAE International, 2020b)
ISO 22737 Low-speed automated driving (LSAD) systems for predefined routes	specifies requirements for the operational design domain, system requirements, minimum performance requirements, and performance test procedures that may be relevant for validating the safety of LSAD systems for operation on predefined routes	2021	Low-speed automated driving systems at SAE level 4	-	Low speed of intended systems is below 32 km/h. Pedestrian pathways explicitly included in scope.	(ISO, 2021a)

Table 6 continued from previous page

ISO/SAE 21434 Road vehicles - Cybersecurity engineering	specifies engineering requirements for cybersecurity risk management regarding concept, product development, production, operation, maintenance and decommissioning of electrical and electronic systems in road vehicles	2021	Road vehicles	Qualitative analyses	Standard is to be used during the entire product lifecycle	(ISO, 2021b)
ISO/4448 draft technical standard for operating automated vehicles and devices at curbs (kerbs) and sidewalks	technical data and communication standard for managing real-time mobility flow among automated vehicles and devices at sidewalk and curb (or pavement and kerb)	Work in progress	Robotic vehicles operating at the curb	-	High-level publication, no publications by ISO itself.	(Grush, n.d.) (Harmonize Mobility, n.d.-a,-b)

5.1 SADR hardware safety risks

The safe functioning of an SADR does not only depend on infrastructure and environmental factors. The technical quality of the SADR itself is at the basis. It is important that all hardware components, such as the actuators, sensors, lights, cameras, radar, Lidar and network communication systems are of high quality, such that as few disengagements as possible are caused by non-functioning or broken hardware components. To guarantee safety at component level, certified components can be used, which is usual practice in high-reliability sectors such as the aerospace industry (SAE International, n.d.). The use of specified standard components with quality control procedures can benefit the safety of the SADR hardware. Next to the hardware safety, the cybersecurity of an SADR must be ensured in order to protect the data, to prevent unintentional control of the SADR by third parties and to prevent the SADR from being intentionally used to send incorrect environmental information to other connected vehicles via V2X communication.

From Table 6 three standards can be identified regarding the technical safety of SADRs: ISO 26262 on functional safety, ISO/PAS 21448 on safety of the intended functionality of vehicles and ISO/SAE 21434 on cybersecurity engineering for road vehicles. According to Bellairs (2019) manufacturers can avoid and control systematic failures and be able to detect and control or mitigate random hardware failures by complying to ISO 26262. In addition, by adhering to ISO/PAS 21448 hazards resulting from functional insufficiencies of the intended functionality or by reasonably foreseeable misuse by persons can be addressed (ISO, 2019), which is referred to as Safety of the Intended Functionality (SOTIF). With ISO/PAS 21448, validation in real-world operating conditions is a key aspect to have sufficient testing under sufficiently random operating conditions to expose unknown unsafe scenario's (Camus, 2019). The cybersecurity of an SADR can be monitored and improved by continuing to review the SADR according to the continuously updated ISO/SAE 21434 standard. It is likely that internationally accepted certificates endorsing cybersecurity quality will be developed within a reasonable period of time, for example TÜV SÜD as an independent specialist in this field is working on such certification (TÜV SÜD, 2018). Since the objective of this research is the development of a dynamic testing framework for SADRs, and the technical safety according to ISO standards should be guaranteed by the manufacturer (ISO, 2018, 2019), no safety assessment as prescribed by ISO 26262 and ISO/PAS 21448 is performed within this research. It should be noted that complying to the aforementioned ISO or SAE standards could form a starting point for Dutch authorities to make an SADR eligible for practical testing on the sidewalk. This will be discussed further in Section 6.

5.2 Current knowledge on ADS performance risks

5.2.1 Scope

The scope of this literature review is to use scientific knowledge on the disengagement of automated driving systems to identify the factors and the impact of the factors underlying this disengagement. The hypothesis is that certain infrastructural and environmental characteristics may hinder the functioning of ADS. Because no publications exist that research this relationship for the case of SADRs, it is chosen to evaluate the relationship for ADS and autonomous vehicles due to the severe overlap with these technologies and SADR technology. Afterwards the found relationships can be translated to the SADR case.

5.2.2 Literature relevance

For this second literature study in this research the same approach has been taken as with the first literature study and is summarized shortly: primarily Google Scholar is used as search engine, a paper is considered relevant when it is published in a scientific journal and is peer-reviewed, most relevant when it is published in a top journal, based on SCImago Journal Ranking and Impact Factor, and papers of a more recent nature are assigned a higher relevance in view of the rapid technological developments in the automation sector. To assess papers on content relevance, the keywords in Table 7 are used. Grey publications are used to clarify matters if no scientific publication is at hand. Based on an informal interview with Prof.dr. Marjan Hagenzieker (April 14th, 2022) it was decided to extend the search queries with more explicit terms on the (un)safe interaction of traffic participants with ADS, because of the high importance of this interaction from a traffic safety perspective.

Table 7: Safety risks literature study concepts and keywords

Concept	Keyword(s)
Driving system	automated driving system ADS autonomous vehicle
Testing phase	testing validating
Safety	safety risk disengagement failure risk
Infrastructure or environment	infrastructure (factor/characteristic/condition/level) road (factor/characteristic/condition/level) environment (factor/characteristic/condition/level)

5.2.3 Literature search and selection

Next to the used search queries that proved to be useful for this study, some sources have been found by backward snowballing. Research has mainly focused on the infrastructure relation with human driver safety risks, little is known about the relation between infrastructure or environment and the ability of an ADS to cope with certain characteristics.

Table 8: Search queries risk factors

Search query	Publications (#)	Search date
"infrastructure risk" AND "automated driving systems"	6	1-4-2022
"automated driving system" AND "safety risk"	210	8-4-2022
"automated driving system" AND "safety risk" AND "infrastructure"	145	8-4-2022
"autonomous vehicle" AND "safety risks" AND "environmental characteristics"	7	8-4-2022
"infrastructure level" AND "automated driving systems"	29	8-4-2022
"road conditions" AND "safety risk" AND "automated driving system"	59	11-4-2022
"cause" AND "disengagement" AND "automated driving system"	259	11-4-2022
"cause" AND "disengagement factor" AND "automated driving system"	1	11-4-2022
"cause" AND "disengagement" AND "factor" AND "automated driving system"	206	11-4-2022
impact of "infrastructure condition" on "automated driving system"	11	11-4-2022

The final literature list used for this literature study consists of the following scientific publications: (Roh & Im, 2020), (Sivak & Schoettle, 2015), (Czarnecki, 2018b), (Czarnecki, 2018a), (Boggs, Arvin, & Khattak, 2020), (Feng et al., 2020), (Storsæter, 2021), (Mihalj et al., 2022), (Ren, Yin, Ge, & Meng, 2019), (Leroy, Gruyer, Orfila, & El Faouzi, 2020), (Thorn, Kimmel, Chaka, & Hamilton, 2018), (Rao, Deosthale, Barickman, Elsasser, & Schnelle, 2021), (Hillman & Capaldi, 2020),(Farah, Erkens, Alkim, & Arem, 2018) and (Schwall, Daniel, Victor, Favaro, & Hohnhold, 2020) The following publications were added to the literature list based on the input from Prof.dr. Hagenzieker: (Theeuwes & Hagenzieker, 1993), (Pokorny, Skender, Bjørnskau, & Hagenzieker, 2021), (Tabone et al., 2021), (Heikoop, Velasco, Boersma, Bjørnskau, & Hagenzieker, 2020), (M. Hagenzieker et al., 2020) and (M. P. Hagenzieker et al., 2020).

5.3 ADS disengagements from infrastructure and environment

First the risk factors originating from the interaction with infrastructure and environment will be separately discussed. Thereafter various studies that implicitly discuss combinations of the infrastructure and environment regarding safety risks will be assessed. The research of Leroy et al. (2020) established an overview of the known risk components associated with autonomous driving, classified into five concept groups: ego-vehicle related risks, environment related risks, road related risks, obstacles related risks and driver related risks (Figure 13). Analysis of the ego-vehicle risks has already been dealt with in Section 5.1, the concept group of driver related risks will not be taken into account in this research. The overview by Leroy et al. (2020) (Figure 13) has significant overlap with the research positioning figure used throughout this research (Figure 12) and is used in the following sections.

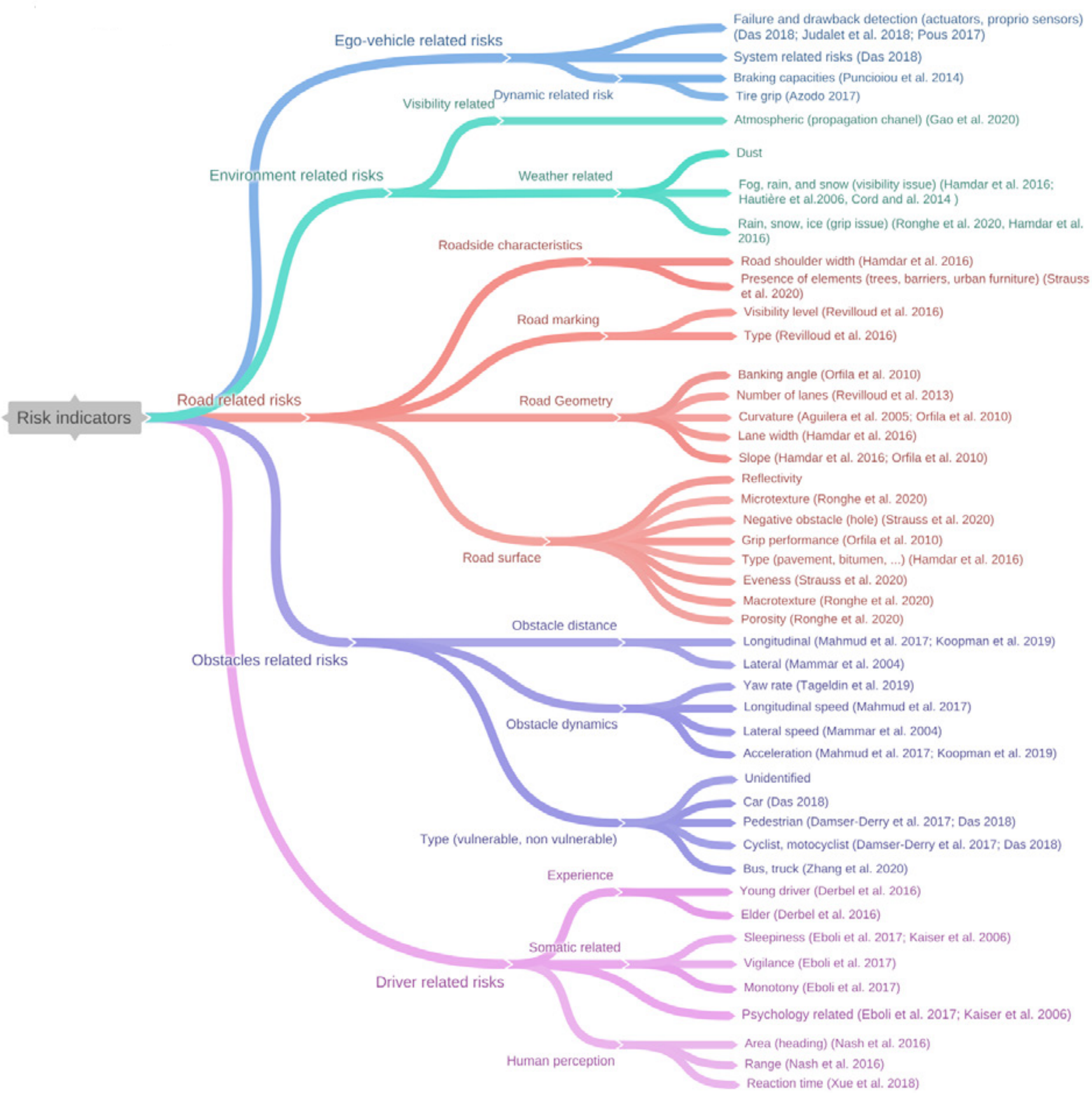


Figure 13: Risk components associated with autonomous driving, adapted from Leroy et al. (2020).

5.3.1 Infrastructure risks

As outlined in Section 4 the Operational Design Domain is the collection of operating conditions under which a given driving automation system is designed to function. When a given driving automation system finds itself in operating conditions for which it was not designed to function or detects a failure of the driving system, the driving task is handed over to the human driver (Boggs et al., 2020). This is referred to in the literature as disengagement of the automated driving system. Such situations are very likely to occur with road vehicles at SAE level 3 of driving automation where it is not possible to operate vehicles fully autonomous (see Section 3.6.1). According to Hillman & Capaldi (2020, p. 4) “disengagements are triggered when an AV cannot correctly match the perceived information with known datasets, due to the presentation of ambiguous or incomplete stimuli”. The neural networks that AVs rely on for manoeuvring are trained on millions of photos and video frames to enable the correct recognition and identification of a stimulus, but there are few opportunities to train the system to recognize edge situations, because edge situations are so uncommon. It is important to identify what other factors influence the occurrence of vehicle disengagements, so that this insight can be used to distinguish between geographical locations in the likelihood of ADS disengagements at those locations. The research of Mihalj et al. (2022) focused, among other factors, on the desired level of road infrastructure to be reliable for ADS. The researchers found that static elements such as ambiguous traffic signs and deviations from standard road markings have a higher probability to lead to recognition failure by ADS, but do not mention what the critical limits for ambiguous traffic signs or deviations from standard road markings are to confuse ADS. Mihalj et al. argue that using a higher grade of retroreflective material or sheeting on traffic signs will improve the visibility of the signs under all environmental conditions and decreases sign degradation over time. Risks according to environmental conditions will be further elaborated on later in this section.

Storsæter (2021) carried out a PhD research on how to design and maintain roads to facilitate automated driving and concluded that colors, patterns, and textures can be utilized to improve the visibility of existing road infrastructure elements, including guardrails, dividers, and road markings to aid automated detection. Furthermore, for lane (departure) detection with cameras, the contrast between road markings and road surface was found to be more essential than retroreflectivity measures. Czarnecki (2018b) argues that surface damage poses a direct risk for ADS. Research on unintended lane or road departure is mentioned more often in the evaluation of ADS, because lane departure is a clear indicator of unsafe driving. Next to the visibility of road infrastructure elements and the ability to detect them, road geometry, associated visibility distance and the speed of the vehicle are important factors that influence the probability to have a road or lane departure (Leroy et al., 2020; Farah et al., 2018; Czarnecki, 2018b). Different road friction coefficients influence vehicle maneuverability and stability. The literature analysis by Farah et al. (2018) identifies the same infrastructure aspects that affect the ADS's driving ability as discussed in this section.

In 2018, KPMG published a country readiness index for autonomous vehicles, which showed that the Netherlands is ranked first worldwide in general (KPMG, 2018). Part of this analysis consisted of an infrastructure assessment on quality, safety and connectivity, which actually have a relationship with the ability to drive safely from ADS systems. However, a large part of the Dutch readiness is derived from the highest presence of Electric Vehicle charging points (average per 100 km) in the world, for which no relation to driving safety can be found.

5.3.2 Environmental risks

The influence of environmental factors on ADS safe driving is recognised by several studies. The earlier cited research by Leroy et al. (2020) only includes weather conditions in their environment related risks concept group. Specific weather conditions have a direct impact on the road surface state and the visible distance for object detection. Rain, fog, sleet, snow and dense dust clouds will reduce the visibility distance and specific light conditions such as dusk or the setting sun can significantly influence the visibility level. The United States National Highway Traffic Safety Administration (NHTSA) notes that roadway conditions can change induced by the weather (Thorn et al., 2018; Rao et al., 2021). For example heavy rain or snow can flood the roadways or reduce the visibility of the road markings. Ren et al. (2019) studied environmental influences on the uncertainty of object detection by a deep neural network methodology. By measuring the average precision of object detection the impact of dark, sunset, rain and motion blur were assessed in comparison to a base scenario. It was found that in comparison to the base scenario the average detection precision decreased at dark, is close to the base scenario at sunset and light snow conditions, seriously decreases at rain and is the worst in the sample for motion blur. The exact average detection precision scores are summarized in Table 9. However it must be noted that for manoeuvring ADS do not solely rely on visible object detection and identification but also on radar and lidar technology, this research provides a detailed insight in the effect of environmental factors on the ability to detect objects with state of the art technology.

Table 9: Average object detection precision for specific environmental conditions by (Ren et al., 2019)

Environmental conditions	Average precision (%)
Base	89.4738
Dark	81.2102
Sunset	86.9811
Rain	62.6450
Snow	89.0789
Blur	56.5843

From practice, conclusions can also be drawn about the impact of environmental factors on the driving ability of autonomous vehicles. Research on data from Waymo, one of the pioneers in ADS, has revealed that their vehicles have not been operated during inclement weather, such as heavy rain and dust storms (Schwall et al., 2020). While it is not stated explicitly, this is most reasonably linked to the safety risks associated with certain environmental conditions. The environmental characteristics that affect the safe movement of an autonomous vehicle also affect other road users. For example, limited object detection capabilities by ADS due to dense fog means that an autonomous vehicle can map less of its surroundings, but it also limits the visibility for other road users. This means that an autonomous vehicle can be poorly visible, which can lead to unsafe situations caused by other road users and should therefor also be taken into account regarding safety risk assessments. The question of guilt in such situations will not be discussed further in this section.

The obstacles related risks concept group by Leroy et al. (2020) is in this research discussed as part of the environment (surroundings) of ADS. Leroy et al. define obstacles as "objects present on the road surface and in the surrounding area of an ego-vehicle" (Leroy et al., 2020, p.213). The risks originating from obstacles depend on three different characteristics of the obstacle: the distance to the obstacle (in both the longitudinal and lateral direction), the dynamics of the obstacle (described by the yaw rate, longitudinal and lateral speed, and obstacle acceleration), and the type of obstacle. The distinction made in type of obstacle is between cars, pedestrians, cyclists, motorcyclists, busses, trucks and unidentified objects. These different obstacle types are of interest to know because different obstacle groups have different expected behaviours, characteristics, goals and, if possible, different traffic rules to abide by. The different obstacle types therefore pose different types of risks.

5.3.3 Vulnerable road user interaction risks

Although other vulnerable road users can be approached as obstacles with position, yaw rate, longitudinal and lateral speed and acceleration, this does not do justice to the mapping of the risks that are and can be associated with the interaction between ADS and vulnerable road users. Vulnerable road users are "non-motorised road users, such as pedestrians and cyclists as well as motorcyclists and persons with disabilities or reduced mobility and orientation" (Tabone et al., 2021, p. 2). The interactions between vulnerable road users and ADS are not always logical or predictable, which poses additional risks that have to be addressed. For example, "the unpredictability of pedestrians makes it almost impossible for people or algorithms to avoid collisions" (Tabone et al., 2021, p. 8). In road traffic, scene dependent scanning behaviour is a phenomenon (Theeuwes & Hagenzieker, 1993). This means that people look for visible elements in places where they expect them, and consequently overlook elements in places where they do not expect them. The risk for ADS that gradually replace regular vehicles of being badly noticed by this phenomenon is small. Most autonomous vehicles have a similar appearance to regular vehicles and are expected to show the same behaviour and follow the same traffic rules, but this phenomenon may not be overlooked for more futuristic and smaller sized autonomous systems, such as delivery robots.

Pokorny et al. (2021) investigated the interaction between vulnerable road users and an ADS equipped shuttle based on video footage of these interactions. The first observation is that the shuttle used had a very defensive driving style. This means that when, for example, cyclists cycled close to the shuttle and came within the safety margins of the shuttle, the hard stop of the shuttle was activated. This created risky situations for motorised vehicles that were driving close to the shuttle, because the presence of cyclists close to a vehicle under normal, human-driven conditions would not lead to this type of hard stop. The hard stop in these cases are unexpected. The defensive driving style also meant that when crossing a zebra crossing, the shuttle reduced speed to such an extent that other road users repeatedly overtook the shuttle using the opposite lane, which is a major safety risk. The low speed of ADS shuttles, which results from safe defensive driving style programming, is recognised in the literature (Heikoop et al., 2020; M. Hagenzieker et al., 2020). For this specific experiment the researchers noted that the system only used pre-trained models and was unable to learn (Pokorny et al., 2021). In combination with incorrect programming of the priority rules for the shuttle, pedestrians approaching a zebra crossing from the left were unfairly denied right of way, proved to be a major safety risk. Throughout the whole range of observations in the experiment several occasions of stalemate situations could be revealed, where the shuttle and vulnerable road user were waiting for the other to make a move. It cannot be said with certainty what traffic situation initially caused this stalemate situations, but it does indicate that there are difficulties in clarifying the intentions of the shuttle for vulnerable road users. Issues in the communication between ADS and other (vulnerable) road users appears several times in the literature. Heikoop et al. (2020) mentions that ADS often follow the formal traffic rules, but is unable to communicate informally. In current traffic situations, many actions are handled by means of eye contact and feedback from other traffic participants (Sivak & Schoettle, 2015). These informal rules are not present in interactions with self-driving vehicles. A further problem in communication is that any communication between ADS and other road users has only been investigated in the context with one vehicle and one road user. It is uncertain how communication should take place when there are multiple road users with whom ADS has to communicate (Tabone et al., 2021).

From practice, examples can be identified where the known defensive driving style of ADS is a reason for other road users to abuse it, for example by taking the right of way, cutting the vehicle off, blocking the vehicle or in some other way testing the braking capabilities (Heikoop et al., 2020; Tabone et al., 2021). According to Heikoop et al. (2020) ADSs are not yet able to deal properly with other road users who deviate from the formal rules in such manner.

5.3.4 Road traffic risks

Apart from controlled (test) areas, the infrastructural and environmental conditions cannot be approached completely separately. Road traffic consists of continuously changing combinations of the static and dynamic factors originating from the infrastructure and environment. M. Hagenzieker et al. (2020) for example note that interactions between ADS and other road users appear to be more risky on shared narrow roads. Roh & Im (2020) identified the road sections and road situations where driving safety is depreciated based on a literature study and expert interviews, and reviewed those identified problematic sections and situations via an analytic hierarchy process (AHP) analysis on driving handicap factors. They came to the following 16 handicap road sections (infrastructure) and road situations (environment):

Table 10: Prioritization overview of handicap sections and situations from Roh & Im (2020).

Main Factor (Hierarchy 1)	Sub Class (Hierarchy 2)	Final Rank
Handicap Section	Areas with skyscrapers (high-rise buildings)	16
	Children and senior protect zone	14
	Section with poor lane condition	8
	Construction section	10
	Merging section	7
	Yellow light dilemma section	11
	Off-road section	15
	Tunnel section	13
Handicap Situation	Direct sun light	9
	Heavy rain and snow	2
	Fine dust and fog	6
	Stopping of large vehicle	12
	Driving of emergency vehicle	4
	Falling objects on road	3
	Abrupt action around road	1
	Pothole	5

Table 10 from Roh & Im (2020) gives a starting insight of the environmental factors that influence the performance of a vehicle for a specific ODD on a public road. Dynamic environmental factors, resulting from the handicap situations, are assigned higher priority than the infrastructural factors, resulting from the handicap sections. This prioritisation can be used to improve the performance of an ADS within the ODD by priority and to further analyse the high-priority ODDs, why these ODDs have a high impact on the ADS's ability to drive safely. It is unclear on what ground the sub classes have been included in the research and what sub classes have been excluded. Next to that, the quality of the infrastructural factors is questionable. The physical order of magnitude of the various factors varies from road sections to total areas. It is unclear what the exact underlying infrastructural characteristics are of the sections and areas in Table 10, which makes the outcomes of the AHP translatable only to a limited extent. In addition, the researchers used weighted values based on expert assessments to determine the hierarchy between the different sections and situations. It was therefore decided not to include these weights in Table 10. It would be optimal to perform a factor analysis on real life data regarding the handicap sections and situations to get an objective overview of the relative differences between the sections and situations with regard to the safety performance of an automated vehicle.

For the area of California Boggs et al. (2020) concluded that ADS-initiated disengagements, the driving system handing back control to the human driver, were more likely to occur in street and road environments compared to high-speed facilities. The built environments of the urban context are more multifaceted than freeways and interstates. Therefor there are more and more diverse interactions with other vehicles, vulnerable road users, intersections, and driveways. The probability that ADS is capable of identifying unforeseen events that might arise on streets and roads is lower compared to actively monitoring humans. A limitation Boggs et al. mention is that the data processing of the ADS cannot be observed and that it is therefore uncertain to pinpoint the cause of disengagements at a parameter level. The authors findings can be used as an hypothesis for other geographic areas, but this cannot currently be substantiated due to a lack of data for other geographic areas.

From the earlier quoted research on Waymo data stood out that other road users were violating at least one traffic rule in the majority of the crash situations Waymo encountered (Schwall et al., 2020). The only incidents that occurred always involved other road users, no unilateral accidents took place. This implies that the Waymo vehicles seem to be able to cope with all infrastructure elements within the test environment and that environmental factors, such as thus other road users behaviour, have a larger impact on ADS safe driving abilities. However, it is likely that Waymo will only deploy its vehicles on roads where it is known that the vehicles should be able to drive safely, to prevent unilateral accidents.

Feng et al. (2020) studied disengagement causes for 16 different ADS companies, including Waymo, and used the following disengagement cause classification:

- disengagement for maps and positioning
- disengagement for perception system
- disengagement for planning and decision system
- disengagement for control system
- disengagement for system overload
- disengagement for non-design traffic environment
- disengagement for weather influence

The planning and decision system was found the have the most prone link to failure and the street was found to be the most likely situated setting to cause disengagement. Perception and control are the second main causes leading to disengagement. The authors assume that the street scenario is more "prone to disengagement probably because of the complexity of the street scenario, which includes more types of traffic participants and more complex forms for intervention" (Feng et al., 2020, p. 37). The research shows that there are serious differences in the disengagement frequency between the different companies studied. The technology of some companies appears to be more reliable than others. The fact that the weather disengagement class does not show any influence on disengagement could be because ADS companies decided not to use their vehicles in inclement weather, as was concluded earlier for Waymo (Schwall et al., 2020).

5.3.5 Safety through digital infrastructure solutions

Several researchers point out that the use of digital solutions can partly reduce safety risks in road traffic. Sensor fusion and road network digitisation via vehicular communication and digital maps provide ways to increase the overall road network's resilience by offering redundancy, according to Mihalj et al. (2022). For example, smart traffic sign technology and smart traffic light technology can ensure that in circumstances where the traffic signs or lights are poorly visible to machine vision systems, a vehicle can still know with certainty what the current speed limit is or whether, for example, an intersection can be crossed safely. Thorn et al. (2018) write that all infrastructural and environmental characteristics information can be digitised so that a vehicle can use this data to make or validate a safe driving choice. Digital dynamic maps can be established to guide connected vehicles through temporary lane closures or variable speed limits. The input from a static lidar sensor can be used to provide a connected vehicle with information that a connected vehicle cannot yet see on the basis of its own sensor information, for example about approaching vehicles around the corner that have priority.

One example that shows how digital solutions can contribute to an increased level of safety is the Infrastructure Support Levels for Automated Driving (ISAD) categorization, established by the INFRAMIX project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 723016 (INFRAMIX, n.d.). The INFRAMIX categorization includes five levels of road infrastructure in its capabilities to support and guide automated vehicles, based on the type of digital information provided to automated vehicles. An overview of these levels and the associated description and provided digital information is given in Figure 14.

	Level	Name	Description	Digital information provided to AVs			
				Digital map with static road signs	VMS, warnings, incidents, weather	Microscopic traffic situation	Guidance: speed, gap, lane advice
Digital infrastructure	A	Cooperative driving	Based on the real-time information on vehicles movements, he infrastructure is able to guide AVs (groups of vehicles or single vehicles) in order to optimize the overall traffic flow	X	X	X	X
	B	Cooperative perception	Infrastructure is capable of perceiving microscopic traffic situations and providing this data to AVs in real-time	X	X	X	
	C	Dynamic digital information	All dynamic and static infrastructure information is available in digital form and can be provided to AVs	X	X		
Conventional infrastructure	D	Static digital information / Map support	Digital map data is available with static road signs. Map data could be complemented by physical reference points (landmarks signs). Traffic lights, short term road works and VMS need to be recognized by AVs	X			
	E	Conventional infrastructure / no AV support	Conventional infrastructure without digital information. AVs need to recognise road geometry and road signs				

Figure 14: ISAD categorization by INFRAMIX (n.d.)

Although the benefits of such digital solutions have been described in a qualitative way, to the authors knowledge there are no publications to date that substantiate these effects in a quantitative way. The publication on expert perspectives presented by Tabone et al. (2021) is the only known publication that also questions the benefits of digital solutions. In particular, the high costs, the need for maintenance and the vulnerability to damage by weather are mentioned as factors that are problematic for the viability of digital solutions in the short term.

5.4 Causal relationship diagram of infrastructure and environmental factors on safe performance

Based on the risks and risk factors that affect ADS in its ability to drive safely identified in the literature study in this section, a causal relationship diagram can be shown (Figure 15). The causal relationship diagram is displayed on the next page and will be discussed in total.

The causal relationship diagram is shown with the aim of providing insight into how the ability to drive safely is achieved for ADSs. The successful functioning of an ADS can be divided into three parts: the degree to which the system is able to correctly detect and recognise environmental elements, the degree to which the system is able to plan a route in real-time and the degree to which the system is able to actually manoeuvre in the physical environment.

The quality of the perception sensors has a positive influence on the ability to perceive correctly. When the quality of the traffic guidance means, such as traffic signs, are of good quality and therefore well visible, an ADS is better able to perceive correctly. According to the research by Mihalj et al. (2022), the use of retroreflective material has a positive influence on the visibility of traffic signs. The presence of non-standard road markings, intended or not, makes it difficult for an ADS system to recognise the road marking and its meaning in traffic. Apart from the (deviating) shape of the road markings, the visibility of the road markings and general road infrastructure has a positive effect on the ability to correctly perceive them. The research by Storsæter (2021) has shown that the use of colours, patterns and textures leads to a higher visibility of road infrastructure elements. Visibility distance is an important construct for the perceptive ability and route planning of ADS. A larger visibility distance offers the opportunity to identify obstacles over a larger distance and to use this information for route planning. Specific road geometry, such as sharp curves, reduces the visibility distance. In addition, sharp road bends increase the risk of a vehicle not being able to perform the manoeuvre properly. Edge situations are situations that statistically occur rarely in road traffic. Because they occur infrequently, there is little input data available to train the Artificial Intelligence (AI) models properly and the risk is greater that an ADS system cannot construct a (good) path for the vehicle with certainty. In addition, the presence of static and dynamic obstacles on and next to the roadway can reduce the visibility distance. The impact of a dynamic obstacle is slightly lower in comparison because this movement blocks other parts of the visibility to a certain extent, so that an overview can still be formed. The consequence, however, is that the risk of incorrect perception increases due to the reduced amount of visibility. The presence of obstacles also affects the ability to manoeuvre of an ADS because it (partially) blocks the infrastructure that can be driven on. The speed of the vehicle, in the literature often referred to as ego-vehicle, affects all three risk categories. A higher speed

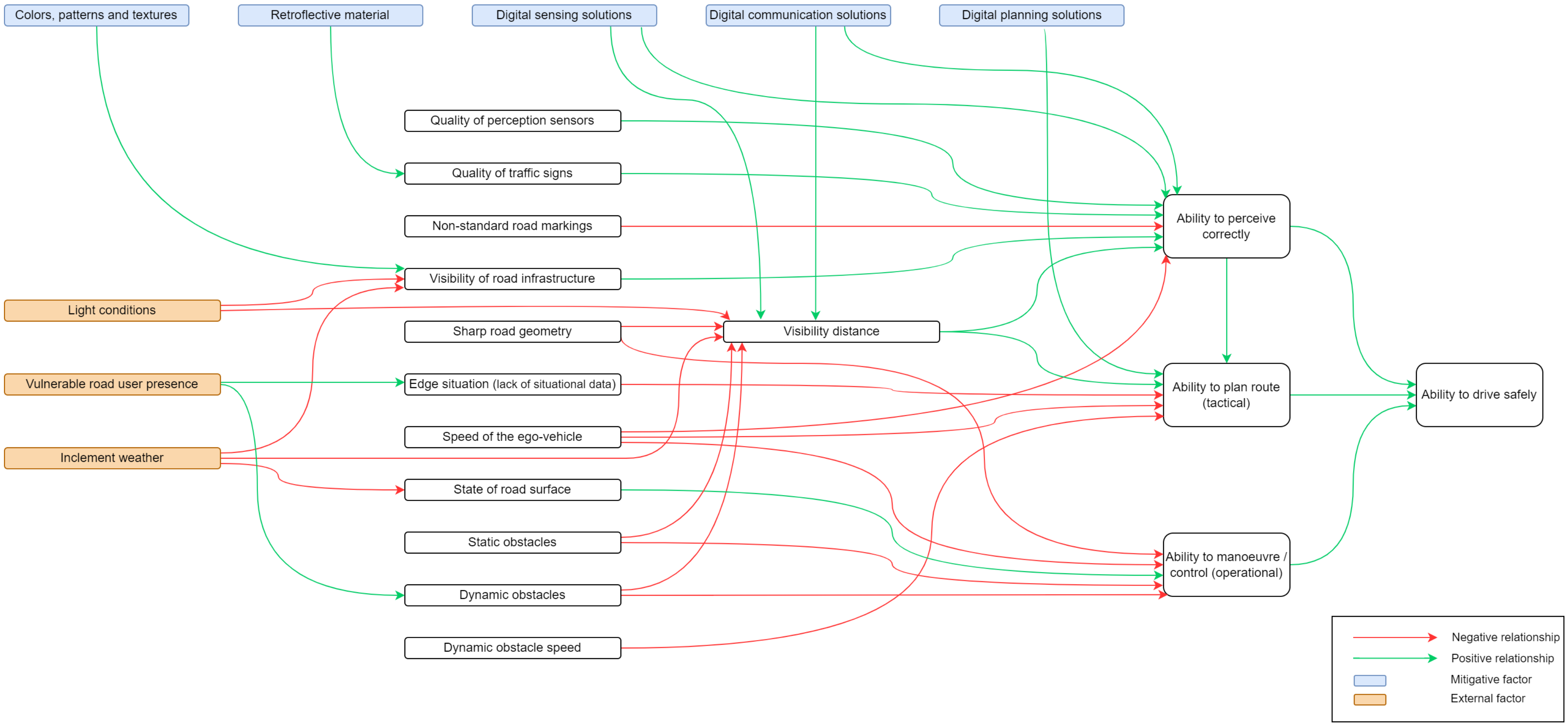


Figure 15: Conceptual causal relationship model

means that the system has less time to make observations with certainty and that possible adjustments to the intended path must be calculated and determined in a shorter time. In addition, at higher speed the vehicle is less controllable and no controlled deviations from the path can be made. Moving dynamic obstacles at a higher speed brings with it the same risk in terms of the possibilities of arriving at a route plan. A safe route must then be determined in less time than when a dynamic obstacle is moving at a lower speed.

External factors are factors that cannot be influenced within the outlined system. In this overview, the presence of vulnerable road users, inclement weather and specific light conditions are external factors. Specific light conditions are for example a bright and low-hanging sun, which makes the various elements of the road infrastructure less visible/identifyable and reduces the visibility distance. This factor does not refer to the difference between day- and nighttime, because however the object identification abilities based on camera information deteriorates, the lidar and radar systems in these conditions still succeed in detection objects because they do not rely on light. Inclement weather does not only decrease the visibility distance and visibility of the road infrastructure elements, but also effects the actual state of the road surface that the vehicle drives on. Inclement weather can increase the risk of the vehicle losing control because of the road surface state quality. It must be noted that these external factors do not only influence the ADS but also other road users in their ability to see and drive safely. The visibility distance for human driven vehicles is even more important, because humans can not detect objects in their vicinity by emitting radio and sound waves. The presence of vulnerable road users is included in the diagram because of the risks associated with the interaction between vulnerable road users and ADS. Due to the unpredictability and sometimes illegitimacy of the behaviour of vulnerable road users, more uncertain traffic situations are created, which can lead to edge situations where ADS does not know how to act correctly. The presence of vulnerable road users is also linked to the presence of dynamic obstacles in the operating environment. For the sake of completeness, this relationship has been included in the diagram.

As described, digital solutions can facilitate ADS. The deployment of lidar, radar and cameras at strategic locations at junctions can artificially increase the visibility distance and contribute to a higher object detection accuracy. This benefits the perceptive capabilities of ADS. By sending a signal when a traffic light is green, an ADS no longer has to rely solely on its own perception. By allowing vehicles to communicate with each other, the separate vehicles in a network can provide each other with geographic specific perceptive information. An ADS system can also compare the speed or direction of a vehicle that it detects with the values that this vehicle communicates, for higher detection accuracy. A digital planning tool could be used to control all vehicles in a network to optimize the network performance, for example in terms of emissions or driving times
It should be noted that this overview only includes the factors and mitigating factors that were identified through the literature study. There are several ways to influence the factors in the causal relationship diagram in the short or long term in order to realise optimal conditions for ADS.

It can be concluded that there exist certain infrastructural and environmental characteristics where the likelihood of ADS disengagements are higher, linked to the ability to perceive correctly, the ability to plan a route and the ability to control the vehicle. This acknowledgement will be used in Section 6 to develop a dynamic testing framework for SADRs on public sidewalks.

5.5 SWOV safety risk assessment Starship delivery robot

When assessing a practical trial with (partially) self-propelled vehicles on the public roads in the Netherlands, SWOV advises the RDW about the human/behavioural aspects of the practical trial concerned. It is of interest to assess how and to what extend risks are determined and classified by SWOV. In the past, SWOV gave advice about a delivery robot once (Petegem et al., 2018). The robot that was investigated in 2018 was a robot from Starship Technologies; it is assumed that a robot from Cartken would have been assessed in the same process manner. The risks the expert committee assessed originated from four risk categories:

- Risks related to the interaction of the operator(s) with the automated system in the vehicle
- Risks related to the interaction of the vehicle with other road users
- Risks associated with the location and timing of the field trial
- General project related risks

Potential risks in those four categories have been identified based on expert knowledge. The same experts then made a qualitative assessment of the probability and consequence of all risks. The most important risks according to the experts are discussed. All information in the next section is from Petegem et al. (2018).

5.5.1 Safety risks according to the SWOV

Safety risks at crossings
Road crossings that the robot has to make to move from sidewalk to sidewalk are considered risky situations, in which there is an increased risk of conflicts between the robot and the road users on the main road. Due to its limited size and height, the robot is less visible than a pedestrian. The robot is also not visible to road users if it is stationary behind parked cars. It is considered as a risk that some road users will not (immediately) recognise the robot as a road user. The visibility and recognisability of a delivery robot of Starships size remain less than that of a pedestrian, however a flag was attached to the delivery robot for visibility purposes.

Distraction of other road users
It is assessed as a risk that people's attention is focused on the robot for too long, causing them to be distracted from the driving task and from paying attention to other road users. Longer distractions and travelling at higher speeds involve higher risks of accidents.

Inattention or distraction of the handler and/or remote teleoperator
During the initial period of the field trial it was intended that for safety monitoring a handler would walk behind the robot. At all times during the field trial the robot has to ask permission to a remote operator to cross a road. It is considered a risk that the handler does not pay attention to the robot and, for example, starts using his phone. Another risk is present when the robot is waiting for permission to cross the road and the handler is distracted and misses the robot driving off, for example if the time to permission takes a while. It is assumed as a risk that both the operator and the handler do not pay enough attention to the situation when intervention is required.

Conflicts on the sidewalk
Due to its limited height and size, the robot does not stand out as much as other traffic participants on the pavement, especially when it is busier. The risk is that pedestrians will overlook the robot and bump into it. In addition, cyclists using the pavement are considered to run the risk of overlooking the robot and colliding with it.

5.5.2 Conclusions on the SWOV assessment

The main drawback of SWOV's risk assessment is that it is a qualitative assessment and that no insight is given into how the risks are generated per risk group. However, the overview of the various risks is extensive and contains clear explanations. It should be taken into account that this risk assessment was done in 2018 with the knowledge on autonomous driving systems applicable at that time and that this assessment was drawn up for a specific geographical location. It is normal that, as the authors themselves mention, the SWOV advice is limited to the circumstances of this specific practical test. In fact, the actual field test never took place, so the experts' qualitative risk assessment cannot be validated. The safety risks according to SWOV can be used for future practical tests with delivery robots. In comparison to the causal relationship diagram stands out that the risk assessment for SADRs by SWOV identifies different risk

5.6 Conclusion

Based on the literature reviewed it can be concluded that there are several factors within the operating domain of Autonomous Driving Systems that affect the sensing, planning and manoeuvring capabilities of ADSs. This understanding can be used in Section 6 to determine relative levels of difficulty within the ODD based on overlapping factors in the operating domain of ADSs compared to SADRs.

The publications examined showed that there are a number of studies that determine the effect of infrastructure and environmental factors on performance, but that these are often focused on small parts of the total operating domain. This may have something to do with the fact that, to date, there are no studies that have first focused on mapping out the operating environment as completely as possible and then looking at the risks that originate from this, as there is not yet an accepted method to completely map out the ODD. On the other hand, it may have to do with the fact that a lot of data on near misses is not made publicly available. Such relationships between factors and performance are therefore more difficult to investigate. Lastly, the publications investigated mainly concluded effects, the direction of the relationship of risk factors on performance, without actually being able to attach a quantitative impact to it.

6 Design suggestion

In this section the design suggestion of this research will be built up progressively. First, the requirements for a dynamic assessment framework for SADRs on public sidewalks in the Netherlands will be outlined. Subsequently, the 6-Layer Model, as motivated in Section 4.4, will be applied to describe the ODD of SADRs and identify the risk factors that are associated with deploying SADRs on public sidewalks. The risk factors identified will then be quantified so that they can be measured objectively. Thereafter the complete design suggestion will be presented and motivated based on the mentioned inputs.

6.1 Dynamic assessment framework design suggestion constraints and objectives

The design suggestion made in this research cannot be drawn up without any justification, but needs guidance in terms of requirements and objectives. The information from the theoretical background step and insights from the problem awareness step of the DSR cycle are at the basis of the requirements and objectives. These narrow down the design solution space and provide an answer to the fourth sub-question:

4. What are the requirements for a dynamic assessment framework for the safe performance of SADRs on public sidewalks in the Netherlands?

The answer to this sub-question will be given by means of an overview of the functional and non-functional constraints and objectives, which is more often used in engineering projects. Functional constraints are basic functions of a system, i.e. the things a system must do. Non-functional constraints are the things a system must have (Vleugel, 2021). If the constraints are too strict, then the solution space is almost empty (van Binsbergen, 2020). If the constraints are not strict enough, the solution may not work in practice. Functional and non-functional objectives describe what the system should have and increase the quality of the system. Justification of the constraints and requirements follows after Figure 16.

Functional Constraints (FC)

1. The framework must use objective descriptions of the infrastructure and environment
2. The framework must use SADR specific risks
3. The framework must allow for performance requirements modifications through time
4. The framework must allow for SADR modifications through time
5. The framework must be able to determine different ODD levels
6. The framework must allow for ODD-level changes over time for geographic locations
7. The framework must store performance data along ODD conditions

Non Functional Constraints (NFC)

1. The framework must be integratable with existing route planning modules
2. The framework data must be protected with state-of-the-art security technologies

Functional Objectives (FO)

1. The framework should use a complete description of the infrastructure and environment
2. The framework should explain a certain ODD level classification
3. The framework should be able to compare different geographic locations on the basis of location characteristics
4. The framework should include a means for incident reporting

Non Functional Objectives (NFO)

1. The framework should be understandable to all different stakeholders
2. The framework should be user-friendly

Figure 16: Overview of design suggestion constraints and objectives

FC1 and FC2 assure that the framework functions based on factual information and actual risks, rather than perceived risks. FC3 ensures that the assessment framework can develop according to what is learned. Because so little is objectively known about the functioning of SADRs in public environments, it cannot be assumed that the assessment framework is immediately perfect. FC4, in line with FC3, ensures that the SADR can be further developed over time. The quality of the dynamic assessment framework would be reduced if performance measurements of different SADR operating systems were compared. FC5 ensures that the relative difference in ODD difficulty is integrated into the framework, supplemented with FC6, justice is done to the dynamic nature of the public environment and environmental conditions that can change over time. FC7 is of importance because the dynamic assessment framework should not only be used as a means to deploy SADRs on public sidewalks, but in a broader context to learn about SADR performance. Therefore performance and ODD condition data should be stored.

NFC1 ensures that the framework can be linked to existing route planning modules, thus avoiding the need to develop additional software to link the framework to SADRs. NFC2 is intended to ensure that third parties who are not entitled to it do not have access to the SADR performance and ODD data.

FO1 is included as an incentive to map locations as complete as possible, but that the dynamic assessment framework is not considered unsatisfactory if data is not (yet) available for some geographical locations. FO2 and FO3 promote system transparency and overcome the issue that knowledge gained at a certain location to date cannot be translated to other locations. FO4 is included because a solution where incident reporting is included ensures a more complete framework system and promotes the learning aspect of the dynamic assessment framework.

With NFO1 it is taken into account that the dynamic assessment framework is designed for the complete set of stakeholders and that it is important that all stakeholders are involved during development and they understand the different aspects of the dynamic assessment framework. NF02 is included to take into perspective that the assessment framework should actually be used and that user unfriendliness of the digital solution is not a reason to not use the assessment framework.

6.2 The 6-Layer Model applied to the ODD of an SADR

The 6-Layer Model redefined by Scholtes et al. (2021) to the urban environment is used in this section as a basis to describe the ODD of the SADR. Each separate layer will be discussed following the layer definitions by Scholtes et al..

6.2.1 Layer 1: Sidewalk network and traffic guidance objects

The first layer for sidewalk network together with all permanent objects for traffic guidance can partially be adapted from the urban road network classification. The sidewalk network refers to the geometry, topology and topography of the sidewalks. Geometric design refers to the dimensions and arrangements of the visible features of a sidewalk (ITE, n.d.). Topology refers to the topological connections and spatial relationships of the aforementioned geometry (QGIS, 2022). The topography is referred to as the "three-dimensional arrangement of physical attributes (such as shape height, and depth) of a surface" (Houghton Mifflin Harcourt, 2006). Based on the former three concepts all relevant elements that make up the sidewalk can be listed and include: sidewalk widths, horizontal and vertical alignment of the sidewalk, slopes channelization, intersections, pedestrian crossings, adjacent infrastructure, sidewalk elevation, sidewalk ramp, road surface material and road surface irregularities.

The traffic guidance objects consists of two categories: the markings on the sidewalk and traffic signs and lights. The markings on the sidewalk differ slightly from the road markings. By means of markings the semantics of the lanes can still be derived, such as surrounding cycle paths or a bus lane. In addition, special areas with a functional purpose can be indicated, such as parking areas and keep-out areas. It can also occur that a different colour of paving stone is used to make a visible difference between adjacent road surfaces. What is different about road and sidewalk markings is that in public road traffic it is the practice to indicate instructions for drivers on the road, in addition to this being accomplished with traffic signs and signals, which is not common practice on the surface of the sidewalk. These road markings include for example stop lines and speed limits. The presence of permanent traffic signs and traffic lights is also described in Layer 1, while their changing states will be described in Layer 3. It should be noted that many different traffic signs and lights are used at the time of road works. These belong in Layer 3 and will be discussed there. In addition there may be traffic signs in the environment which are not applicable to an SADR, but their physical presence is still described within this layer of the 6-Layer Model.

6.2.2 Layer 2: Roadside structures

Layer 2 includes all static elements that can be described in the urban setting, but which are not directly linked to traffic guidance. The elements listed by Scholtes et al. (2021, p. 59137), which also apply to the sidewalk case, comprises: "buildings, vegetation like trees and bushes, walls and fences, street lamps, above ground hydrants, bollards, other types of fixed poles, vehicle restraint systems, guardrails, concrete step barriers and impact attenuators". In the case of the Dutch sidewalk this list may be extended with: bicycle stands, bike racks, benches, fixed waste bins, transformer boxes and letter boxes. Scholtes et al. note that bus shelters and surrounding constructions, such as bridges and tunnels, should also be grouped in Layer 2.

A significant difference with Layer 2 elements as described for the road case is that some of the static objects placed alongside public roads are placed on sidewalks, which makes that these elements limit the drivable surface of the sidewalk. However, from an element or object categorization perspective there is no difference between the layers.

6.2.3 Layer 3: Temporary modifications of Layer 1 and Layer 2

Layer 3 contains temporary modifications of elements of Layer 1 and Layer 2. In accordance with the definition by Scholtes et al. (2021) in Layer 3 only new objects and no new object classes can be introduced. The idea behind the third layer is that all adjustments to the infrastructure and the roadside structures that are not permanent can be designated. in this way, a distinction can be made between infrastructural sidewalk situations of a temporary nature during change and those of a permanent nature. Layer 3 offers the possibility to describe an infrastructurally dynamic situation as static. One Layer 3 example, not present in the road case, is the temporary presence of a sidewalk marking (modification to Layer 2) in the form of chalk.

6.2.4 Layer 4: Dynamic objects

Layer 4 includes "movable objects whose movements could evolve over time and are described by trajectories or maneuvers" (Scholtes et al., 2021, p. 59138). However elements or objects within Layer 4 terminology have the ability to move, they do not necessarily have to. Layer 4 can be best described as the physical layer of potential moving objects on top of the defined road network in Layers 1, 2 & 3. Examples provided by Scholtes et al. are most clearly defined in Figure 10 and categorizes: vehicles (moving and non-moving), pedestrians (moving and non-moving), trailers, animals and miscellaneous objects such as balls, coke cans et cetera. However this last category gives the opportunity the classify any object imaginable within Layer 4, there are a few examples important to mention regarding the specific usage of the sidewalk in the Netherlands: storage of building materials in urban areas, (shared) mobility resources parked on the sidewalk, recreationists on the sidewalk, whether in the form of playing children or people who have set up a terrace on the sidewalk in sunny weather. It must be noted that other delivery robots that an SADR encounters should also be described in Layer 4.

6.2.5 Layer 5: Environmental characteristics

Layer 5 is not different from the urban road classification case and describes the environmental conditions and resulting environmental road conditions. Environmental conditions include: (artificial) illumination, precipitation, visibility, cloudiness, wind and temperature. The environmental road conditions are a result of a combination of the aforementioned and describe if the sidewalk infrastructure is for example dry, wet or icy. The environmental road conditions are also the result of the implementation of any mitigating actions, such as the spreading of road salt on frozen roads. These actions do not change the objective road condition description.

6.2.6 Layer 6: Digital information

Layer 6 focuses on all possible digital communication exchange possibilities for a vehicle with other vehicles, infrastructure, environment or combinations thereof. The collective name for this type of digital communication is Vehicle to Everything (V2X). According to the Corporate Finance Institute (CFI) (n.d.) "V2X is a communication system that supports the transfer of information from a vehicle to moving parts of the traffic system that may affect the vehicle". Underlying technologies are vehicle-to-vehicle (V2V) communication, vehicle-to-infrastructure (V2I) communication, vehicle-to-network (V2N) communication and vehicle-to-pedestrian (V2P) communication (Mahmood, Zhang, & Sheng, 2019). A graphical representation is displayed in Figure 17. The overview is extended with the communication between the vehicle and the control room

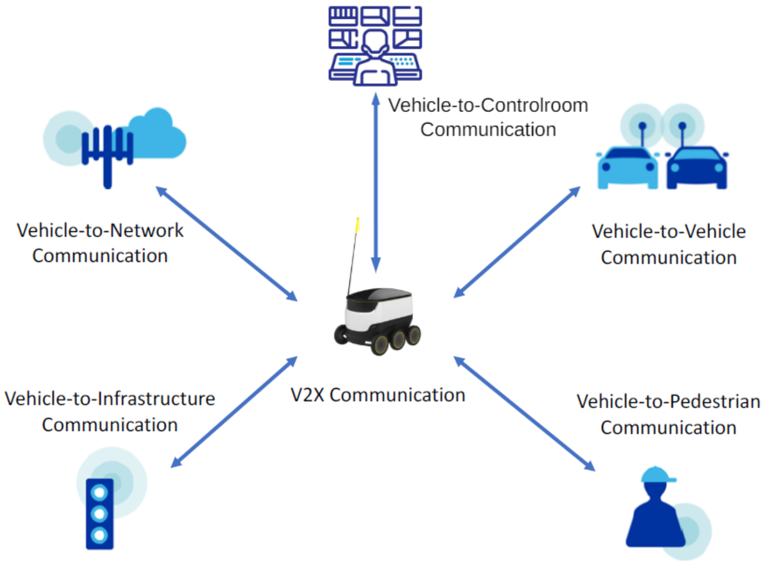
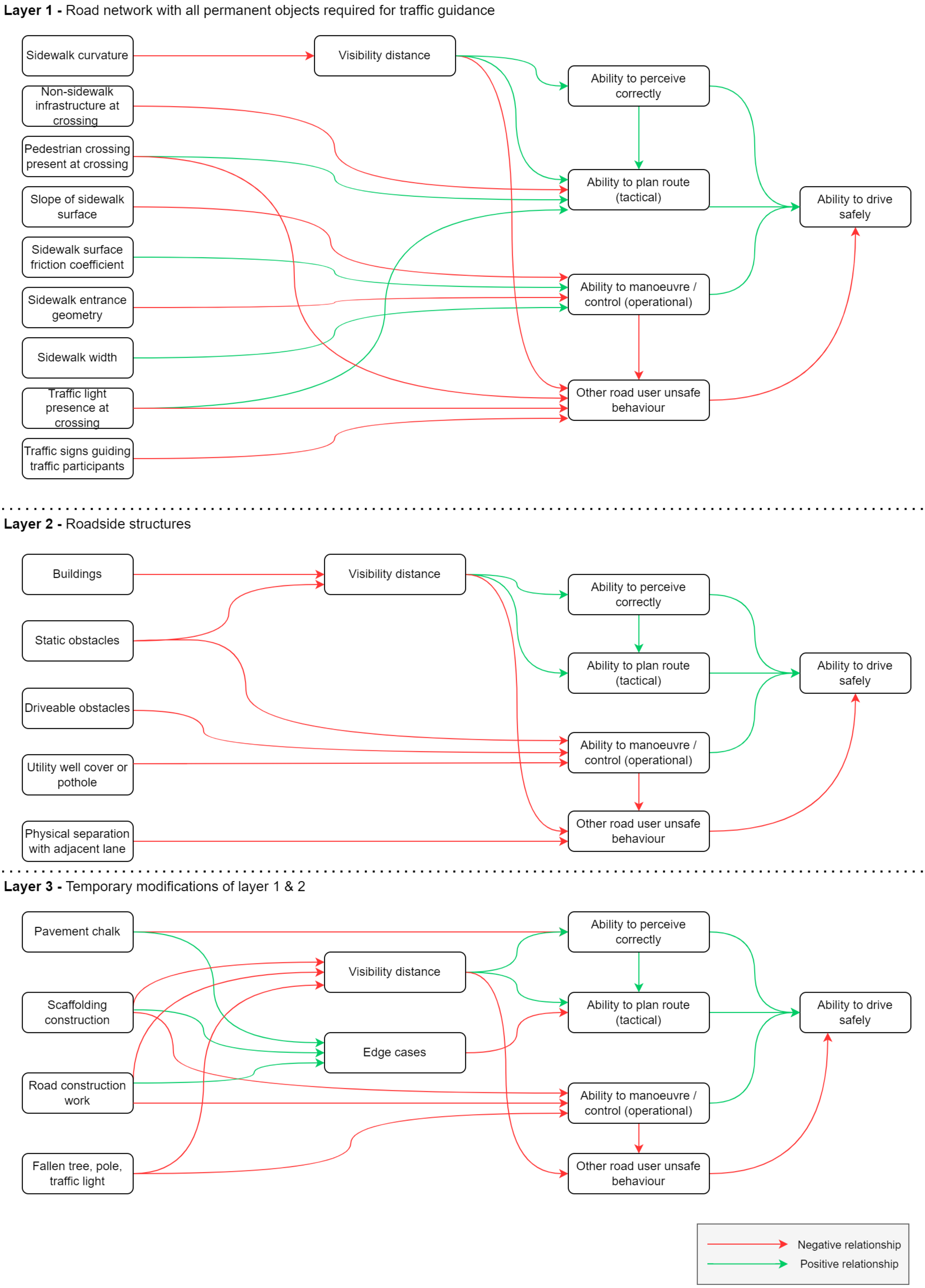


Figure 17: Vehicle to Everything clarification, adapted from Mahmood et al. (2019).

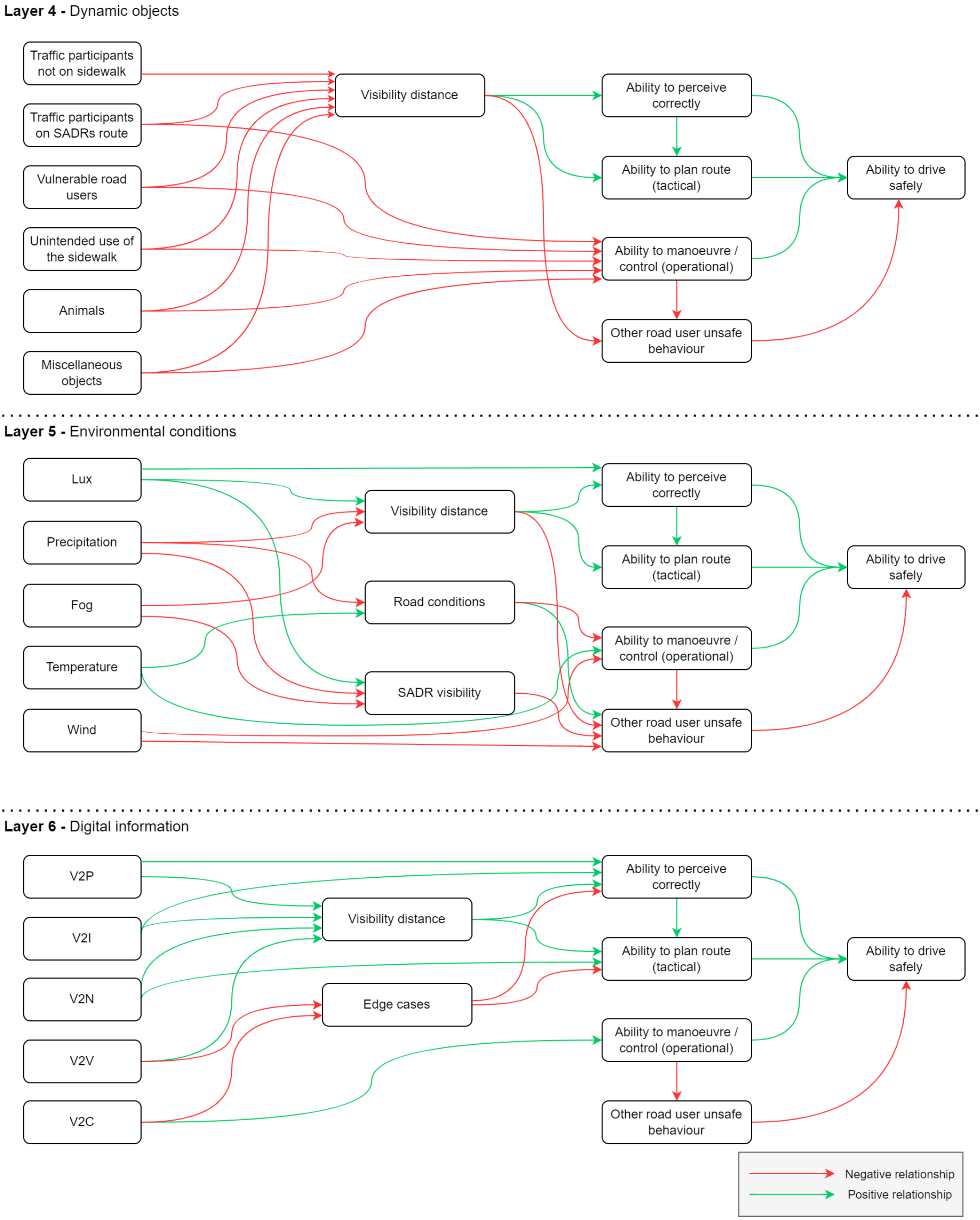
Examples of relevant information to be shared to an SADR in the context of this research are: the closure of sidewalk segments, extreme weather conditions, trajectory data of a crossing vehicle, data from intelligent traffic management systems (Scholtes et al., 2021) and an updated sidewalk conditions map according to a previous passing SADR. With increasing V2X communication SADRs (and vehicles in general) will become less dependent on their sole sensor data processing abilities. Digital information exchange with the remote teleoperator in the control room is important for the constant monitoring by the teleoperator as well as when the teleoperator has to drive the SADR.

6.3 SADR safe performance risk factors

The risk factors for ADS identified in Section 5 and the ODD specified for delivery robots in Section 6.2 can be used to distinguish between different levels of ODD difficulty for a delivery robot in a structured way. To do so, first the risks associated with the different layers of the 6-Layer model specified to SADRs have to be drafted for SADRs. The specifications of the robot of focus in this research are described in more detail in Appendix E. While for ADS the built environments of the urban context are more multifaceted than freeways and interstates, the same difficulty classification can be thought of for delivery robots. There exist grades of difficulty in types of sidewalks, in characteristics and obstacles, and difficulty in types of crosswalks. To illustrate this relative difficulty within the ODD, a conceptual model is used in which positive and negative relationships indicate the influence of the presence of elements within the ODD on the performance of SADRs. The performance of SADRs, which is represented as the factor 'the ability to drive safely', is determined by 'other road user unsafe behaviour' and the three main tasks an SADR has to fulfill. These three tasks are to perceive correctly, to plan a route/path and to actually manoeuvre in the physical environment.



(a) SADR performance risks guided from the 6-Layer Model: layers 1, 2 and 3



(b) SADR performance risks guided from the 6-Layer Model: layers 4, 5 and 6

Figure 18: SADR performance risks guided from the 6-Layer Model

The risk notion for SADRs is extended with a graph on the risks originating from the SADR in its appearance and operating characteristics in relation to other traffic participants, as is identified important in Section 5.3. This certain group of risks does not originate only from the infrastructure or environment, but also from the SADR.

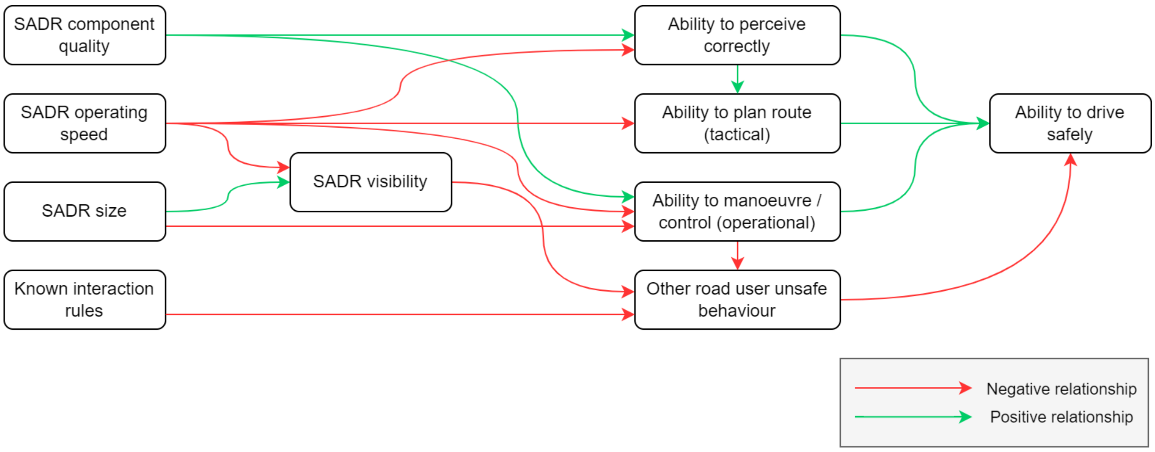


Figure 19: SADR performance risks guided from ego-vehicle

Based on Figure 18 & 19, the third sub-question can be answered:

3. What are the risk factors associated with driving SADRs on public sidewalks?

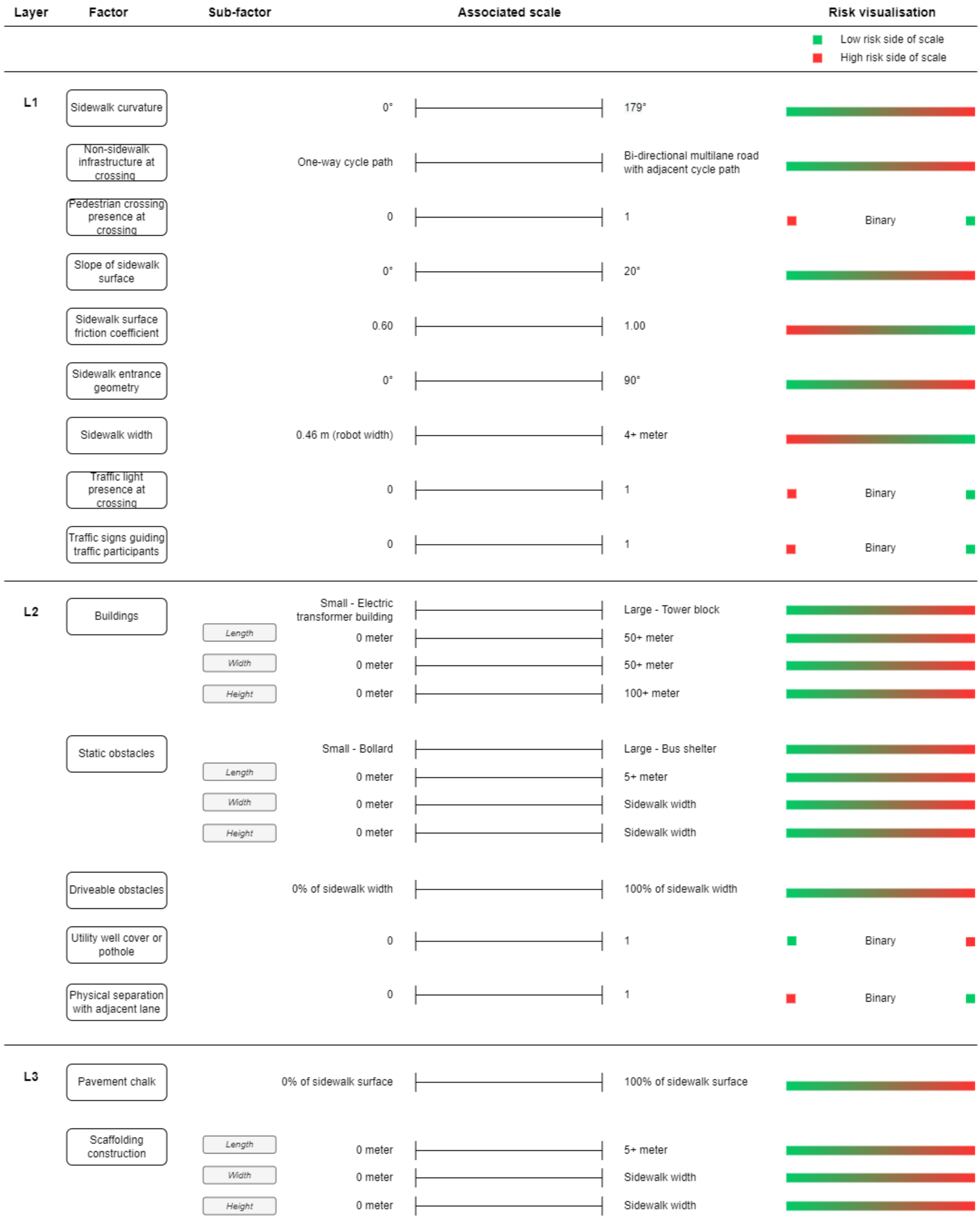
The main factors influencing the ability to drive safely of an SADR are: a low visibility distance and edge case occurrence, preventing an SADR from perceiving or planning correctly, reduced sidewalk conditions, specific sidewalk geometries and the presence of static and dynamic objects, making an SADR technically unable to plan or drive a path. In addition, a reduced visibility distance and a low visibility of an SADR cause an increased risk in the interaction with other road users.

The above will be further explained based on the layers of the 6-Layer Model. From Layer 1 can be concluded that infrastructural elements such as a pedestrian crossing, traffic signs and traffic lights at intersections help to ensure that an SADR is able to plan a route and that other road users show safe driving behaviour at these locations. The sidewalk geometry and friction coefficient of the sidewalk surface material determine whether an SADR is technically able to manoeuvre. The curvature of the sidewalk impacts the visibility distance of an SADR. From Layer 2 it follows that the presence of roadside structures reduces the visibility distance. In addition, roadside structures influence SADR manoeuvrability. Temporary modifications, as described with Layer 3, result in reduced visibility and a lower manoeuvrability due to their presence, but also make it more difficult for an SADR to find landmarks in the surrounding area. This increases the occurrence of edge cases. From Layer 4 it follows that all dynamic objects that can be identified on the sidewalk result in a reduced visibility distance and a lower manoeuvrability due to their presence. From Layer 5, it can be noted that, in general, the better, clear and calm weather conditions provide the best visibility distance and road conditions. This also ensures that an SADR is most visible to other road users, which improves overall traffic system safety. Layer 6 covers all the digital means that can assist an SADR in perceiving and planning by artificially increasing the visibility distance and reduce the number of edge cases that an SADR has to deal with. From the last added layer with ego vehicle characteristics it becomes clear that SADR characteristics can affect the ability to perceive, plan and manoeuvre, and that established rules for the driving behaviour of SADRs and for the interaction with SADRs benefit the (unintended) unsafe behaviour of other road users.

6.4 Qualified and quantified risk factors

The next step to work towards a classification of different ODD levels is by qualifying and quantifying the different factors from Figure 18 & 19. By objectively measuring the different factors, they can then also be compared with each other. Where possible, risk factors have been quantified, as this is most useful when carrying out data analyses at a later stage (see Appendix G). Otherwise, risk factors have been qualitatively described. In the complete overview (Figure 20) for every risk factor present in the ODD of SADRs is described on what scale the risk factor is measured and a visualisation of the low and high risk side of the associated scale has been added. It must be noted that the risk factor visualizations cannot be compared to each other on impact.

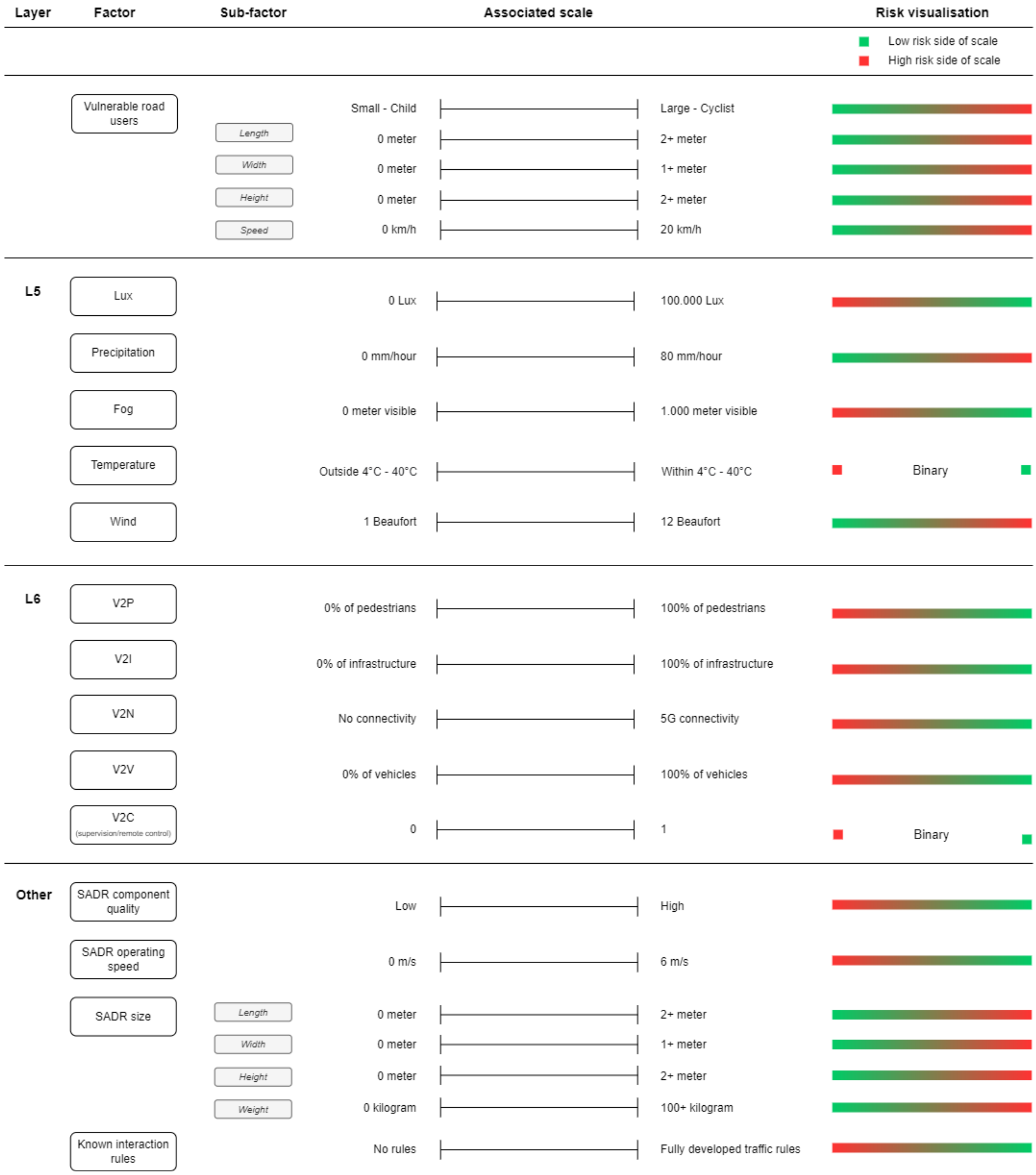
To give readers and insight into what logically falls within a specific risk factor, a qualitative scale with size examples has been added.



(a) Scale overview (1/3)

Layer	Factor	Sub-factor	Associated scale		Risk visualisation
					Low risk side of scale High risk side of scale
	Road construction work	Length	0 meter	5+ meter	
		Width	0 meter	Sidewalk width	
		Height	0 meter	Sidewalk width	
	Fallen tree, pole, traffic light	Length	0 meter	6+ meter	
		Width	0 meter	2+ meter	
		Height	0 meter	2+ meter	
L4	Traffic participants not on sidewalk		Small - Moped	Large - Truck	
		Length	0 meter	20+ meter	
		Width	0 meter	3+ meter	
		Height	0 meter	4+ meter	
		Speed	0 km/h	50 km/h	
	Traffic participants on SADRs route		Small - Moped	Large - Truck	
		Length	0 meter	20+ meter	
		Width	0 meter	3+ meter	
		Height	0 meter	4+ meter	
		Speed	0 km/h	50 km/h	
	Unintended use of the sidewalk		Small - Parking of modalities	Large - Outdoor terrace	
		Length	0 meter	20+ meter	
		Width	0 meter	3+ meter	
		Height	0 meter	4+ meter	
		Speed	0 km/h	15 km/h	
	Animals		Small - Bird	Large - Horse	
		Length	0 meter	2+ meter	
		Width	0 meter	1+ meter	
		Height	0 meter	2+ meter	
		Speed	0 km/h	25 km/h	
	Miscellaneous objects		Small - Coke can	Large - Waste Cannister	
		Length	0 meter	20+ meter	
		Width	0 meter	3+ meter	
		Height	0 meter	4+ meter	
		Speed	0 km/h	15 km/h	

(b) Scale overview (2/3)



(c) Scale overview (3/3)

Figure 20: Risk factor qualification and quantification according to the 6-Layer Model

From Figure 20 follows that almost every risk factor can be quantitatively described. The risk factors that have only been qualified are: 'Non-sidewalk infrastructure at crossing', 'V2N', 'SADR component quality', 'SADR size' and 'Known interaction rules'. Another interesting conclusion that can be drawn from the overview is that many elements present in the ODD can be reduced to a description of their length, width and height. If those elements are dynamic also their speed is important. This generalisation can be well used in the design suggestion.

6.5 Dynamic assessment framework design suggestion

As has been made clear more often in this research, testing SADRs on the public sidewalk was not that easily solved. By identifying an objective manner to describe the ODD of an SADR, identifying the associated risks in the ODD and providing guidance on quantifying them, a major step has been taken towards a system where different degrees of difficulty in the ODD can be identified. Based on difficult ODD levels, a safe assessment framework for SADRs can be designed and a secure proof-of-concept method to actually test SADRs on public sidewalks, where SADRs are allowed more as the technology is proven to be safe (Humblet, 2021), can be rolled out in phases in the Netherlands. The actual design suggested in this research is a digitised solution in line with the intended outcome of the Design Science Research cycle and will be described further in this section. To do so, the structure in Figure 21 is used.

Figure 21: Buildup Section 6.5 dynamic assessment framework design suggestion

As mentioned in Section 6.4 the risk visualisations from the different risk factors in Figure 20 cannot be compared to each other. That is because to date there has been no research that has evaluated the impact of the risk factors on the ability of an SADR to drive safely. However, a starting insight into different ODD difficulties and the opportunities that arise from them can be given. Based on these difficulty differences between ODDs, a digital dynamic assessment framework is proposed and further developed using a Living Lab approach, on the basis of which, from a risk minimisation perspective, SADRs are tested in public spaces according to a proof-of-concept method. The digital system can on the one hand be a means for human stakeholders to understand the ODD difficulty and what exactly causes that relative difficulty. On the other hand, the system can provide geofenced guidance to an SADR so that it knows exactly when and where it is allowed to drive. Performance data of an SADR is stored alongside the exact conditions under which that performance was achieved. That way lessons can be learned from the relationship between these environmental factors and the performance of SADRs, and the assessment framework can be updated accordingly. Additionally, because operating conditions are accurately monitored, knowledge is gained that can be translated to other geographical locations. The full elaboration in Section 6.5 will answer the fifth sub-question:

> *5. What dynamic assessment framework can be designed to assess if an SADR is able to operate safely at different ODD-levels?*

6.5.1 Dynamic assessment framework structure

The analysed layers of the 6-Layer Model can be divided into a spatial axis (Layer 1, 2 and 3), a dynamic objects axis (Layer 4) and an environmental characteristics axis (Layer 5). The sixth layer, dependent on the functionalities the sixth layer includes for a specific geographic location, can affect all three axes. This resulting matrix can form the basis for the proof-of-concept methodology (Figure 22). This matrix is the complete abstract space in which every geographic location can be classified in relative difficulty regarding spatial elements, dynamic objects and environmental characteristics. For illustrative purposes has been chosen to present the three axes on a scale from 0 to 5. To date these scores cannot be benchmarked against one another.

First the 6-Layer Model has been used to objectively map the complete ODD of a delivery robot. As follows from Figure 11 the Layers 1, 2 and 3 can be aggregated as they all relate to the spatial component of the ODD. The aggregation of Layers 1, 2, and 3 will be further discussed in Appendix G, where, by means of a regression analysis, the difference in effect on the ability to drive safely of an SADR is demonstrated based on the underlying risk factors from the three layers. The choice for the remaining two axes is based on the difference between the content of the layers and the ease with which these underlying risk factors can be analysed separately.

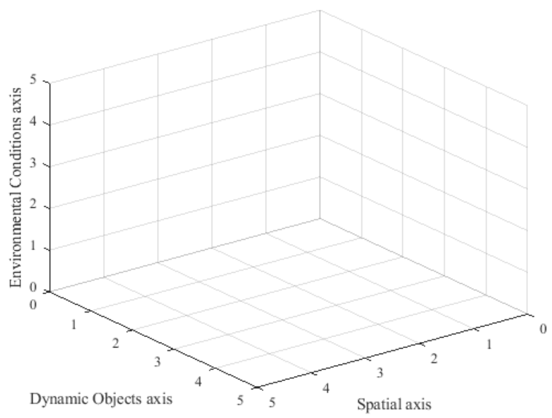


Figure 22: 3D ODD classification space

By distinguishing between these three axes, the relative difficulty along each axis can be determined separately and then combined into a unified ODD difficulty classification.

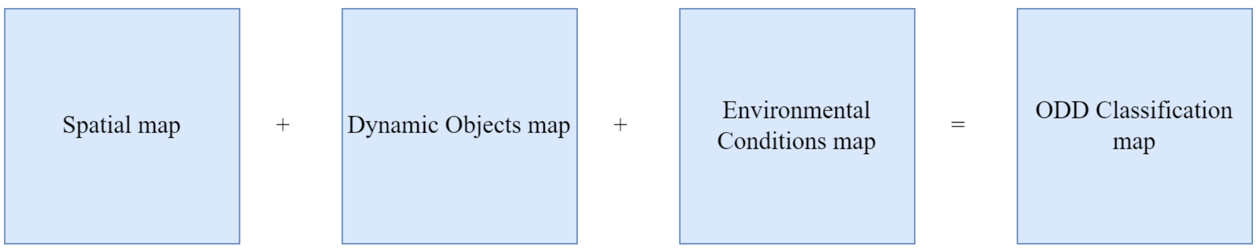


Figure 23: ODD classification map structure

This structure results in the following three sub-maps and the final map for the dynamic assessment framework:

(a) Spatial ODD classification map - illustrative

(b) Dynamic objects ODD classification map - illustrative

(c) Environmental conditions ODD classification map - illustrative

(d) ODD classification map - illustrative

Figure 24: Illustrative sub-maps and final ODD classification map

In order for users and stakeholders within a practical test to understand how the dynamic assessment framework map functions, it is important that the digital system is able to indicate why a geographical location belongs to a certain ODD level at a certain point in time. For example, by clicking on a specific link in the network in the dynamic assessment map, the digital solution can display on the right-hand side the three main factors that, in percentage terms, make up the largest proportion of the final ODD level. The example in Figure 25 shows the same link in a network twice: once where the sidewalk is classified as relatively easy because the sky is clear and the weather is calm, and once where the sidewalk is classified as relatively difficult because it is snowing and there are vulnerable road users (VRUs) on the sidewalk. This example also shows well that the ODD level of a geographical location can change over time.

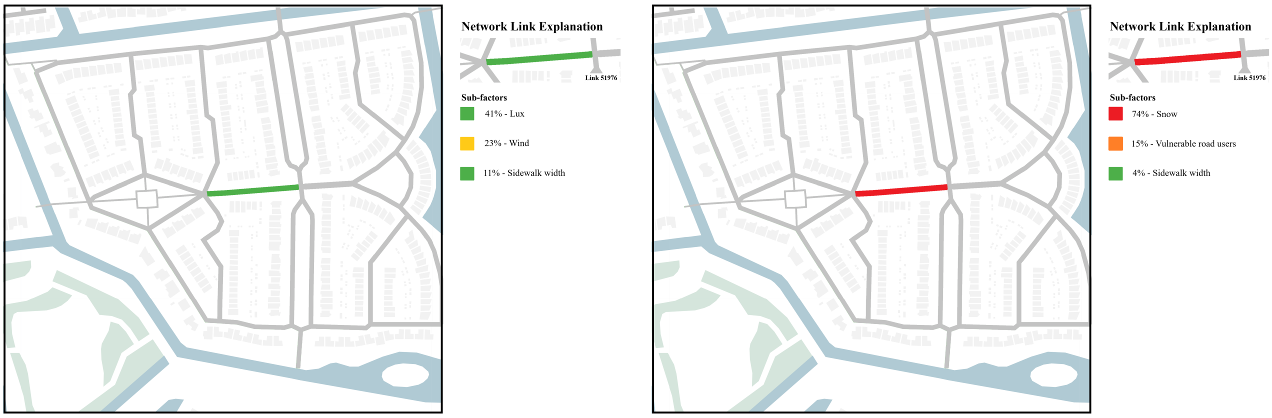


Figure 25: ODD level explainability

6.5.2 Rationale to build framework on a link level

The first argument for presenting the dynamic assessment framework proposed in this study at link level is that such a system has never been designed and there are still uncertainties before it can actually be put into operation. By making the dynamic assessment framework transparent at a link level, on the one hand it can be made insightful how the framework can work and can be used, and on the other hand, no claims have to be made at a more detailed level that cannot be substantiated.

Second, this research does not focus on the deliveries associated with SADRs, but on their feasibility to drive safely on Dutch sidewalks and therefore justifies to generalize the dynamic assessment framework at a link level. It is common standard that a route consists of several adjacent links, connected by nodes (Lee, 2016). If for example a delivery should be made to a location at the beginning of a street, whilst a road obstruction is present at a more distant location in that same street, the SADR will not have to interact with the road obstruction. To include this in the dynamic assessment framework more detailed maps of every geographic location would be required, which is not deemed feasible at this stage. It does, however, raise the question of how this data should be handled after the proposed dynamic assessment framework system is implemented: what does it say if an SADR has driven a stretch in a 'difficult' ODD level, but never had to deal with the factors that actually cause the difficult ODD because it never got there? A possible solution for this is that the database with SADR performance data only stores measurements of driven sidewalk sections that an SADR has completely covered. In this way it is possible to learn from the largest possible part of the travelled route of the SADR, without storing improper data. Improper data could potentially falsely classify locations as easier than they really are.

If a link is assigned a certain ODD level based on characteristics this implies that along a certain link the characteristics are completely equal. However, if a sidewalk width differs from 1.00 meter to 1.10 meter over a certain network section, this assumption no longer holds. This can be mitigated by abstracting the quantitatively defined scales in Section 6.4 to scales of equal interval. The advantage of this abstraction is that network links that differ to the finest detail can still be assessed as one link and that due to a higher degree of abstraction geographical locations can be compared more easily (see Section 6.5.9). The disadvantage of this abstraction is that less detailed insight is obtained into the impact of risk factors on the SADR ability to drive safely. It is recommended that follow-up research focuses on the questions of whether and how the various defined scales of the risk factors should be abstracted so that, on the one hand, network links can be assessed unambiguously and factually but, on the other hand, enough detail remains to learn from the experiences with SADRs.

Link-node representation and dynamic assessment framework map

In Figure 26 the link-node representation associated with the dynamic assessment framework on a link level is visualised at a fictitious intersection. In this figure, the nodes are placed at the corners of sidewalks, where SADRs must cross the road using the zebra crossing or remain on the sidewalk, dependent on their destination. In Figure 27 a selection of the underlying links and nodes of the dynamic assessment framework maps used in this research are presented. In the network, an SADR travels its route by successively travelling through different links. This small cutout of the chosen network shows that the sidewalk network is very unstandardised. There are streets in the focus area that have a sidewalk on one side of the street, and streets that have a sidewalk on both sides of the street. Although not shown on this figure, there are differences in the presence of pedestrian crossings, and some sidewalks are or are not elevated relative to the adjacent road surface used by cyclists and vehicles. This link-node representation can be used to map the infrastructural, dynamic objects and environmental characteristics of each link in the network detail. Note that the red color for the links and nodes has been chosen because of the contrast with the aerial photo in the background and not to indicate (relative) difficulty of the ODD.

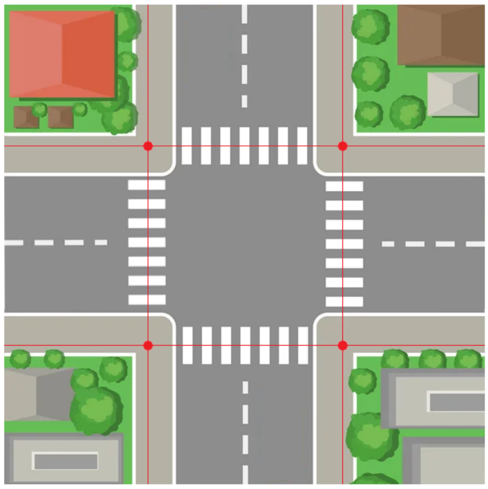


Figure 26: Schematic node-link representation applied at an intersection

Figure 27: Node-link representation applied to dynamic assessment framework map

6.5.3 Incident reporting

Some countries outside the Netherlands are less restrictive when it comes to testing ADS-equipped vehicles on public roads. An important requirement in this respect is the mandatory publication of crash reports in the United States and mandatory collision or incident reporting in Canada (NHTSA, 2021; Canada Transport, 2021). These reports include information on the incident situation and should include ADS data. This data could reveal whether there are any common patterns in self-driving vehicle crashes or if there are any operational issues (NHTSA, 2021). According to Canada Transport (2021) to expand common knowledge of the current level of ADS testing and deployment and increase public trust, trial organizations should consider sharing data with researchers and the general public. It cannot be assumed that incidents will never happen, and it is important to learn from the incidents that do happen, which justifies the obligation to publish certain reports.

It is very likely that Dutch authorities will demand similar publicly accessible incident reports during field tests, so it is good to take this into account already in this development phase of the dynamic assessment framework. Because the performance of an SADR is monitored in a complete and objectively described ODD, a lot of information about an incident location is already present and can be stored in a database system. This data can be supplemented with a qualitatively described report by the SADR operator on what has taken place, a possible cause and what mitigating measures (if any) have been or will be taken to prevent a similar incident from occurring in the future. On the one hand, the structured collection and storage of incident data and reporting over time can be used learn from the performance of SADRs. On this basis, the assessment framework can be further improved, as wil be further discussed in Section 6.5.5. On the other hand, publicly accessible incident reports improve system transparency (NHTSA, 2021). It can be verified for example if the aforementioned mitigating measures are indeed implemented. Incident reporting can be integrated into the digital system as follows: an overview of incidents can be displayed on the map shown previously (Figure 24), by clicking on an incident the full incident report can be displayed. A visual representation is given in Figure 28. The overview in the digital system can be made as detailed as desired by the relevant project stakeholders.

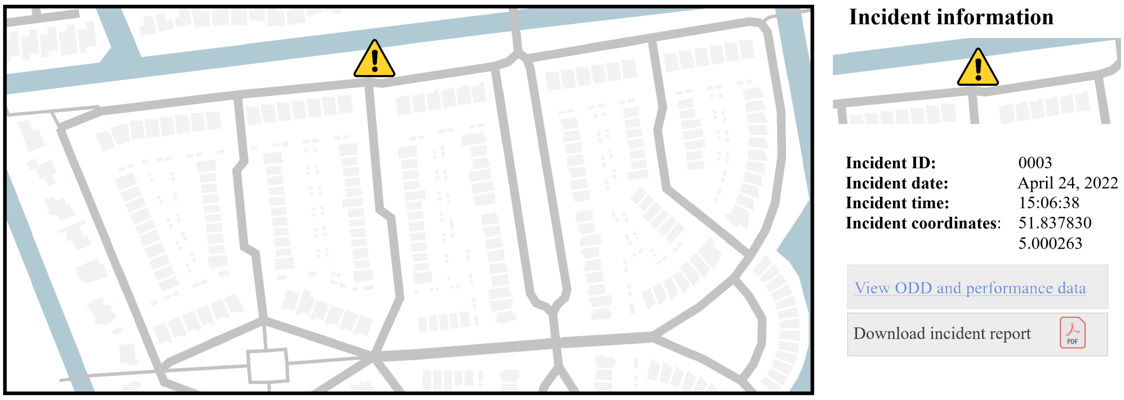


Figure 28: Incident reporting visualisation

6.5.4 Proof-of-concept dynamic assessment framework

Separating the static spatial elements from the dynamic elements and objects makes it possible to test the impact of infrastructural elements on an SADR's ability to drive at a (closed) test location. The first advantage is that by testing on a closed terrain, there is no safety risk to the public. The second advantage is that the outcomes of those tests provide an objective way to assess and qualify spatial differences because no disturbances were caused by dynamic objects. This objective assessment forms the basis for the dynamic assessment framework along the spatial axis (Figure 22). Tests on a closed test site can also provide insight into the effect of environmental conditions on the performance of an SADR, because it is possible to experiment with different amounts of (artificial) light, precipitation, wind, etc. The safe measurements of the complete solution space along the spatial and environmental axes can be visualised as follows:

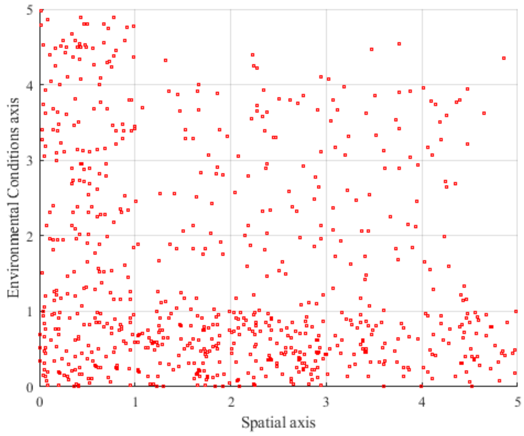


Figure 29: Test location experiment results - illustrative

Proving that an SADR can drive safely on a closed test site does not in itself prove that an SADR can drive safely on the public sidewalk. Manufacturers that bring SAE Level 4 or Level 5 autonomous driving innovations to the market can indicate that, for example, SADRs are capable of handling the entire ODD of the sidewalk. However, this is not a valid reason to allow SADRs everywhere on the sidewalk. This research argues that there are different degrees of difficulty within the ODD when it comes to the extent to which an SADR is able to manoeuvre safely. By actually determining these different levels and recognising them as such, the deployment of SADRs on public sidewalks can be gradually controlled, once it has been proven that they can safely handle a certain ODD level. This so-called proof-of-concept approach is explained using the infrastructural (Figure 30), dynamic (Figure 31) and environmental characteristics (Figure 32) axes. The argument that 'the number of kilometres driven by an autonomous vehicle says nothing about the safety of the vehicle because the context is not known' can be refuted by starting from the objective conditions and their actual impact. The safest approach is to allow an SADR in the easiest locations first. By making safe kilometers in this specific, monitored setting, more difficult settings can then be gradually allowed. In the figures, this is represented by the red and blue points: the red points are 'proven safe', for example proven already at a test location, and the blue points are 'unknown safe'. By then, not enough kilometres have been driven in those specific conditions to be able to argue that the deployment of SADRs on the sidewalk is feasible under such conditions. When a sufficient number of kilometres has been driven, the measurements marked as 'unknown safe' can be marked as 'proven safe', after which progressively more difficult settings can be allowed until the system limits are reached. First, it should be noted that the data points in Figures 30 to 32 are illustrative to show how the system could possibly work. Secondly, it should be noted that SADRs can gradually be allowed to operate in more difficult infrastructure and dynamic conditions, but it cannot easily be allowed to operate in more difficult environmental conditions, because in order to do so these weather conditions must be present. Data from the Royal Netherlands Meteorological Institute (KNMI) shows that these conditions occur less frequent in the Netherlands (KNMI, 2022). A criterion to determine when the 'unknown safe' data points change to 'proven safe' cannot be given at this time. On the one hand, this will depend on the number of ODD intermediate levels in the assessment framework that follow from the regression analysis on the risk factors. On the other hand, it is an operational choice of the stakeholders involved in initial field test: a more risk averse approach will require more data and vice versa.

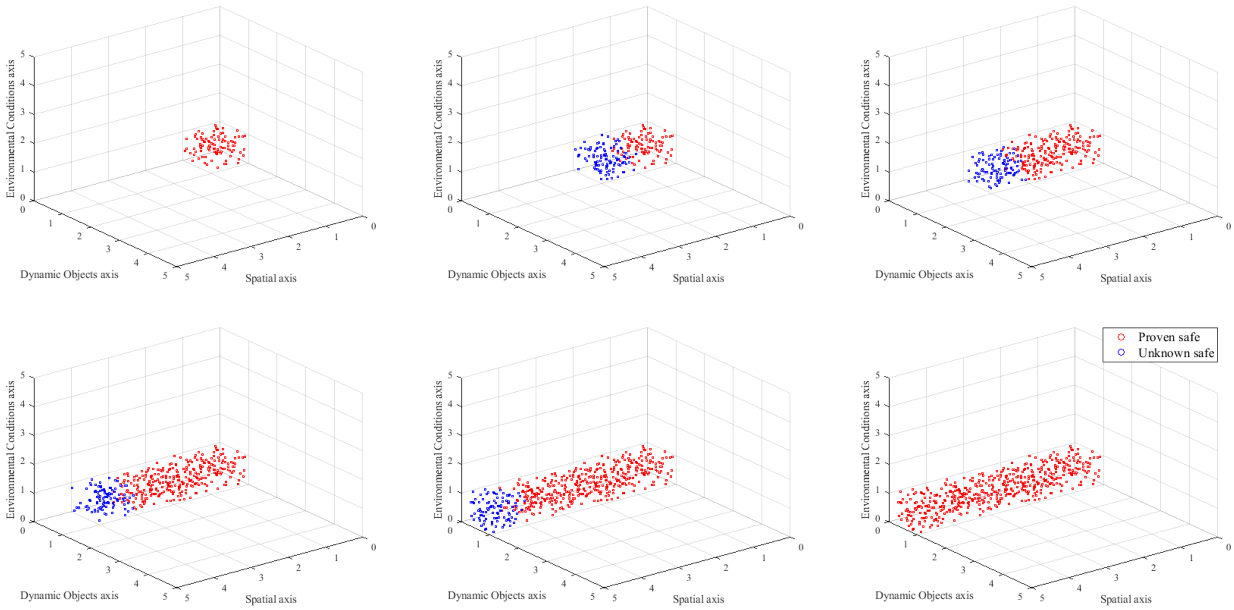


Figure 30: Proof-of-concept along the spatial axis

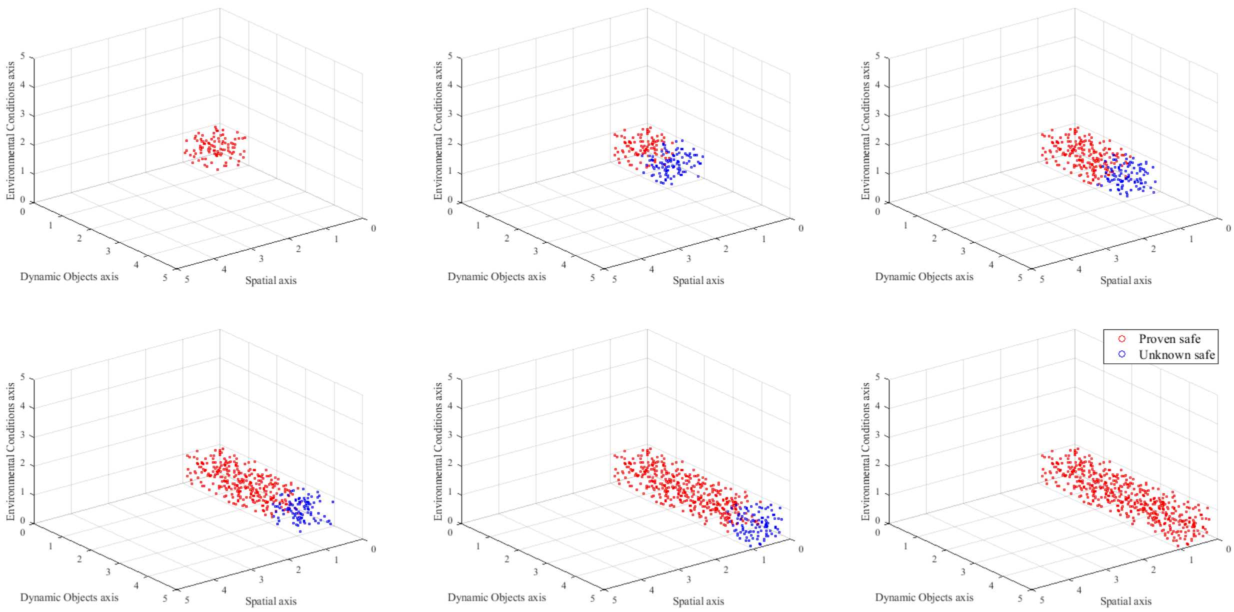


Figure 31: Proof-of-concept along the dynamic objects axis

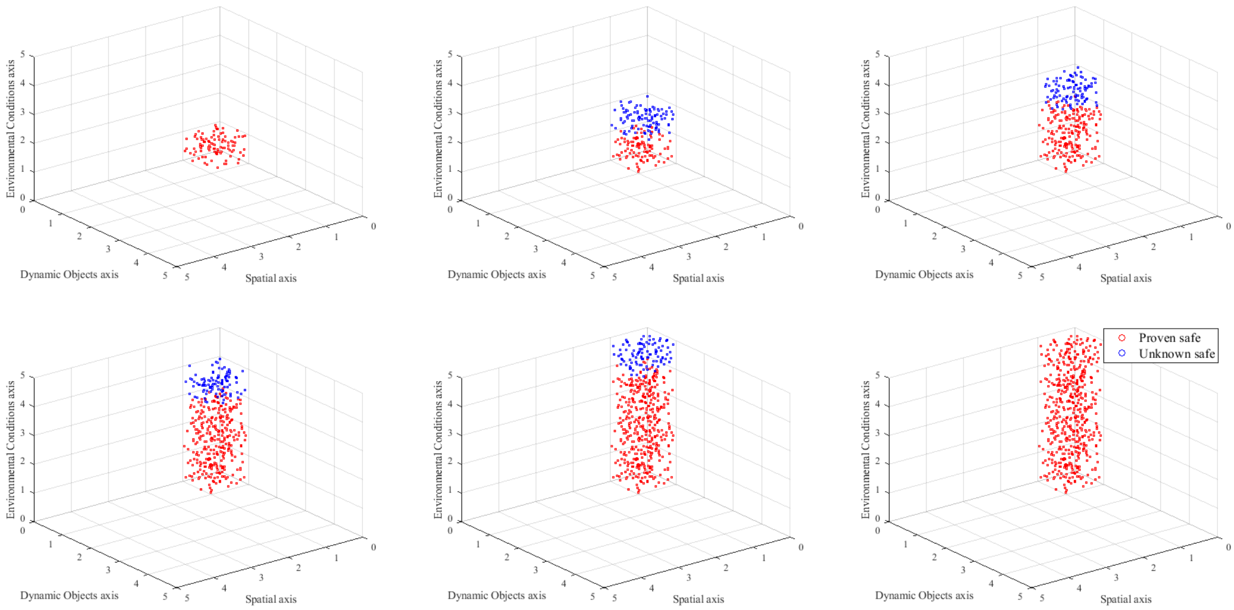


Figure 32: Proof-of-concept along the environmental characteristics axis

This proof-of-concept method can be used not only to initially demonstrate the safety of SADRs, but also in the case of operating system updates. When the operating system is updated, it cannot be assumed with certainty that the safety of the SADR on the public sidewalk will remain the same in all cases. By starting at the easiest setting and gradually admitting the SADR to more difficult ODD levels again, this safety can be proven again, without a large part of the network being unavailable for a while: the old operating system had already proven safe and can remain operational until the newer software update has demonstrated the same level of safety.

One could argue that it is almost impossible to find infrastructure in the Netherlands and other areas that gradually changes from easy to difficult, to gradually expand the rideable network, and that this is exactly what would be needed in this proof-of-concept approach. There are multiple solutions to this. For example, it is possible to allow a delivery robot to drive a maximum of 5% of a route outside the 'known safe' operating area. In addition, a teleoperator could be asked to monitor driving under these specific operating conditions in order to ensure that the robot is stopped if necessary for safety reasons. Dependent on the number of ODD difficulty levels that will eventually be used in the dynamic assessment framework system, it could also be decided to only allow deviations within the 5% deviation that are, for example, a maximum of one level away from the 'known safe' level. The easiest solution is to let the SADR take up a safe position on the sidewalk and wait until conditions have changed such that it can be reasonably expected that the SADR can continue safely on its way. That SADRs can interrupt their route is an advantage SADRs have over AVs.

6.5.5 Living lab approach

According to Bergvall-Kareborn, Hoist, & Stahlbrost (2009, p. 1) "a living lab is a gathering of public–private partnerships in which businesses, researchers, authorities, and citizens work together for the creation, validation, and test of new services, business ideas, markets, and technologies in real-life contexts" and is a Living Lab "an environment in which people and technology are gathered and in which the everyday context and user needs stimulate and challenge both research and development, since authorities and citizens take active part in the innovation process". In this section will be argued why the living lab approach suits the further evolution and use of the design suggestion for a dynamic SADR assessment framework as proposed in this research. The actual development of a living lab is left open for further research.

Quak & Nesterova (2021) argue that the focus of a living lab is on practical implementation, learning and improvement, which is in line with what is proposed in this research. Because to date very little is known about the actual performance of SADRs in real world conditions, and simulated performance cannot be assumed to be safe with certainty (Shetty et al., 2021), structured sidewalk testing can overcome this challenge. Not only about the SADR performance, but about numerous aspects of SADRs in the public environment in the Netherlands little or nothing is known. A living lab is more sustainable, more educational and more adaptive to changes due to the involvement of different stakeholders, the predefined goals, the longer duration and the iterative development approach of living lab projects compared to standalone pilots (Quak & Nesterova, 2021). The advantage of the dynamic assessment framework proposed in this research is that it focuses on learning about SADR performance and the collection of data, which combined with the living lab approach allows to learn about the complete traffic system that includes SADRs. The third and perhaps most important focus of living labs is the possibility for improvement. Improvements to the dynamic assessment framework can be made as more is learned about the performance of the robot in real world environments. SADRs are also far from being fully developed and improvements need to be properly evaluated in public environments. In addition, from a technical perspective SADRs can perform safely, but there are social developments which could make the system requirements stricter or less strict. The ongoing development of the innovation, the assessment framework and the system requirements fit in well with the iterative approach used in living labs. The presence of a learning and development approach is also an important requirement for the Netherlands Vehicle Authority to test automated functions on public roads (van der Stoep, 2022). Adopting a living lab approach ensures this learning and development approach. Additionally, a long-term test can more easily open up research opportunities and subsequent follow-up studies.

6.5.6 Dynamic assessment framework functionalities benefits per user group

This section explains by whom and in what possible ways the dynamic assessment framework can be used. As described with the Living Lab approach (Section 6.5.5), a field test with the dynamic assessment framework is envisaged, in which businesses, researchers, authorities and citizens work together. These stakeholders will use the digital system in different ways and for different purposes. First of all, the visualisation of relative ODD difficulty on the digital maps is important for all stakeholders because it gives them an insight into the relative difficulty of geographical locations and how this may vary over time. The overview of ODD level explainability at link level also helps stakeholders understand why or why not an SADR is allowed to a certain sidewalk section at a certain moment in time. For researchers and the SADR operator, the data side of the dynamic assessment framework system is important, because insights from this data can be used to further improve the dynamic assessment framework, to better determine relative ODD difficulty and to improve the SADR performance. The additional advantage for the SADR operator is that, if the dynamic assessment framework is built in the same programming language as the navigation system of the SADR, the digital assessment framework can be directly linked to the SADRs navigation system. Subsequently, based on geofencing (for example with the Google Geofencing API (Google, n.d.-c)), SADR access to specific network links can be restricted if these links are too far outside the 'known safe' operating area in terms of conditions. In addition, gaining practical experience on SADR performance is a good way for SADR operators to verify and improve their SADR simulation models. Improvements to these models can in turn lead to a better performance of SADRs in public space. The incident reporting overview has added value for authorities because it provides an overall view on the locations where incidents with SADRs occur more frequently. Analysis of the underlying data, i.e. the system conditions monitored at the time of the incident, and a qualitative description of the incident provided by the SADR operator can then be used to determine whether action should be taken in response to the incident. The storage of video footage of the incident would be interesting for research purposes, but this could prove problematic in terms of privacy legislation. The incident report overview should be publicly accessible, because it is also good for other stakeholders, such as citizens, to know why a certain incident occurred and whether (incorrect) human actions influence the likelihood of such incidents. In this way, citizens can learn from the incident reports on how to interact with SADRs as well. The public accessibility of the incident reports is an incentive for the relevant parties to take mitigating actions to prevent similar incidents from happening.

6.5.7 Manoeuvrability and obstacles

According to the Cambride University Press (n.d.), manoeuvrability is the "quality of being easy to move and direct". In this research the manoeuvrability construct is used to indicate the degree to which an SADR is operationally competent to move, dependent on spatial, dynamic objects and environmental conditions. Manoeuvrability can be objectively measured, as will be discussed in Section 6.5.8. From sub-question three was concluded that the visibility distance, edge case occurrence and ability to manoeuvre are important, ego-vehicle related constructs that determine the safe driving ability an SADR. Edge case occurrence is a direct consequence of a lack of data, and is a problem whose impact will likely decline over time when more data is gathered. Visibility distance however, is a very difficult construct to measure and assess. The visibility distance that is reduced by the presence of obstacles is context dependent, as will be demonstrated with the following example. In the first case, an SADR drives and during its movement gets ever closer to a possible obstacle. This means that the visibility distance is gradually reduced further, but until that moment the SADR has had time to observe and monitor its surroundings. The second case is the occasion when an SADR drives around a corner and encounters the same obstacle that it must deal with. This is not a gradual reduction in visibility distance, but a sudden large reduction, which might have a different effect on the detecting and planning abilities of an SADR in comparison to the case where the robot approached the obstacle on a straight section. Thus, one of the reasons why it is difficult to determine the effect of a reduction in visibility distance is because the effect is dependent on what has happened prior. A second difficulty is determining the effect of obstacle size on visibility distance reduction. It is likely that taller objects will take away a larger part of the field of view of an SADR. However, from a certain height, this reduction will be marginal and there is no linear relationship between obstacle height and the reduction in field of view and therefore visibility distance. It is currently not possible to make valid statements about the actual effect of the visibility distance reduction on the SADRs ability to detect and plan a path. The assumption for now is that larger objects lead more quickly to a reduction in visibility distance. As already mentioned, manoeuvrability can be objectively measured, which is why in this research manoeuvrability will be used as a starting point to determine ODD difficulty. Gradual further development of the dynamic assessment framework and collected data can then be used to improve the assessment framework and for example include visibility distance when determining ODD difficulty.

It can be assumed that obstacle size has an effect on the manoeuvrability of an SADR. Larger obstacles are more likely to block the passage of an SADR than smaller obstacles. However, the width of the sidewalk is an important component: if the sidewalk is ten metres wide, the impact of a three meters wide obstacle is smaller than it would be if the sidewalk were only three and a half metres wide. In addition to the size of an obstacle, the rotation of the obstacle on the sidewalk also influences whether an SADR can drive over it, as visualised in Figure 33. The two objects displayed are identical, however the right obstacle has been rotated 90°, which is why the right, green sidewalk is passable and the left, red sidewalk is not. Next to obstacle rotation, the position of the object on the sidewalk affects whether the sidewalk is passable (Coppola & Marshall, 2021).

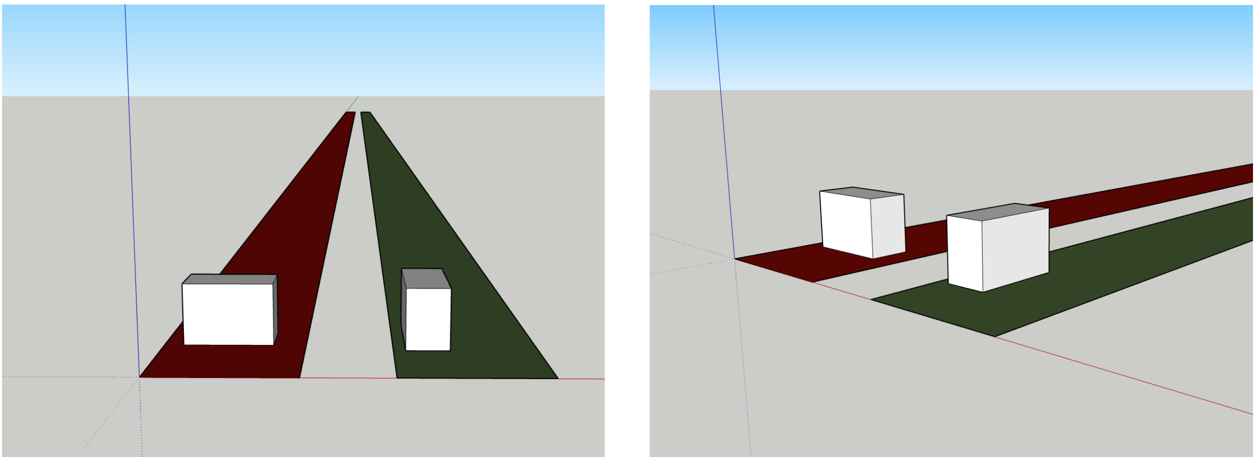


Figure 33: Illustration impact obstacle rotation

In the proposed assessment framework is therefore the following relation for 'sidewalk obstruction' caused by a single object assumed:

$$\textit{Every obstacle is generalised to its length, width and height} \tag{1}$$

$$Obstacle\ rotation = \alpha \tag{2}$$

$$\beta = 180^\circ - \alpha \tag{3}$$

$$Sidewalk\ obstruction = width * cos(\alpha) + length * cos(\beta) \tag{4}$$

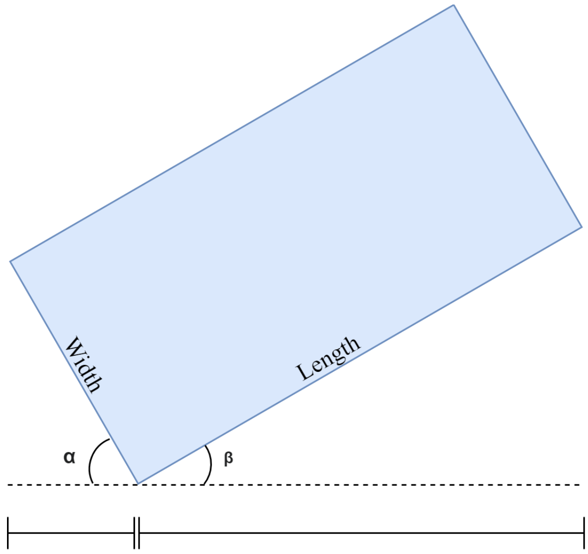


Figure 34: Top view of sidewalk obstruction

Obstacles and Spatial Axis

The direct obstruction effect of static objects on the sidewalk, resulting from Layer 2 in the 6-Layer Model, can be determined with the above insight. The research by Coppola & Marshall (2021) investigated the effect of static objects on the clear width of the sidewalk. Their focus was on the extent to which a disabled person in a wheelchair could drive over the sidewalk, which can be compared to the passage of a delivery robot. Two metrics used in the research are the 'Sidewalk Minimum Clear Width', defined as the narrowest passage point within a sidewalk polygon, and the 'Sidewalk Minimum Clear Width Accounting for Static Obstructions', which additionally to the Sidewalk Minimum Clear Width takes the impact of obstacles on the clear width for pedestrian access into account. This latter metric does not only take the positioning of one obstacle on the sidewalk into account, but also the possible presence and obstruction of other objects. This metric can be put into words as the maximum minimum passageway for a sidewalk polygon. A visualisation of this metric in various contexts is given in Figure 35.

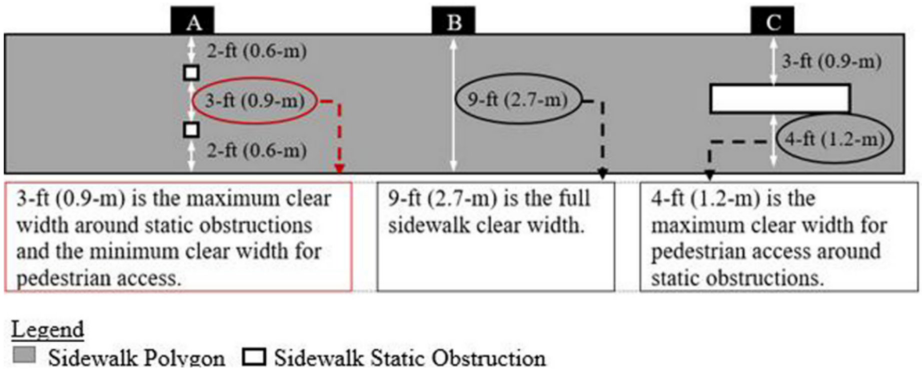


Figure 35: Top view maximum clear width visualisation, copied from Coppola & Marshall (2021, p. 207)

Coppola & Marshall (2021) used existing Geographical Information System (GIS) data from Cambridge and Quantum GIS software to successfully determine the sidewalk clear width at link level for a complete city. For passage with a wheelchair, the narrowest passage on a sidewalk section is the most telling, because any wider passage should not pose a problem for passage. From an SADR, as mentioned several times, nothing is known yet about this kind of relationship. Of course, the narrowest passageway is also important for an SADR to determine whether an SADR can physically fit between obstacles. In addition, we do not yet know whether a sidewalk at link level with four identical obstacles in terms of narrowest passageway can be classified as equally difficult ODD as if there were only one obstacle, as compared to the wheelchair case. This relationship can be further investigated by not only taking the smallest value for the Sidewalk Minimum Clear Width Accounting for Static Obstructions, but to take for example the four smallest values and determine whether those three additional parameter values in a regression analysis (see Appendix G) are statistically significant in terms of their impact on the manoeuvrability of an SADR. The approach established by Coppola & Marshall is a promising method to objectively map all static Layer 2 objects on the sidewalk, from which subsequently an ODD difficulty for the spatial axis can be determined.

Obstacles and Dynamic Objects Axis
Although the approach for mapping roadside structures initially also seems applicable for determining the ODD difficulty along the Dynamic Objects Axis, this is not entirely the case. The approach would in principle be applicable to all obstacles resulting from Layer 4, which do not move at a specific moment in time (velocity = 0). These objects have a similar impact on the manoeuvrability of an SADR as explained above, because they are static. However, difficulties arise when objects have a speed, move, and thus do not form a constant obstruction. Context is again very decisive for the manoeuvrability of an SADR. Consider, for example, pedestrians walking next to each other and forming an obstruction, but in the case of an approaching SADR they start to walk behind each other to make space. The associated clear width changes immediately. Pedestrians can also step off the sidewalk in a quiet environment, and suddenly no longer form an obstruction. Additional metrics to the 'Sidewalk Minimum Clear Width Accounting for Static Obstructions' are necessary to determine an accurate ODD classification for the Dynamic Objects Axis. Two additional metrics will be proposed: the 'average busyness by VRUs' and the 'total occupied sidewalk area'.

The average busyness by VRUs can be measured as the number of VRUs that have passed a sidewalk section per unit of time divided by the surface of the sidewalk polygon. The smaller the unit of time specified, the more accurate the monitored conditions are for an SADR while driving. The advantage of this metric is that it is objective, easy to measure and takes the sidewalk polygon into account. The disadvantage of this metric is that it does not provide insight into the exact positioning of the VRUs and that therefore smaller values of the average busyness metric can in reality be more obstructing than higher values.

The total occupied sidewalk area can be measured as the percentage of the sidewalk polygon that is occupied by obstacles. The advantage of this metric is that it is again objective and easy to measure. In addition, if the metric is updated in the same time frame as the average busyness by VRUs is determined, and a standard calculation size is used for each VRU, it is possible to calculate which part of the sidewalk is occupied by non-VRUs. Further research can reveal whether this division in human and non-human obstacles has a different effect on the ability to manoeuvre by an SADR. How to obtain data on the proposed metrics will be further discussed in Section 6.5.10.

6.5.8 Assessment framework performance indicators

To prove an SADR as safe it is important that there are metrics to objectively measure this performance. The body of scientific literature on performance metrics for ADS-equipped vehicles is growing, and some of these performance metrics can be used for SADRs as well. However, ADS-equipped vehicles are being developed to gradually replace human driven vehicles. It makes sense to have performance indicators that compare the performance of AVs to the performance of a human driver. However SADRs in the future will replace part of delivery vans or other vehicles used for delivery, their performance cannot one on one be compared to these vehicles. SADRs compared to delivery vans drive on a different road surface and have a different operating speed. In addition, ADS-equipped vehicles are subject to the same regulations as the human-driven vehicles they replace. For SADRs on the sidewalk, there are no traffic rules yet, which means that violations of them are not possible either. Wishart et al. (2020) created a comprehensive list of driving safety performance metrics for AVs based on an evaluation of 50 references that included one or more performance metrics. In this section, the performance metrics deemed relevant during an initial field test with the dynamic assessment framework are explained in more detail. In Appendix H, a motivation is given for the metrics identified by Wishart et al. that are not (yet) taken into account for an initial field test with the dynamic assessment framework. This section concludes with a passage on how the development of performance metrics over time can contribute to increasing the safety of SADRs on the public sidewalk in the Netherlands.

- Minimum Safe Distance Violation (MSDV), "an instance in which the actions of the ego vehicle result in encroaching upon its safe boundaries with another (safety-relevant) entity within the scenario environment, as defined by current velocities and acceleration capabilities of both entities" (Wishart et al., 2020, p. 3).
- Proper Response Action (PRA), "an instance of an action (longitudinal and/or lateral acceleration) taken by the ego vehicle to restore itself to its calculated safety boundaries after a safe distance violation has occurred." (Wishart et al., 2020, p. 4).

- Minimum Safe Distance Calculation Error (MSDCE), "the accuracy of the ADS calculation of its safety boundaries with respect to other safety-relevant entities in comparison to ground truth" (Wishart et al., 2020, p. 5).
- Collision Incident (CI), "an instance of the ego vehicle being at fault in a collision, as determined by an examination of the data from the on-board and/or off-board sensors, potentially including the vehicle event data recorder, along with the police report" (Wishart et al., 2020, p. 6).
- Rules-of-the-Road Violation (RRV), "an instance of the ego vehicle violating a traffic regulation that would result in an infraction or citation" (Wishart et al., 2020, p. 6).
- ADS Active (ADSA), "a confirmation that the ADS is active while executing the behavioral competency" (Wishart et al., 2020, p. 7).
- Modified Time-to-Collision (MTTC), "the time until a collision between two entities in the scenario environment would occur if both continue with the present velocities and accelerations" (Wishart et al., 2020, p. 9).

From the list of performance metrics it follows that it is important to assess the performance of the SADR itself. By only assessing situations that are the result of actions of the ego vehicle, this condition is met. The ADSA metric is important because in this way the actual driving behaviour of an SADR is assessed and control by a teleoperator is not wrongly used to classify the autonomous driving system as safe. This does not alter the fact that unsafe situations can occur as a result of steering by a teleoperator, but it is important to find out what causes this and whether the quality of the driving system or that of the teleoperator needs to be improved. The MSDCE metric provides insight into the average deviation in the perceived and actual geographic location of an SADR and can be used to determine applicable safety margins. The RRV metric indicates the need for (basic) traffic rules for SADRs. Violations and obeying to these rules can be objectively measured. The MTTC can provide insight into how often unsafe situations occur that do not immediately lead to a CI. All of the above performance metrics can be calculated. For the exact formulas is referred to the work of Wishart et al. (2020). It must be noted that data regarding the velocity, speed and exact location of individual SADRs are to be stored in the database system as well. Although their parameter values on its own do not give any information about SADR performance, the data can be analysed to investigate what has led to for example an incident.

With the performance metrics that are now considered relevant, a good first impression can be gained of the safety performance of SADRs. The living lab approach is also desirable in this context because more is learned about the performance of SADRs over time. For example, if undesirable situations arise that do fall within the rules or for which there are no rules, rules can be adjusted or drawn up. Subsequently, objective measurements can be made again. This iterative approach is necessary for an innovation that we as a society know so little about, but which we do potentially want to see implemented in urban systems. Gradually, therefore, according to what is being learned during the practical tests, performance indicators can be added that show that they contribute to the sustainable integration of SADRs in road traffic in the Netherlands. In the longer terms, these requirements could also include the subjective interaction between SADRs and VRUs, which could be measured with a feedback system integrated with a smartphone application. This line of reasoning is not further elaborated on in this research.

6.5.9 Transfer knowledge to other geographic locations

Transferability of knowledge to other geographic locations is an important requirement of the dynamic assessment framework in this research. By identifying all ODD factors with the 6-Layer Model, there is a method to map locations completely and objectively. The outcomes of the proposed regression analysis (see Appendix G) are the basis for comparing actual geographical locations at ODD difficulty for an SADR. When given two geographic locations classified on ODD difficulty in the *xyz*-space, the distance between the two points can be calculated with the three dimensional Euclidian distance (see Equation 5 and 6). If the two points in the three dimensional space are close enough to each other, one could argue that if an SADR manages to drive safely at the first location, it will most likely do so as well at the second location (see Equation 7). The knowledge gained at the first location can then be transferred to the case at the second location.

$$P_1 = (x_1, y_1, z_1) \text{ and } P_2 = (x_2, y_2, z_2) \tag{5}$$

$$d(\mathrm{P}_1, \mathrm{P}_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{6}$$

Knowledge gained at two geographic locations is deemed to be interchangable if:

$$d(\mathrm{P}_1, \mathrm{P}_2) \leq \textit{comparability threshold} \tag{7}$$

An important requirement for comparing geographical locations in terms of difficulty is that all axes in the 3D space must correspond in scale to the actual impact that the scale has on the ODD difficulty. If, for example, environmental conditions have a relatively smaller impact than the presence of dynamic objects, this should be reflected in the scales of the corresponding axes. At the moment, for illustrative purposes, the scales in this research are shown between 0 and 5. A future regression analysis can confirm or invalidate this choice, as well as that an applicable comparability threshold can be agreed upon with relevant stakeholders.

6.5.10 Required data

The proposed digital dynamic assessment framework depends on the availability of data. For the spatial data needed to assess a geographical location on ODD difficulty, it is sufficient to map the risk factors from Layer 1 and Layer 2 once, as these are in principle fixed. The earlier cited methodology from Coppola & Marshall (2021) is a promising means to do so. Temporary modifications, on the other hand, would have to be constantly monitored and updated in order to determine the most objective ODD difficulty for the spatial component. A cost-effective approach is to have one involved stakeholder responsible for keeping track of the locations and duration of temporary modifications (Layer 3) in the digital system. In this way, an SADR is informed about that the actual state of a location deviates from what is expected, and that a certain sidewalk is less or no longer passable. The latter information can also be added to the system. Data on current weather conditions (Layer 5) such as temperature and wind speed is nowadays locally available in real-time. Current precipitation data and predictions for precipitation are also getting better and better, think for example of the Dutch application of buienalarm.nl, which warns people based on their geographical location when rainfall approaches (Buienalarm, n.d.). The use of such a predictive system can prevent an SADR from travelling under conditions that lie outside the known safe conditions. The reason why existing datasystems on environmental conditions can be used is twofold. First, because weather conditions differ only marginally over certain geographical areas, there is no need to measure the environmental conditions on every street corner. Second, by using appropriate abstraction of the scales on which the environmental conditions are measured, as argued in Section 6.5.1, this need to measure the environmental conditions on every street corner further declines.

Measuring or digitising the elements originating from the dynamic object axis in the proposed assessment framework, to come up with an objective determination of ODD difficulty related to dynamic objects, will be the biggest challenge. It would be most detailed if all dynamic objects with real-time associated characteristics in the environment of an SADR were included in the digital system, as if a digital twin of the physical environment was made. According to Jones, Snider, Nassehi, Yon, & Hicks (2020, p. 36) a digital twin consists of "a physical entity, a virtual counterpart, and the data connections in between" and can a digital twin be a means to improve "the performance of physical entities through leveraging computational techniques". Although it is a means of mapping a space as objectively and completely as possible, creating a digital twin implies that all dynamic objects must be connected and share real-time data. This could potentially be possible on a small scale in a test environment, but is not feasible for a large network or, for example, an entire city. Again the real question is: what actually needs to be measured to be able to make statements about dynamic objects that influence the driving ability of SADRs?

As with the monitoring of weather conditions, consideration can be given to the use of existing data on dynamic objects. A good example is the visit data that Google has been making available to companies via Google Maps for a few years now (Google, n.d.-b). Visit data includes how busy a business location or shop typically is during different times of the day, how much time customers on average spend at the location and how long customers usually have to wait before they receive service during different times of the day. More recently, Google has expanded this service with live visit data, which provides real-time insight into how active a location is. It should be noted that this data is only available for businesses and retail locations, and only if Google has sufficient data. In addition to this data for shops and businesses, Google also provides data on area busyness, which is determined by live busyness trends from local places to determine an overall busyness level for the specific area (Google, n.d.-a). Although this information, due to privacy considerations, is not available at a link level, as posed in this research, it provides insight into the data possibilities that do exist. A solution for monitoring current traffic density on a sidewalk (link) level is the use of digital tools, such as scanners, counters and counting sensors (CityTraffic, n.d.). With these tools the average busyness by VRUs can be determined, as discussed in Section 6.5.7. The disadvantage of all the possibilities mentioned in this paragraph is that only the number of VRUs can be measured in this way and that this says nothing about their positioning nor about other ways in which the sidewalk is used and obstructed, such as by the parking of modalities. A means to map this last category of objects is to place lidar sensors at strategic points near intersections. These can detect within a range up to more than 200 metres, on the basis of which the Sidewalk Minimum Clear Width Accounting for Static Obstructions and the occupied sidewalk area can be determined (Bosch, n.d.). P. Lassner notes that Bluecity's lidar solution has a 100 meter radius and that detection and classification is possible closer to the 75 meter radius (personal communication, May 30, 2022).

The technologies mentioned range from high level insight to detailed insight, but therefore also from cheap to expensive to integrate. To date, it is too expensive to place a lidar sensor on every street corner, for example, but this could change in the future with technological advances. In addition, certain technology would not only benefit operating SADRs, but autonomous vehicles could also reap the benefits of such digital safety tools. The use of such tools also leads back to the central follow-up question of this research: namely, determining what factors actually increase the risk of errors being made by an SADR and determining the actual impact of using digital technology on the abilities of SADRs (see Appendix G). When more clarity is obtained on this, the question can also be answered as to what should be digitised, how it should be digitised and to what extent measurements should be real-time. Gradual testing of digital tools such as lidar fits well with the living lab approach proposed in this research. Taking into account that it is difficult to determine a suitable method to map the presence of obstacles on the sidewalk, it can be decided to test in practice on geographic locations where statistically the probability of obstacles on the sidewalk is smaller. For example, residential areas with many low-rise buildings and detached houses, which means that there are fewer inhabitants per square kilometre. Also in residential areas where people have a front and back garden the probability that people park their bicycles on the sidewalk is reduced.

6.5.11 Overview of dynamic aspects

In this research, the term 'dynamic' is cited and discussed on several levels. Therefore, this final section that discusses the complete design suggestion will briefly conclude with a clear overview of which aspects of the dynamic assessment framework are dynamic, and in what way. The four concepts that are posed as dynamic in this research are: the ODD difficulty of a geographic location, the authorised ODD difficulty which an SADR may undergo, the capabilities of SADRs and the performance indicators that are used to assess SADRs on their safe performance.

First, the ODD difficulty of a geographic location can change over time because of the spatial, dynamic objects and environmental location characteristics that can change. In addition, the use of (external) data can ensure that this associated ODD difficulty level is lowered for a specific location. From the three axes in the ODD classification space (see Figure 22) the dynamic objects axis is the most dynamic, after that the environmental conditions axis, and the spatial axis is the least dynamic axis of the three. Whether this spatial axis data from Layer 1 and Layer 2 should be updated weekly, monthly, yearly or bi-yearly is left for further research. Layer 3 information should be updated as real-time as possible because it has a direct effect on the known network of the SADR. A second way in which ODD difficulty can change is when, by using the database system over time, data is collected and the ODD difficulty classification method is recalibrated, for example by means of the regression analysis (see Appendix G).

Second, the authorised ODD difficulty that an SADR may undergo changes over time. As explained with the proof-of-concept approach (see Section 6.5.4), when an SADR proves to be able to drive safely in easy conditions, progressively more difficult levels are allowed to drive. This permitted ODD difficulty can also be restricted over time. For example, after an update to the operating system or after an unsafe performance or incident where it is decided to relegate an SADR to easier ODD difficulty.

Third, the capabilities of SADRs will change over time. On both the hardware and the software side, manufacturers will continue to modify their delivery robots. This will not always necessarily improve performance, but should be thoroughly evaluated in its objective context. When delivery robots are technically able to undergo more difficult terrain, for example dirt tracks or inclined slopes, the ODD difficulty level of a geographic location changes because of a change to the capabilities of the delivery robot.

Fourth, the performance indicators and thresholds that are used to assess SADRs on their safe performance can change over time. The fact that a delivery robot performs safely from a technical point of view does not provide the complete overview. It could be that a relevant performance indicator is missing. Additionally, a technically safe performance does not mean that other road users will also experience this performance as pleasant. Rules of conduct and traffic rules will develop over time, which means that the robot's performance will also be assessed against other rules over time. In addition, it is not yet known which threshold values should be used for the

performance indicators drawn up in Section 6.5.8. All these unknowns will be discovered and coloured with the proposed living lab approach.
As can be noted from the above, the different dynamic aspects are interrelated and the complete system with SADRs on public sidewalks can evolve over time. A high-level Process Map where these interrelations become clear is given in Figure 36, where the operational functioning of the living lab is visualised in a Process Map. .

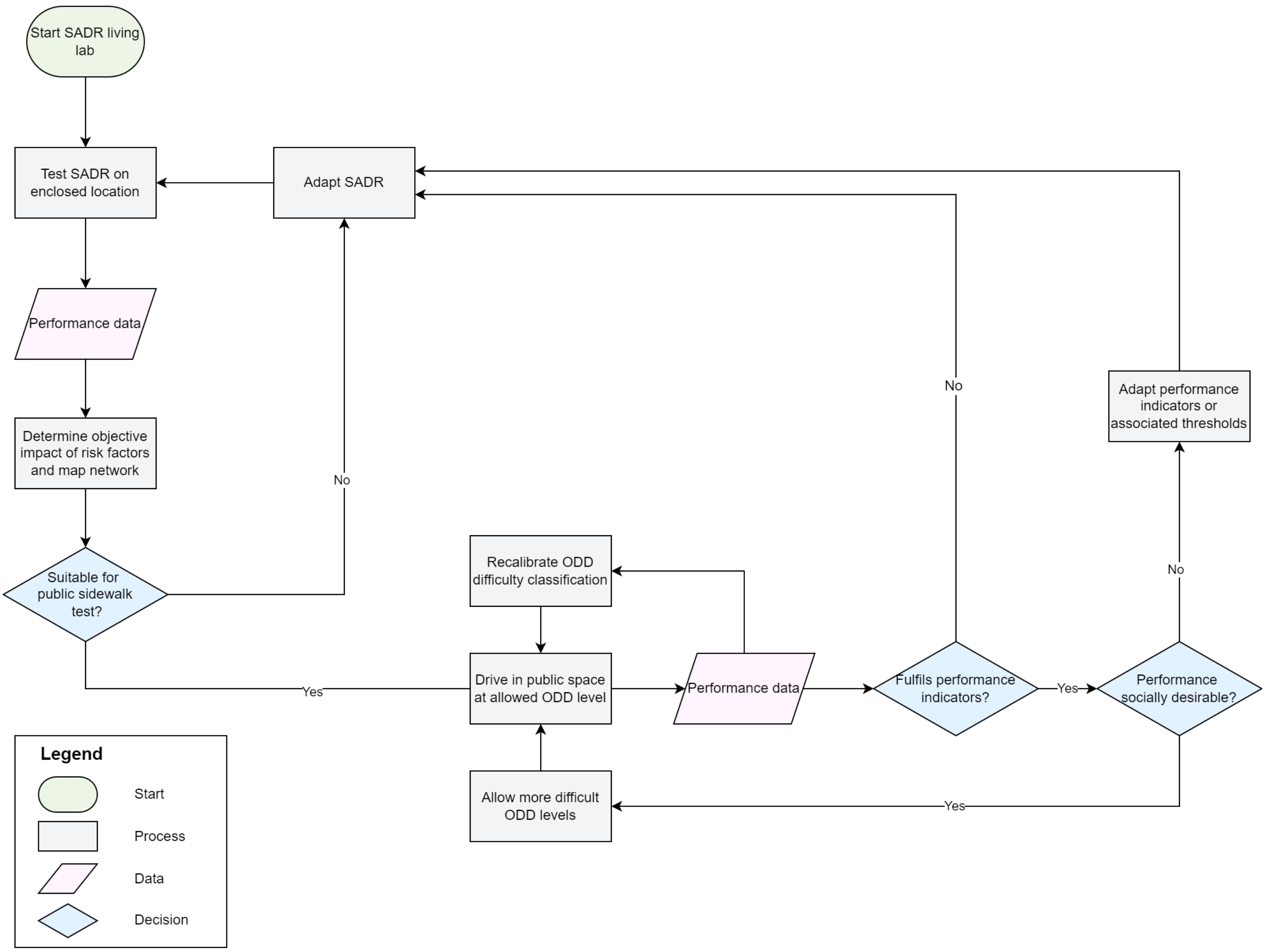


Figure 36: Process Map dynamic living lab

From the figure all mentioned dynamic aspects related to the dynamic assessment framework can be identified. By following the adjacent arrows it can be concluded which operational decisions originate at certain process steps.

6.6 Design requirements evaluation

In this section will be briefly reflected on the drafted design requirements and objectives in order to verify whether the drafted design actually meets the requirements set. This will be done by presenting the requirements and objectives in table form, Table 11, and indicating in which section of the report the design requirement is implicitly referred to. Below the table key insights will be elaborated upon.

Table 11: Design requirements evaluation overview

Requirement / Objective	Report Section
FC1. The framework must use objective descriptions of the infrastructure and environment	Section 6.2
FC2. The framework must use SADR specific risks	Section 6.3
FC3. The framework must allow for performance requirements modifications through time	Section 6.5.11

Table 11 continued from previous page

FC4. The framework must allow for SADR modifications through time	Section 6.5.4
FC5. The framework must be able to determine different ODD levels	Section 6.5.1
FC6. The framework must allow for ODD-level changes over time for geographic locations	Section 6.5.1
FC7. The framework must store performance data along ODD conditions	Section 6.5
NFC1. The framework must be integratable with existing route planning modules	Section 6.5.6
NFC2. The framework data must be protected with state-of-the-art security technologies	
FO1. The framework should use a complete description of the infrastructure and environment	Section 6.2 Section 6.5.10
FO2. The framework should explain a certain ODD level classification	Section 6.5.1
FO3. The framework should be able to compare different geographic locations on the basis of location characteristics	Section 6.5.9
FO4. The framework should include a means for incident reporting	Section 6.5.3
NFO1. The framework should be understandable to all different stakeholders	partially Section 7
NFO2. The framework should be user-friendly	

The first thing to note is that the FCs are fully developed, albeit theoretically. No functioning dynamic assessment framework has been developed in this research. NFC2 is therefore not further discussed in this research. From the actual development phase, the step that should be carried out after validation according to the DSR cycle, data protection is an important component. Both non-functional objectives, NFO1 and NFO2, are not fully explained in this research. In the expert validation, with a group of stakeholders, an advance is made to the understandability of the proposed dynamic assessment framework. User-friendliness (NFO2) is something that can only be determined when a functioning framework has actually been developed. Here also applies: if the dynamic assessment framework is developed, methods will have to be developed to determine how user friendliness and the understandability of the framework for all relevant stakeholders should be shaped.

7 Validation

In this section the validity of this thesis research by means of semi-structured interviews with different relevant actors in this research field is discussed, to answer the sixth and seventh sub-questions:

6. What expert perspectives exist regarding the establishment of different levels of difficulty within the Operational Design Domain of SADRs?

7. What expert perspectives exist regarding the applicability of the proposed dynamic assessment framework design suggestion?

Relevant actors in this study can be divided in two groups: the academic audience and the practitioner audience (Hevner, 2007). The stakeholders are relevant because they are involved in the assessment of innovative vehicles, their approval, they have expertise in the field of road safety and/or they have expertise in the field of automotive innovations. It has been chosen to conduct semi-structured interviews because the research area is complex and little about it is known (Saks & Allsop, 2012), because the open nature of the research questions can lead to an in-depth conversation pursued from the given answer (Newcomer et al., 2015) and because the flexibility of the semi-structured interview approach can lead to new insights (Saks & Allsop, 2012; Newcomer et al., 2015), It should be noted that the initial intention of the study was to conduct two focus group sessions, but due to the busy schedules of experts it was decided to switch to a semi-structured interviews approach. The purpose of the validation is to evaluate whether experts think the research contributes to the problems identified. The interview questions are divided into a part where the research steps and line of reasoning followed in the research is validated and a set of questions that focuses on the digital suggestion for a dynamic assessment framework for SADRs. If there are inaccuracies in the process that have led to the design suggestion, this is important to consider for follow-up research. According to the outcomes of the validation process step the theoretical design suggestion can be adapted accordingly before proceeding to the actual development and implementation of the design suggested in this research.

7.1 Validation interview

Prior to the interview the interviewees received a video per email of about seven and a half minutes explaining the broad outlines of this research[1]. This way, all interviewees have the same information provided by the researcher at the start of the interview and no interview time is lost. If the interviewees native language was Dutch, the interview was conducted in Dutch. The interview questions that were prepared for the semi-structured interviews can be found in Appendix I. The abbreviation SADR was avoided during the interview, the term delivery robot was used. All interviewees are familiar with delivery robots and stated that they watched the introductory video before the actual interview started. Multiple interviewees explicitly mentioned that they considered the introductory video interesting and a pleasant way to be informed about the contents of the thesis validation. The interviewees and their relation to this research topic are summarized in Table 12

Table 12: Experts interviewed for validation

Name	Position
Michael Chin	Freelancer
Tom Alkim	MAP Traffic Management
Kirsten Pouwels	Ministry of Infrastructure and Water Management
Pieter van der Stoep	RDW
Nicole van Nes	SWOV and TU Delft
Haneen Farah	TU Delft
Marjan Hagenzieker	TU Delft
Bart van Arem	TU Delft
Frans Tillema	HAN University of Applied Sciences

[1]The video can be accessed via: https://youtu.be/PMDGgR4x5wI

7.2 Validation research process and ODD difficulties

7.2.1 Existing barriers

Nine out of nine interviewees agree with the barriers presented to them. An additional barrier mentioned by one of the interviewees is that "there is a lack of political leadership", which lags behind innovation in the mobility field. Another interviewee called this "a cultural difference with other countries". In addition, one of the interviewees nuanced the barrier that delivery robots cannot be tested on public sidewalk. "A means to test delivery robots on the public sidewalk exists through the Experimental Act (Experimenteerwet), which is intended for vehicles without a driver". However, the interviewee acknowledges that "the procedure in that act is very difficult and that this indeed creates a barrier". Two interviewees added that the lack of objective performance data in the procedural application for a test under the Experiments Act is a reason for rejection. One interviewee mentioned that "the identified barriers contribute well to a better understanding of the positioning of this research".

7.2.2 Different levels of difficulty within the Operational Design Domain

Eight of the nine interviewees agreed that there can be different levels of difficulty identified within the Operational Design Domain, one interviewee replied that (s)he had no opinion on this. A selection of the interviewees' statements is: "I agree, and this is a very interesting insight regarding the term ODD", "the hypothesis that there are different degrees of difficulty seems reasonable, but that has to be backed up by empirical research", and "I agree, but that is partially the case because autonomous driving technology is not yet capable of everything".

The approach where risk factors for delivery robot performance have been identified from the literature on AVs and AV system components is assessed reasonable, but the effects of the risk factors could be a lot smaller / less critical then they are for AVs due to the lower operating speed and shorter braking distance of delivery robots in comparison to AVs, as multiple experts notice. One of the experts noted that "you should reflect on the use of international research publications in defining risks for the case of the Netherlands". Additionally there will be additional risk factors for delivery robots that cannot be subtracted from literature because delivery robots drive on different road surface, from which other risks will originate. Because of the lower operating speed of the delivery robot and users of the sidewalk, there is more time to process the available information. One of the experts explicitly disagreed with "comparing AVs and delivery robots on risk profile because in terms of size, weight and speed the vehicles are physically not in the same risk profile. Only at system level can it be said that they use the same technical components such as lidar and radar." One interviewee mentioned that using AV risks is "a good approach, but that on the other hand, a minimum acceptable level of functioning of a delivery robot and traffic rules should also be taken into account". This notion will be further discussed in Section 7.2.3.

As multiple interviewees address, risks do not only arise from the physical presence of traffic participants. It is also about the human interaction and human expectations of delivery robots, such as: do pedestrians know that a delivery robot could be approaching? Also, "consistency by design in the delivery robot behaviour will affect the ODD difficulty by increasing the expectancy of people". The number of events that conflict with road users is reduced and the delivery robot performance is improved. One of the practitioners additionally stated that, regarding the human robot interaction "we will not know what the human interaction with delivery robots will be until we observe it in practice". Cultural differences have not been taken into account in this framework currently. For example a MSDV critical threshold could be different between different countries. Additionally this relates to Section 6.5.9, where it is argued that if two locations are close to each other in the three dimensional representation, it is likely that if an SADR manages to drive safely at the first location, it will most likely do so as well at the second location. Two interviewees stated that this might be the case within countries, but that this cannot be that easily substantiated across country borders.

Most of the interviewees are not familiar with the 6-Layer Model, but they mention that it looks complete, which is mostly because of the presence of the digital connectivity layer. One of the interviewees noted that "companies active in the AV industry see increasing value in connectivity and cooperation in that field", and it is discussed that connectivity will most certainly be a means to improve vehicle performance in the future. The same interviewee gave an example from his/her own experience where an AV could not proceed its route due to a temporary modification to the infrastructural elements (Layer 3). Another interviewee stated that (s)he "really likes the static and dynamic division, and that by using the 6-Layer model I [, the researcher,] do capture most of the things that would effect a delivery robot in its environment". One interviewee commented that "the causal relations work in both ways: given the delivery robot a certain ODD can be difficult or easy, but the delivery robot can be modified and improved over time and change the ODD difficulty of the exact same location". Another interviewee described the causal representation as "super" and notes that the next step is to empirically find out "if it works, maybe not all relationships will be causal or direct, some will have correlations, but starting this way is very appealing". The part of the conceptual causal model on risks that originate from the delivery robot characteristics itself is mentioned as a good addition to the 6-Layer Model according to several interviewees. One interviewee adds that "next to the quality of the delivery robot components, the position of sensors and cameras on a delivery robot also affects its ability to detect" and should therefor be taken into account. Currently this risk factor is not included in the overview. One of the interviewees noted that "other delivery robots and glare should be added as factors to the causal conceptual representation". The same interviewee noted that "the factors in Figure 19 are not necessarily risks but elements that could potentially deteriorate the SADR performance and the figure title is not correct". Seven out of nine experts reacted positively to the fairly easy way in which the different risk factors present in the complete ODD on an individual level can be measured. The other two mentioned that it is good to measure factors objectively, but that they question some of the factors presented. With their current unit they may be too generic.

When asked the question if mapping and measuring the risk factors in the presented way can contribute towards a more standardised description of the ODD, seven out of nine experts agreed, two experts were in partial agreement. The following responses emerged:

- "Yes. Although qualitatively it can be visualised, empirical data is missing."
- "Compliments on the stepped nature of the research. Methodologically this is 100% a very good way of peeling this problem off. I do have doubts whether this is going to lead to a viable methodology, especially if we have to do this for every sidewalk section in the long run. It is still very complex, but the methodology is good."
- "Yes, certainly this contributes to standardisation. Why do you doubt the question? Any degree of nuance you have when describing something, especially including standards, is useful. The follow-up question is: is that standard well chosen? From a continuous learning perspective, this is less relevant. The standard may change over time as a result of new knowledge gained."
- "This certainly contributes to standardisation, taken into account that you have to test and verify everything. It is especially nice that you can measure factors in an easy way. Measuring is a continuation to the conceptual model. you can then actually work with it. Over time you will find out what is important and what is not, and step by step you can make the conceptual model better, perhaps simpler if you know the connections. What you propose is very appealing to me."
- "The layering of what is proposed certainly helps."
- "Yes, absolutely. The factors identified are also things that if you have high definition digital maps, you want to present in these maps. To describe objects in a standardised way, you also have to agree on a language. That language is universal. This is a European development and is not determined by a local party. Standardisation is very important and so is a system to do so. I see this as a complete information architecture with building blocks."
- "Any measurement method is interesting if you have fixed values in a method. But you have to verify that method empirically. This representation will not directly lead to standardisation, which is agreed at a higher decision-making level."

7.2.3 Dynamic assessment framework

The visualizations explaining the dynamic assessment framework are helpful to make sure stakeholders are on the same page. A subset of responses regarding the first impression of the proposed dynamic assessment framework is: "3D comparison is very nice and in mind a very good step to avoid having to map every street against all others. What I see as problematic is the need for data and the actual estimation of factor impacts remains a litmus test". "It is a combination of infrastructural characteristics, traffic and weather conditions that cause gaps in the ODD, there is defragmentation. That 3D representation is good". "The visualisation is very easy to understand. 3D is very nice, because it makes it manageable." Nine out of nine interviewees agreed that starting to test in the easiest operating domains and gradually scaling up to more difficult domains is the right approach. Two interviewees mentioned that "through this proof-of-concept method, system boundaries can literally be stretched".

Several interviewees mentioned that the demand for data to get and keep this system operational is a huge challenge. Especially mapping the real-time dynamic objects axis conditions will be a barrier to realising the dynamic assessment framework in the short term. As one of the interviewees mentioned, it quickly becomes expensive or incomplete, and if it becomes incomplete a regulator will demand the highest level of safety, the delivery robot must then be able to operate under the most difficult conditions before it will be allowed to enter the sidewalk for testing. However, "theoretically the proposed dynamic assessment framework looks like a sound methodology to assess the performance of delivery robots". Another interviewee said: "It is difficult to monitor the identified factors separately while driving, because one never finds the factors so bare. There is always something on or along the infrastructure." Another interviewee argued that: "the question now should not be how to represent all the elements in the ODD as completely as possible in a digital system. (S)he reasoned that at the front end, it must first be determined where testing can and will take place, and then tailored to the impact of specific risks and tailored to that specific geographic location, it must be determined which elements from that environment must be digitised." Another interviewee mentioned something similar, namely: "you have to avoid knowing everything in detail about individual ODDs, it has to be usable". Experts see it as interesting to use relevant factors for further development, but it must be borne in mind that factors that are not relevant now may become so in the future.

The link-level explainability function of the proposed framework is mostly regarded as a base functionality. Regarding the incident reporting function of the proposed framework two of the nine experts highlight the need for only an objective description of what happened. According to them a qualitative (incident) report drawn up by an operator cannot be trusted to be objective in terms of possible causes and there are several sides to a story, which should be taken into account. One of the experts calls the translatability of knowledge to other geographical locations "an exciting step". (S)he additionally states that: "The way this works is so generic that you might also be able to use it for other self-driving vehicles. You will have to adjust some things, but generically the system works like this on paper." One interviewee noted rightfully so that the comparability of locations should take into account that if the two locations plotted in the three dimensional space, and one of the two points is in a discretised class that is not yet proven safe or unknown safe, and therefore lies, as it were, two classes outside the proven safe system boundary, it cannot yet be assumed that a safe performance at the first location leads to a safe performance at the second location. Another expert mentions: "You have to test the translatability in isolation. It can be an indication of an expected performance, but it will not be directly leading. Now it is still too far away, but over time it could be used to scale up more quickly" This aspect of the dynamic assessment framework was not discussed further with two experts because by then they had already indicated that the proposed framework is far too abstract for how things actually are in reality.

The living lab approach of adapting the dynamic assessment framework over time according to what is learned is well received by the majority of those interviewed. It provides an opportunity to adapt the assessment framework to what is learned and because currently we know very little, it is not it is not likely that every assumption is a right one. According to one interviewee, "having a method is most important, because with that we can get started and we can stretch the system boundaries while learning, which is very meaningful". Another interviewee indicated that "Virtual Reality could be a possible addition to this living lab, it offers an extra element to drive around, for example if something is too uncertain to be tested in the real world". Several interviewees wonder how the dynamic assessment framework can ultimately be put into practice, because there are still many uncertainties and loose ends. The stakeholders in the living lab, assuming that all relevant parties in the field are involved, are mentioned by the researcher as parties who have operational control over the design of the system with delivery robots in their neighbourhood, including set operational boundaries. It should be noted that several operational questions are asked by the interviewees. The answer to these questions leads back to what those involved in the development and sustainment of the living lab will decide on.

The last question posed to the interviewees was if they thought that the actual development of the proposed dynamic assessment framework could contribute to allowing public sidewalk tests of delivery robots in the Netherlands. An overview of the interviewees perspectives is as follows:

- "Yes, of course. It could contribute. That whole qualitative thing we have now, which some organisations say is arbitrary, can be countered by applying your way of working in a very structured and measurable way. We must make good use of digital information. This is better than what we have now."
- "This system can contribute to tests with delivery robots, but this question should be put to the RDW. It brings a lot of nuance to the questions the RDW asks. This system can also be implemented on the road, in terms of the ODD approach. I find it valuable. Perspective is promising. You would like to lift it to a higher level immediately. European, above European is United Nations. because it is very relevant. ERTICO also tries to standardise and collect lessons learned. Incident reporting may not be their thing, but you want something that makes it much easier to document what actually happened. There are good opportunities for follow-up studies."

- “Yes, it definitely contributes. Perhaps it has already become too complicated. I understand that the factors are relevant, but where do you put the testing framework? What level of supervision is needed? Also consider the theory of meaningful human control.”
- “Yes, I think so, methodologically I think it is good to follow what you propose. This system makes that everything you make explicit you can also have discussions, for example we can find more factors. The combination of preprocessing and the safety notion that we have defined the ODD and therefore understand the risks, and a vehicle that actually causes few risks, should be able to do it.”
- “No, it does not contribute in this way. An important aspect here is who is developing this framework and on what basis? What are the interests involved? There is a scientific and a commercial side to this. It is most important that an assessment framework is and remains objective. At the moment there is a problem because there are different interests, so a neutral party is needed.”
- “Yes, especially that it is empirically quantifiable. It does need to be worked out in detail, to test how it works in practice. What helps is to visualise things.”
- “Yes, it is clear to everyone. Setting up the process step by step, stretching it and the living lab approach are exactly what this innovation requires.”

7.3 Conclusion

Based on the insights gathered from the expert interviews the sixth and seventh sub-questions of this research can be answered.

6. What expert perspectives exist regarding the establishment of different levels of difficulty within the Operational Design Domain of SADRs?

It can be concluded that concerning the development of different levels of difficulty within the ODD the experts are predominantly very positive. Based on the materials available during this study, the methodological approach is rated as complete. In terms of exact ODD difficulty content, most experts mention that there is a need for empirical evidence that quantitatively supports the hypothesis of relative levels of difficulty within the ODD. About whether the presented method is going to contribute to standardization regarding the mapping of the ODD there are two different perspectives: one is that the proposed method is a good start now and will develop over time, the other is that the method will need further development and more detail before it can be used. Several experts mention additionally that the human interaction component regarding the risks that exist for SADRs is currently under researched. This is partly due to the scoping of this research, which, because of a simultaneous graduation project on these human and delivery robot interactions, is deliberately focused more on the broader overview of infrastructural, dynamic objects and environmental performance risk factors.

7. What expert perspectives exist regarding the applicability of the proposed dynamic assessment framework design suggestion?

It can be concluded that there are two perspectives regarding the applicability of the dynamic assessment framework designed in this research. These perspectives are in line with the conclusion of the establishment of different levels of difficulty within the ODD. The majority of the experts state that they like the risk minimising and learning approach that is proposed with the living lab and proof of concept methodologies that go together with the dynamic assessment framework. Over time ODD difficulty will be mapped as accurately as possible according to the performance data gathered on public sidewalks. On the other hand, the experts who felt that the risks, on which ODD difficulty in this research is theoretically based, were not specific enough also felt that the development of the theoretical assessment framework on the basis of these factors was not good enough yet to be brought to practice. Again, empirical evidence that quantitatively supports the hypothesis of relative difficulty within the ODD will bridge the gap between the theoretical framework and its further development in practice.

8 Discussion and conclusion

8.1 Discussion

The three main points that will be discussed are the accuracy of the risk factors used to explain ODD difficulty, the difficulties when bringing the dynamic assessment framework from theory to practice, and the expert validation carried out.

8.1.1 Accuracy of ODD difficulty risk factors

In this study, the theory is explored in which the ODD is no longer approached as binary, but where, through the presence of environmental elements, a relative level of difficulty can be found between different ODDs. This theory is backed up by theoretical evidence and validated by experts in the field. During this research the line of reasoning is used that risks that deteriorate the performance of an AV can pose a similar, not necessarily equal, risk to deteriorate the performance of delivery robots. This line of reasoning is followed because AVs and delivery robots use comparable hardware and software components to execute the autonomous driving tasks. The presence or absence of these risk factors leads to relative difficulty of ODDs. However, as several interviewees in the validation process of this research noted, this insight must be substantiated with empirical data. Because of the lower operating speed of delivery robots in comparison to AVs it might be that risk factors that do deteriorate the AV performance do not necessarily deteriorate the performance of an SADR. In addition, because the operating environment of an SADR differs from the operating environment of an AV, there might be risk factors of influence on the SADR performance which have not been identified in the complete literature study on ADS risk factors. Also, there are no (scientific) publications as of yet that specify the risk factors associated with delivery robots on public sidewalks. Therefor the risk factors for delivery robots identified in this research might not be complete or risk factors might be included that do not affect the SADR performance at all. According to the interviewees this does not invalidate the theory of relative difficulty levels existing within the ODD. Risk factors may be adjusted or added to the presented overview over time. This research contributes to partially removing the encountered issues for follow-up research.

The overview of delivery robot risk factors and their causal impact on the performance of a delivery robot are identified and displayed per layer of the 6-Layer Model. In practice, this separation between the six layers is not so pronounced and the different layers may influence each other. Think of how Layers 1, 2 and 3 in the design suggestion are aggregated or think of the sixth layer with digital communication means that can possibly reduce the impact of other road user's presence (Layer 4) on the delivery robot performance. Because the different layers have been deliberately taken apart and depicted separately, Figure 18 and 19 are a simplified representation of reality. Again, it must be noted that empirical research is the next step to find out how and which factors contribute to relative ODD difficulty. The qualification and quantification of the different delivery robot risk factors identified in this research is tailored to the specific robot central in this research. For that reason, four risk factors are shown with an associated scale that accounts for the delivery robot of focus in this study. These risk factors are: the 'slope of the sidewalk surface', the 'sidewalk width', the 'temperature' and the 'SADR operating speed'. Because the ODD of a specific delivery robot was mapped in this study, the associated scale is justified, but it does not contribute to the fact that this method can be directly copied to other types of delivery robots because they may have different characteristics and operating limits. Minor adjustments must be made.

8.1.2 Dynamic assessment framework to practice

A current limitation of the proposed dynamic assessment framework is that it assumes that there are indeed different levels of difficulty within the ODD and that this difficulty can be objectively determined. The experts interviewed for the validation of this research agreed that the theory of existence of different ODD levels is correct, eight out of nine experts agreed and one stated to have no opinion. However, most of the experts also stressed the importance of empirical data to support this theory and stated that it is not possible to actually develop this assessment framework on a theory until it has been proven. According to some of the interviewees, what is proposed is correct and feasible from a theoretical perspective. Especially since the steps that are followed are logically connected and reproducible. However, from a practical perspective, there are still many barriers to be overcome before the proposed dynamic assessment framework can be used to guide delivery robot tests on public sidewalks. As discussed in Section 6.5.10 on the data required for a fully functioning dynamic assessment framework, once empirical relationships between risk factors and delivery robot performance have been found, an important consideration is how to digitise these risk factors to be able to use them in the dynamic assessment framework. If the impact of monitored risk factors on ODD difficulty changes over time, this does not pose a problem because the framework can be altered, but when the impact of new, not yet network-wide monitored risk factors on ODD difficulty become statistically significant over time, these risk factors should also be digitized. How best to approach this needs further research.

A strength of the dynamic assessment framework is its associated living lab approach. This recognises that the traffic system is dynamic, but also that system requirements and user preferences can change over time. The current knowledge on autonomous innovations is too little to develop a rigid assessment system for autonomous innovations. The living lab approach contributes to the assessment framework and the deployment of delivery robots developing in a sustainable way and to ensuring that the assessment method does not become obsolete in the short term.

8.1.3 Expert validation

A drawback of the validation in this research is that the intended workshops, where experts would evaluate the contents of this research together, could not take place due to the availability of the experts. Instead, semi-structured interviews were conducted. Because of the semi-structured approach more qualitative data was gathered, however due to the unstructured nature of the conversations, not every interview had exactly the same degree of depth in each research area as it would have if all interviewees had attended the same workshop. Two interviewees preferred that only notes were taken during the interview and not an audio recording. Therefore the quality of the interview data relied more on the ability of the researcher to take good notes and process them directly after the interview took place. In addition it must be noted that approximately one hour per interview is limited given the new and layered nature of this research, especially when two experts joined the same interview. One interviewee has been interviewed for two hours. The experts interviewed for this validation process are well balanced in terms of academics and practitioners and are authoritative in their field. The fact that these experts all have different backgrounds and a different field of expertise, but were mostly in agreement on the research process and findings indicates that the research is complete, clear and theoretically well substantiated to these experts.

8.2 Conclusion

The design-science contributions of this research are: (i) the development of the theory of different levels of difficulty within the Operational Design Domain, (ii) an approach to map and measure the factors that cause Operational Design Domain difficulty for delivery robots, and (iii) the theoretical development of a dynamic assessment framework based on which delivery robots can be assessed on their safe and sustainable performance on public sidewalks in the Netherlands.

This research has focused on the design of a dynamic assessment framework to assess delivery robots on safe performance on the public sidewalks in the Netherlands. This framework design is based on the theory that there exist different degrees of difficulty within the Operational Design Domain, caused by the presence of specific characteristics of the environment. The development of this theory of relative ODD difficulty is the first conclusion of this research. The associated level of difficulty for a specific geographic location can vary over time due to the dynamic nature of the public environment. A dynamic operating domain that changes in difficulty is much better suited to the development and assessment of delivery robots than the traditional binary approach to the concept of the Operational Design Domain and the rigid safety testing procedures of vehicles. By using different levels of difficulty within the ODD, and having insight into what actually causes this relative difficulty objectively, it is possible to steer delivery robot development in a much more structured way towards the critical parts of a delivery robot that actually need further development, while at the same time, by applying this nuance within the ODD, knowledge can be gained about the performance of delivery robots in the public domain in a structured and risk-minimising way. The proposed assessment framework is also innovative in the way it not only investigates whether delivery robots perform technically safe, but also works towards the most socially acceptable integration of delivery robots into the existing traffic system.

The proposed dynamic assessment framework can be of value because it provides guidance to integrating delivery robots into the Dutch traffic system and society over time. This integration is important because delivery robots are a means to mitigate the negative externalities associated with last-mile delivery. According to experts interviewed the theory of different levels of difficulty within the Operational Design Domain is promising and worthy of further elaboration. The actual development of the designed digital dynamic assessment framework can contribute to tests with delivery robots on public sidewalks in the Netherlands. As emerges from the expert validations as well, empirical evidence is the missing link between the theoretically established ODD difficulty levels and the envisaged dynamic assessment framework. Development of the dynamic assessment framework should be driven by factors that actually impact the performance of delivery robots. This and related follow-up questions arising from this research are discussed in detail in the recommendations section in Section 9.

9 Recommendations

In this section the implementation recommendations for practice and scientific recommendations for further research will be discussed.

9.1 Implementation recommendations

The first practical recommendation is that it is opportune to record the exact location characteristics and the performance of delivery robots of current pilot projects for all their movements from now on. In this way, a start is made towards a system where the performance of delivery robots is evaluated on the basis of their objective context. From a system transparency perspective it would be good to publish publicly accessible reports regarding these circumstances and performance (Dawes, 2010; Hollyer, Rosendorff, & Vreeland, 2014), however it must be noted that current strategic interests and competitive advantage are at odds with this.

The most important recommendation from this research is to verify whether there is indeed an empirical relationship between the location characteristics identified in this study and the performance of delivery robots. These empirical relationship(s) are not necessarily linear and self-contained. Thorough research needs to be done on the best models to explain ODD difficulty. Based on the insights from these studies, the performance of the delivery robots and the quality of the digital simulation environment for training delivery robots can be improved in an iterative way. The factors that turn out to be statistically significant in relation to delivery robot performance can be further investigated in such simulation environments without posing risks to other traffic participants. When enough objective and independently verified data has been gathered to classify geographical locations on static elements for good infrastructure for delivery robots, all locations can be listed that offer a good case for a living lab. The goal and further implementation of this living lab can then be determined with all stakeholders as follows:

All stakeholders in relation to the deployment of delivery robots on public sidewalks could be involved in the living lab from the start. The parties that might be included are at least: the robot operator, the road authority, party/parties that want to use robot delivery, local residents and researchers. Together, these stakeholders should further shape the goal of integrating delivery robots into the current traffic system in a safe and sustainable manner, and decide on the initial performance indicator thresholds used to evaluate the performance of delivery robots. A valuable robot operator is one who is able and willing to provide all the robot performance data needed to determine the performance indicators in Section 6.5.8 and is also willing to work with the living lab for a longer period of time, to develop the delivery robot performance into a technically safe and sustainable logistics solution. This robot performance data from the operator does not necessarily have to be shared publicly, but it must be able to be used to accurately determine ODD difficulty. Long-term commitment is not only needed from the robot operator, but from all stakeholders involved in order to work together towards the goal of the living lab on a continuous basis, opposed to short term small scale pilot projects. The representation of residents should be a good reflection of the residents in the operating area to secure social support for delivery robots. Regarding the included researchers, it is important that they are from different research disciplines so that they can ensure that all research questions that will be investigated in the living lab are also investigated in a scientific way. In addition, this stakeholder group can ensure that decisions within the living lab are made on the basis of objectively gathered knowledge. It should not be overlooked that a prerequisite to operate on the public sidewalk is to first test in an enclosed area, where it is verified that the robot is technically capable of doing what it is supposed to do in the first place. Lastly, it is important that all stakeholders agree on the living lab course as depicted with the process map in Figure 36, such that it is clear for everyone on what basis changes to the allowed operating area, performance indicators and delivery robots are made.

A recommendation regarding the strategic selection of a living lab location is that the possible emergence of delivery robots, and other innovative vehicles, can be included in urban planning. If, from the infrastructure side, it is known which factors shape a passable infrastructure for delivery robots, safe locations can also be designed that way. For example, a wider total space for the sidewalk and the bicycle path can be considered. In this way, the policy choice for allocation to the sidewalk, cycle path or separate carriageway in between can be made in time. In the long term, if no delivery robots or other innovative vehicles are allowed to the existing traffic system, the infrastructure for vulnerable road users is only very generous in some locations. The interesting advantage of an area that is still to be developed is that if delivery robots are already on the road when this urban area is completed, there will be no problems with road users who are already used to the area without delivery robots.

A final practical recommendation is that the innovation of a delivery robot could be separated from the delivery function in view of the dynamic assessment framework. In this way, as much knowledge as possible can be gained about the driving performance of self-driving robots before they are actually used for delivery. The locations where it would be safest to drive, for example, a sparsely populated area with many low-rise buildings, is not, from a business model point of view, the location that will be most profitable to carry out deliveries. There are various business models that can be implemented with exactly the same size robot, such as a tool for reversed logistics, small (electronic) waste collection, marketing, scanning number plates for paid parking fees, sidewalk mapping for pedestrians etc. The dynamic assessment should be focused on the assessment of the autonomous driving systems related to their operating environment.

9.2 Scientific recommendations

As noted during this research, the literature on delivery robots is almost non-existent. In this section recommendations for further research will be discussed that follow directly from the research conducted. It must be noted that the implementation recommendations for practice should also be conducted in a scientific way, and can certainly be seen as scientific recommendations as well. From the research conducted numerous follow-up research possibilities emerge, both qualitative and quantitative.

A first recommended research direction focuses on the possibilities for standardisation, which was mainly found to be lacking in this study. Standardisation is missing in the description of the Operational Design Domain, in the way objects and elements in the Operational Design Domain should be measured, in the way incident reporting is done and in the way signals and objects are digitised. Regarding the dynamic assessment framework there are also no standards to assess autonomous driving systems on safety. When the empirical relation between ODD factors and delivery robot performance is confirmed, and testing on public sidewalks is allowed based on the proposed proof of concept methodology, a subsequent study could focus on the usefulness of different performance indicators for delivery robots.

A second line of research focuses on the research that has to be conducted regarding delivery robots and the human interaction. Finding out about user acceptance is essential to make the deployment of delivery robots in public space a success. Research can focus on the appearance of delivery robots to increase their acceptance. There are studies in this direction (e.g. de Groot (2019)), but this has not been widely studied for delivery robots. The other side of this research direction can focus on: what is the most desirable behaviour of a robot for other road users? The results of this research can on the one hand be used by a delivery robot manufacturer or operator to increase the acceptance by adjusting the behaviour, on the other hand the results can provide policy makers with tools to define initial traffic rules and guidelines for delivery robots (and similar innovations). By drawing up these rules according to the experiences of road users, acceptance of delivery robots is increased. It would be beneficial to investigate this by means of revealed preference research in a real life context, for example in the proposed living lab(s). Stated preference research poses the issue of hypothetical bias (Van Cranenburgh, Chorus, & van Wee, 2014).

A third line of research could focus on the theory of different difficulty levels within the Operational Design Domain. First of all, it would be scientifically useful if this theory is also confirmed by other studies and researchers. Currently, this theory has only been examined for the ODD of the sidewalk and the innovation of a delivery robot. It is scientifically relevant whether this theory also applies within other operating domains and for other innovations. The follow-up question could be whether the notion of different ODD difficulty degrees is also a useful insight for the performance assessment of other vehicles, or whether this is only true for delivery robots because of their unique characteristics.

A fourth line of research opens up after the empirical relationships are confirmed. It is then possible to look at which elements actually impact the performance of delivery robots and how these elements can be digitally captured as efficiently as possible to actually realise the digital dynamic assessment framework proposed in this research. Additionally, the research could focus on which Layer 6 solutions at which geographical locations should be used to improve the performance of delivery robots as efficiently as possible.

Food for thought on responsible engineering

The contents of this section are based on a conversation about ethical aspects of this research with Dr. Filippo Santoni de Sio, who is associate professor in Ethics of Technology at TU Delft, and an adjunct professor in Ethics of Transportation at the Politecnico di Milano. From June 2019 to 2020 he has been Rapporteur of the EU Commission Expert Group to advise on ethical issues raised by driverless mobility (E03659) and from 2017 to 2020 he was co-director of the NWO-funded interdisciplinary project Meaningful Human Control over Automated Driving Systems (TU Delft, n.d.).

With this section, ethical matters regarding the deployment of delivery robots and the development of the dynamic assessment framework not explicitly touched upon during the main research can be explored. Engineers developing certain technologies should take into account moral values and ethical principles. Note that this section is purely meant to give readers a broader perspective on this research topic and provide them with 'food for thought'. The fact that something is possible from a technological perspective does not necessarily mean that it has to be realised. The topics that will be briefly covered in this section are: sustainable integration, safety, and data and ownership.

Sustainable integration

By applying the proposed assessment framework in a living lab context operational aspects that are not necessarily considered when admitting new vehicles to public roads or sidewalks can be explored in a structured way. Consider the recent international developments in shared mobility. Shared mopeds, shared bikes and shared scooters cause operational issues that were not taken into account when these innovations were admitted to public space. Residents in cities are complaining about the nuisance caused by such mobility services, and regulators in some countries or cities have already banned the services because of the problems and unsafety caused (Buckley, 2019). Value sensitive design and participatory design are means to embed the human values of relevant stakeholders to include the sustainable integration of those innovations with society. In daily life people are not constantly reasoning on safety based on scientific models, so other perspectives have to be taken into account. Both value sensitive design and participatory design are in line with the Living Lab approach which is advocated for in this research.

A technologically safe and socially accepted implementation of delivery robots in the existing traffic system is a long and intensive process. From a philosophical perspective it is of interest to focus on the added value of delivery robots, before this effort is made. Questions that have to be evaluated are: 'what is the problem we are solving with delivery robots?', 'why is solving that problem valuable?', 'under which conditions are delivery robots as a solution for this problem a good idea'? Care should be taken not to invent a problem for a solution that we have. It should also be taken into account that there could be ways to solve the identified problem with a different technical or non-technical solution. Answering the aforementioned questions contributes to the development of a delivery robot system that actually adds value to society. In addition, answering the questions could help in the formulation of a clear goal for the intended living lab proposed in this research.

Safety

As identified in this research, there are no existing systems or vehicles to which the safe performance of delivery robots can be compared. That delivery robots, given the function associated with them, could replace delivery vans in the long term is no reason to take the safe performance of delivery vans as a benchmark for delivery robots. And this benchmark should not be sought either. The existing traffic system, where accidents occur daily, is not easily solved. But that does not necessarily mean that every innovation that is slightly safer than what currently exists should be allowed to the public space just like that. Precisely because no benchmarks and expectations exist yet, there is an opportunity to set up a well-functioning and sustainable delivery robot system from scratch. The existing barriers identified in this research can be seen as restrictive and unnecessary, but can also be seen as an incentive to push delivery robots to the highest possible performance level.

The construct of safety is hard to define, because what is the definition of safety? And how should we measure safety? Is safety for example 'not being killed by a delivery robot', or is safety 'not feeling threatened by a delivery robot'? Additionally it may be that elements that are not really about safety are transferred into general concerns about safety because people have a disattitude towards new technology such as delivery robots. For example, people that do not like the idea of sidewalks crowded with delivery robots could state that they do not want delivery robots on sidewalks because they are not safe. Also the slippery slope argument is applicable in this context, because it could be that people fear that the deployment of delivery robots is just the beginning of more public space being given away to such innovations and of potential mass surveillance. Safety is not just a technical concept that can be measured with strict performance indicators. There may be differences in the views of different groups in society and between different societies as to what is socially desirable and perceived as safe. This relates well to how the development of the dynamic assessment framework with stakeholders is envisaged in this research, where the representation of residents and traffic participants should be a good reflection of society.

Data and ownership

The dynamic assessment framework proposed in this research combines different data sources to define different Operational Design Domain difficulty levels. In addition, delivery robots themselves generate a lot of data about its performance and its surroundings. There should be a debate on who owns the data that is stored in the intended database system, who has the right to exploit that data, and what can and cannot be done with that data? These questions also have to be answered for other data driven technologies that have been or are being developed, however the new thing about delivery robots is that they drive into public spaces and collect information there, where people are less able to guard their privacy than when people choose to use online data-driven technologies. On the one hand you would like to have the complete dataset of location characteristics and delivery robot performance publicly available to learn from and prevent accidents from happening, but on the other hand you do not want the same mass of private data to be publicly available to companies. One solution posed is a sort of auditing system, where companies would be constrained not to collect certain data and would also be forced to share the collected data with some specific agency under specific circumstances, like the government. This would prevent the objection of the companies that this data is their competitive advantage, but the data should at least be confidentially shared with some agency to be audited. The follow up point of attention is the trustworthiness of this agency. Another solution is to have an agency that is responsible to convert raw data into usable data to draw conclusions without sharing privacy-sensitive information with commercial parties.

It should be carefully considered how the ethical aspects discussed are going to be managed responsibly in the public space, because:

"Trust in technology is fragile: slowly to build and easy to destroy" ∼ Robotics Vision Team TU Delft (2021)

References

ANWB. (n.d.-a). *Aansprakelijkheid wegbeheerder.* Retrieved March 11, 2022, from https://www.anwb.nl/juridisch-advies/aanrijding-en-dan/aansprakelijkheid/aansprakelijkheid-wegbeheerder

ANWB. (n.d.-b). *Wat is een APK keuring?* Retrieved March 11, 2022, from https://www.anwb.nl/auto/onderhoud-en-reparatie/apk/wat-is-een-apk-keuring

Bagschik, G., Menzel, T., & Maurer, M. (2018). Ontology based scene creation for the development of automated vehicles. In *2018 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1813–1820). https://doi.org/10.1109/IVS.2018.8500632.

Bellairs, R. (2019, January 3). *What Is ISO 26262? Overview and ASIL.* Retrieved April 7, 2022, from https://www.perforce.com/blog/qac/what-is-iso-26262

Bergvall-Kareborn, B., Hoist, M., & Stahlbrost, A. (2009). Concept design with a living lab approach. In *2009 42nd Hawaii International Conference on System Sciences* (pp. 1–10). https://doi.org/10.1109/HICSS.2009.123.

Bock, J., Krajewski, R., Eckstein, L., Klimke, J., Sauerbier, J., & Zlocki, A. (2018). Data basis for scenario-based validation of had on highways. In *27th Aachen colloquium automobile and engine technology* (pp. 8–10). https://doi.org/10.1080/15389588.2019.1630827.

Boggs, A. M., Arvin, R., & Khattak, A. J. (2020). Exploring the who, what, when, where, and why of automated vehicle disengagements. *Accident Analysis & Prevention*, *136*, 105406. https://doi.org/10.1016/j.aap.2019.105406.

Bosch. (n.d.). *Long-range lidar.* Retrieved May 27, 2022, from https://www.bosch-mobility-solutions.com/en/solutions/sensors/long-range-lidar/

Boysen, N., Schwerdfeger, S., & Weidinger, F. (2018). Scheduling last-mile deliveries with truck-based autonomous robots. *European Journal of Operational Research*, *271*(3), 1085–1099. https://doi.org/10.1016/j.ejor.2018.05.058.

Bright. (2021, January 4). *Zelfrijdende auto's mogen nu de weg op: waar blijft tesla?!* [Video]. YouTube. Retrieved March 11, 2022, from https://www.youtube.com/watch?v=bxb3TUr2QOs

Buckley, J. (2019, November 22). *E-scooters suddenly appeared everywhere, but now they're riding into serious trouble.* Retrieved July 17, 2022, from https://edition.cnn.com/travel/article/electric-scooter-bans-world/index.htmle

Buienalarm. (n.d.). *Buienalarm.* Retrieved May 26, 2022, from https://www.buienalarm.nl/

Cambride University Press. (n.d.). Manoeuvrability. In *Cambridge Dictionary.* Retrieved May 30, 2022, from https://dictionary.cambridge.org/dictionary/english/manoeuvrability

Camus, J.-L. (2019, January 21). *Unifying control and verification of cyber-physical systems.* Retrieved April 7, 2022, from https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c3b28d28&appId=PPGMS

Canada Transport. (2021, August 6). *Guidelines for testing automated driving systems in canada.* Retrieved May 31, 2022, from https://tc.canada.ca/en/road-transportation/innovative-technologies/connected-automated-vehicles/guidelines-testing-automated-driving-systems-canada

Cartens, R. (2020, September 29). *Proef met bezorgrobot op hogeschool in Breda: 'Dit is de toekomst'.* Retrieved March 1, 2022, from https://www.starship.xyz/b2b/

CBR. (n.d.). *Je autorijbewijs halen stappenplan.* Retrieved March 11, 2022, from https://www.cbr.nl/nl/rijbewijs-halen/auto/je-autorijbewijs-halen/stappenplan.htm

CFI. (n.d.). *Vehicle to Everything (V2X).* Retrieved March 23, 2022, from https://corporatefinanceinstitute.com/resources/knowledge/other/vehicle-to-everything-v2x/

Cho, H. (2020). *Operational Design Domain (ODD) framework for driver-automation integrated systems.* [Thesis]. Massachusetts Institute of Technology. Retrieved from https://hdl.handle.net/1721.1/129156

CityTraffic. (n.d.). *How do we measure.* Retrieved May 27, 2022, from https://www.citytraffic.nl/en/how-do-we-measure/

Coppola, N. A., & Marshall, W. E. (2021). Sidewalk static obstructions and their impact on clear width. *Transportation research record*, *2675*(6), 200–212. https://doi.org/10.1177%2F0361198121991833.

CROW. (n.d.-a). *About CROW.* Retrieved April 5, 2022, from https://www.crow.nl/english-summary

CROW. (n.d.-b). *De zelfrijdende auto.* Retrieved April 5, 2022, from https://www.crow.nl/testenzelfrijdendeauto/zelfrijdende-auto-s

Czarnecki, K. (2018a). Operational Design Domain for Automated Driving Systems. *Taxonomy of Basic Terms, Waterloo Intelligent Systems Engineering (WISE) Lab, University of Waterloo*.

Czarnecki, K. (2018b). Operational world model ontology for automated driving systems–part 1: Road structure. *Waterloo Intelligent Systems Engineering Lab (WISE) Report, University of Waterloo*.

Czarnecki, K. (2018c). Operational world model ontology for automated driving systems–part 2: Road users, animals, other obstacles, and environmental conditions,". *Waterloo Intelligent Systems Engineering Lab (WISE) Report, University of Waterloo*.

Dawes, S. S. (2010). Stewardship and usefulness: Policy principles for information-based transparency. *Government Information Quarterly*, *27*(4), 377–383. https://doi.org/10.1016/j.giq.2010.07.001.

de Groot, S. (2019, August). *Pedestrian acceptance of delivery robots.* [MSc Thesis]. Delft University of Technology. Retrieved February 24, 2022, from http://resolver.tudelft.nl/uuid:f9e8c003-c8fc-4075-bff3-0d54e0f0fecb

Dutch Safety Board. (2019, October). *Veilig toelaten op de weg - lessen naar aanleiding van het ongeval met de stint.* https://www.onderzoeksraad.nl/nl/page/12408/veilig-toelaten-op-de-weg---lessen-naar-aanleiding-van-het-ongeval.

Erasmus University Rotterdam. (2021, December 1). *Primeur: robot Rosie bezorgt boodschappen op campus Woudestein.* Retrieved February 24, 2022, from https://www.eur.nl/nieuws/primeur-robot-rosie-bezorgt-boodschappen-op-campus-woudestein

European Commission. (n.d.). *Decision making process.* Retrieved March 11, 2022, from https://ec.europa.eu/info/strategy/decision-making-process_en

Farah, H., Erkens, S. M., Alkim, T., & Arem, B. v. (2018). Infrastructure for automated and connected driving: State of the art and future research directions. *Road vehicle automation 4*, 187–197. https://doi.org/10.1007/978-3-319-60934-8_16.

Feng, J., Yu, S., Chen, G., Gong, W., Li, Q., Wang, J., & Zhan, H. (2020). Disengagement causes analysis of automated driving system. In *2020 3rd World Conference on Mechanical Engineering and Intelligent Manufacturing (WCMEIM)* (pp. 36–39). https://doi.org/10.1109/WCMEIM52463.2020.00014.

Geyer, S., Baltzer, M., Franz, B., Hakuli, S., Kauer, M., Kienle, M., ... Bruder, R. (2014). Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance. *IET Intelligent Transport Systems*, *8*(3), 183–189. https://doi.org/10.1049/iet-its.2012.0188.

Ghaffarzadeh, K. (2019, December 23). *Sidewalk last mile delivery robots: A billion-dollar-market by 2030?* IDTechEx. Retrieved March 2, 2022, from https://www.idtechex.com/de/research-article/sidewalk-last-mile-delivery-robots-a-billion-dollar-market-by-2030/19278

Google. (n.d.-a). *Get information about busy areas from Google Maps.* Retrieved May 27, 2022, from https://support.google.com/maps/answer/11323117?hl=en

Google. (n.d.-b). *Popular times, wait times, and visit duration.* Retrieved May 27, 2022, from https://support.google.com/business/answer/6263531?hl=en

Google. (n.d.-c). *Provide useful information to your users when they are near an area of interest.* Retrieved June 4, 2022, from https://developers.google.com/location-context/geofencing

Griffor, E., Wollman, D., & Greer, C. (2021). Automated Driving System Safety Measurement Part I: Operating Envelope Specification. *NIST Special Publication*, *1900*, 301. https://doi.org/10.6028/NIST.SP.1900-301.

Grush, B. (n.d.). *Getting ready for sidewalk robots - what's a smart city to do? (2021-2025).* Retrieved April 6, 2022, from https://citm.ca/wp-content/uploads/2021/02/Harmonize-Mobility_Getting-Ready-for-Sidewalk-Robots_2021.02.20.pdf

Hagenzieker, M., Boersma, R., Velasco, J., Ozturker, M., Zubin, I., & Heikoop, D. (2020). Automated buses in europe. *An Inventory of Pilots ", Version 0.5. TU Delft*.

Hagenzieker, M. P., Van Der Kint, S., Vissers, L., van Schagen, I. N. G., De Bruin, J., Van Gent, P., & Commandeur, J. J. (2020). Interactions between cyclists and automated vehicles: Results of a photo experiment. *Journal of Transportation Safety & Security*, *12*(1), 94–115. https://doi.org/10.1080/19439962.2019.1591556.

Harmonize Mobility. (n.d.-a). *Robots on city streets and walkways.* Retrieved April 6, 2022, from https://harmonizemobility.com/sidewalkandcurb/

Harmonize Mobility. (n.d.-b). *White paper - the last block - towards an international standard to regulate & manage sidewalk robots.* Retrieved April 6, 2022, from https://citm.ca/wp-content/uploads/2021/02/Harmonize-Mobility_The-Last-Block_21.02.01.pdf

Heikoop, D. D., Velasco, J. P. N., Boersma, R., Bjørnskau, T., & Hagenzieker, M. P. (2020). Automated bus systems in europe: A systematic review of passenger experience and road user interaction. *Advances in Transport Policy and Planning*, *5*, 51–71.

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems*, *19*(2), 4.

Hillman, R., & Capaldi, R. (2020). *Test methods for interrogating autonomous vehicle behaviour.* [Video]. HORIBA MIRA. Retrieved March 14, 2022, from https://www.horiba-mira.com/webinars/

Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. (2014). Measuring transparency. *Political analysis*, *22*(4), 413–434. https://doi.org/10.1093/pan/mpu001.

Houghton Mifflin Harcourt. (2006). Topography. In *American Heritage Dictionary.* (4th ed.)

Humblet, M. (2021, November 17). *What is proof of concept & how to actually use it.* [Video]. YouTube. Retrieved June 1, 2022, from https://www.youtube.com/watch?v=3BRrxjHGy6U

INFRAMIX. (n.d.). *Infrastructure Categorization: ISAD levels.* Retrieved April 16, 2022, from https://www.inframix.eu/infrastructure-categorization/

ISO. (2018). *ISO 26262-1:2018 Road vehicles — Functional safety — Part 1: Vocabulary).* Retrieved from https://www.iso.org/standard/68383.html

ISO. (2019). *ISO/PAS 21448:2019 Road vehicles — Safety of the intended functionality.* Retrieved from https://www.iso.org/standard/70939.html

ISO. (2021a). *ISO 22737:2021 Intelligent transport systems — Low-speed automated driving (LSAD) systems for predefined routes — Performance requirements, system requirements and performance test procedures.* Retrieved from https://www.iso.org/standard/73767.html

ISO. (2021b). *ISO/SAE 21434:2021 Road vehicles — Cybersecurity engineering.* Retrieved from https://www.iso.org/standard/70918.html

ITE. (n.d.). *Geometric design.* Retrieved March 23, 2022, from https://www.ite.org/technical-resources/topics/geometric-design/

Jennings, D., & Figliozzi, M. (2019). Study of sidewalk autonomous delivery robots and their potential impacts on freight efficiency and travel. *Transportation Research Record*, *2673*(6), 317–326. https://doi.org/10.1177%2F0361198119849398.

Johannesson, P., & Perjons, E. (2021). A method framework for design science research. In *An introduction to design science* (pp. 77–93). Springer. https://doi.org/10.1007/978-3-030-78132-3_4.

Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the digital twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, *29*, 36–52. https://doi.org/10.1016/j.cirpj.2020.02.002.

Khatun, M., Glaß, M., & Jung, R. (2021). A systematic approach of reduced scenario-based safety analysis for highly automated driving function. In *Proceedings of the 7th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2021)* (pp. 301–308). http://dx.doi.org/10.5220/0010397403010308.

KNMI. (2022). *Daggegevens van het weer in Nederland.* Retrieved May 25, 2022, from https://www.knmi.nl/nederland-nu/klimatologie/daggegevens

Kocsis, M., Zöllner, R., & Mogan, G. (2022). Interactive system for package delivery in pedestrian areas using a self-developed fleet of autonomous vehicles. *Electronics*, *11*(5), 748. https://doi.org/10.3390/electronics11050748.

Koopman, P., & Fratrik, F. (2019). How many operational design domains, objects, and events? In *Safeai@aaai.*

KPMG. (2018, January). *Autonomous vehicles readiness index: Assessing countries' openness and preparedness for autonomous vehicles.* KPMG International Cooperative, Report No. 135006-G.

Kroesen, M. (2021). *SEN1721 Econometric modelling paradigm.* [PowerPoint slides]. Delft University of Technology. Retrieved June 4, 2022, from https://brightspace.tudelft.nl/d2l/le/content/401532/viewContent/2241472/View

Kuutti, S., Fallah, S., Katsaros, K., Dianati, M., Mccullough, F., & Mouzakitis, A. (2018). A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications. *IEEE Internet of Things Journal*, *5*(2), 829–846. https://doi.org/10.1109/JIOT.2018.2812300.

Lee, J. M. (2016). A design of road database for self-driving vehicles. *Indian Journal of Science and Technology*, *9*, 19. http://dx.doi.org/10.17485/ijst/2016/v9i20/94704.

Leroy, J., Gruyer, D., Orfila, O., & El Faouzi, N.-E. (2020). Five key components based risk indicators ontology for the modelling and identification of critical interaction between human driven and automated vehicles. *IFAC-PapersOnLine*, *53*(5), 212–217. https://doi.org/10.1016/j.ifacol.2021.04.141.

Mahmood, A., Zhang, W. E., & Sheng, Q. Z. (2019). Software-defined heterogeneous vehicular networking: The architectural design and open challenges. *Future Internet*, *11*(3), 70. https://doi.org/10.3390/fi11030070.

Manjunath, A., Liu, Y., Henriques, B., & Engstle, A. (2018). Radar based object detection and tracking for autonomous driving. In *2018 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)* (pp. 1–4). https://doi.org/10.1109/ICMIM.2018.8443497.

Martin, H., Winkler, B., Grubmüller, S., & Watzenig, D. (2019). Identification of performance limitations of sensing technologies for automated driving. In *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE)* (pp. 1–6). https://doi.org/10.1109/ICCVE45908.2019.8965181.

Merwe, A. v. d., Gerber, A., & Smuts, H. (2017). Mapping a design science research cycle to the postgraduate research report. In *Annual Conference of the Southern African Computer Lecturers' Association* (pp. 293–308). https://doi.org/10.1007/978-3-319-69670-6_21.

Mihalj, T., Li, H., Babić, D., Lex, C., Jeudy, M., Zovak, G., ... Eichberger, A. (2022). Road infrastructure challenges faced by automated driving: A review. *Applied Sciences*, *12*(7), 3477. https://doi.org/10.3390/app12073477.

Molin, J. (2019, October 15). *Multiple regression analysis - an introduction to multivariate analysis.* [PowerPoint slides]. Delft University of Technology. Retrieved May 23, 2022, from https://brightspace.tudelft.nl/d2l/le/content/401532/viewContent/2241495/View

Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). Conducting semi-structured interviews. *Handbook of practical program evaluation*, *492*, 492. https://doi.org/10.1002/9781119171386.ch19.

NHTSA. (2021, June 29). *NHTSA Orders Crash Reporting for Vehicles Equipped with Advanced Driver Assistance Systems and Automated Driving Systems.* Retrieved May 31, 2022, from https://www.nhtsa.gov/press-releases/nhtsa-orders-crash-reporting-vehicles-equipped-advanced-driver-assistance-systems

Petegem, J. v., Nes, C. v., Boele, M., & Eenink, R. (2018, April). *Advies praktijkproef - Starship bezorgrobot.* Instituut voor Wetenschappelijk Onderzoek Verkeersveiligheid (SWOV). https://swov.nl/nl/publicatie/advies-praktijkproef-4.

Pokorny, P., Skender, B., Bjørnskau, T., & Hagenzieker, M. P. (2021). Video observation of encounters between the automated shuttles and other traffic participants along an approach to right-hand priority t-intersection. *European Transport Research Review*, *13*(1), 1–13. https://doi.org/10.1186/s12544-021-00518-x.

Politie. (2022). *Personeelstekort politie nog steeds hoog, neemt af na 2022.* Retrieved April 2, 2022, from https://www.politie.nl/nieuws/2022/januari/13/personeelstekort-politie-nog-steeds-hoog-neemt-af-na-2022.html

Ponn, T., Gnandt, C., & Diermeyer, F. (2019). An optimization-based method to identify relevant scenarios for type approval of automated vehicles. In *Proceedings of the ESV—International Technical Conference on the Enhanced Safety of Vehicles, Eindhoven, The Netherlands* (pp. 10–13).

QGIS. (2022, March 20). *7. topology.* Retrieved March 23, 2022, from https://docs.qgis.org/3.22/en/docs/gentle_gis_introduction/topology.html

Quak, H., & Nesterova, N. (2021). *Living labs for transitions in urban freight transport systems.* (Final draft paper)

Rao, S. J., Deosthale, E., Barickman, F. S., Elsasser, D., & Schnelle, S. C. (2021). *An approach for the selection and description of elements used to define driving scenarios.* (Report No. DOT HS 813 073). United States. Department of Transportation. National Highway Traffic Safety Administration.

RDW. (n.d.-a). *Normverwijzingen accreditatie rdw testcentrum.* Retrieved March 4, 2022, from https://www.rdw.nl/zakelijk/branches/fabrikanten-en-importeurs/typegoedkeuring-aanvragen/testen/normverwijzingen-accreditatie-rdw-testcentrum

RDW. (n.d.-b). *Organisatie.* Retrieved March 11, 2022, from https://www.rdw.nl/over-rdw/organisatie

RDW. (n.d.-c). *Wetgeving per voertuigcategorie.* Retrieved March 4, 2022, from https://www.rdw.nl/zakelijk/branches/fabrikant/wetgeving-en-geldigheid-typegoedkeuring/wetgeving-per-voertuigcategorie

RDW. (n.d.-d). *ZZ-Kenteken.* Retrieved April 2, 2022, from https://www.rdw.nl/particulier/voertuigen/auto/het-kentekenbewijs/bijzondere-kentekenbewijzen/zz-kenteken

RDW. (2019, July 2). *Experimenteerwet voor zelfrijdende voertuigen van kracht.* Retrieved February 24, 2022, from https://www.rdw.nl/particulier/nieuws/2019/experimenteerwet-voor-zelfrijdende-voertuigen-van-kracht

RDW. (2020, December 2). *RDW - Goedkeuringsplicht.* [Video]. YouTube. Retrieved March 4, 2022, from https://www.youtube.com/watch?v=6Ta_R7MYDcs

Regeling vergunningverlening experimenten zelfrijdende auto, nr. IENW/BSK-2019/134685. (2019, June 20). https://wetten.overheid.nl/jci1.3:c:BWBR0042343&z=2019-07-01&g=2019-07-01.

Ren, L., Yin, H., Ge, W., & Meng, Q. (2019). Environment influences on uncertainty of object detection for automated driving systems. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 1–5). https://doi.org/10.1109/CISP-BMEI48845.2019.8965948.

Riedmaier, S., Ponn, T., Ludwig, D., Schick, B., & Diermeyer, F. (2020). Survey on scenario-based safety assessment of automated vehicles. *IEEE Access*, *8*, 87456–87477. https://doi.org/10.1109/ACCESS.2020.2993730.

Rijksoverheid. (n.d.). *Bijzondere bromfietsen.* Retrieved February 17, 2022, from https://www.rijksoverheid.nl/onderwerpen/bijzondere-voertuigen/bijzondere-bromfietsen

Rijksoverheid. (n.d.). *Wanneer wordt mijn rijbewijs ingevorderd?* Retrieved March 11, 2022, from https://www.rijksoverheid.nl/onderwerpen/rijbewijs/vraag-en-antwoord/wanneer-wordt-mijn-rijbewijs-ingevorderd

Robotics Vision Team TU Delft. (2021). *Vision Team Robotics.* Retrieved July 17, 2022, from https://d2k0ddhflgrk1i.cloudfront.net/TUDelft/Over_TU_Delft/Strategie/booklet-robotics-final-losbladig.pdf

Roh, C.-G., & Im, I.-J. (2020). A review on handicap sections and situations to improve driving safety of automated vehicles. *Sustainability*, *12*(14), 5509. https://doi.org/10.3390/su12145509.

SAE International. (n.d.). *SAE Aerospace Quality Standards.* Retrieved April 6, 2022, from https://www.sae.org/publications/collections/content/asquality/

SAE International. (2020a, April 15). *AVSC Best Practice for Describing an Operational Design Domain: Conceptual Framework and Lexicon.* Retrieved March 21, 2022, from https://www.sae.org/standards/content/avsc00002202004/

SAE International. (2020b). *Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles J3018_202012.* Retrieved May 4, 2022, from https://www.sae.org/standards/content/j3018_202012/

SAE International. (2021a, April 30). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles.* Retrieved March 7, 2022, from https://www.sae.org/standards/content/j3016_202104

SAE International. (2021b, July 15). *Taxonomy & Definitions for Operational Design Domain (ODD) for Driving Automation Systems J3259.* Retrieved March 14, 2022, from https://www.sae.org/standards/content/j3259/e

Saks, M., & Allsop, J. (2012). *Researching health: Qualitative, quantitative and mixed methods.* SAGE.

Scholtes, M., Westhofen, L., Turner, L. R., Lotto, K., Schuldes, M., Weber, H., . . . Körtke, F. (2021). 6-layer model for a structured description and categorization of urban traffic and environment. *IEEE Access*, *9*, 59131–59147. https://doi.org/10.1109/ACCESS.2021.3072739.

Schuldt, F. (2017). *Ein beitrag für den methodischen test von automatisierten fahrfunktionen mit hilfe von virtuellen umgebungen* (Doctoral dissertation). https://doi.org/10.24355/dbbs.084-201704241210.

Schwall, M., Daniel, T., Victor, T., Favaro, F., & Hohnhold, H. (2020). Waymo public road safety performance data. *arXiv preprint arXiv:2011.00038*.

Shetty, A., Tavafoghi, H., Kurzhanskiy, A., Poolla, K., & Varaiya, P. (2021). Risk assessment of autonomous vehicles across diverse driving contexts. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)* (pp. 712–719). https://doi.org/10.1109/ITSC48978.2021.9564744.

Sivak, M., & Schoettle, B. (2015). *Road safety with self-driving vehicles: General limitations and road sharing with conventional vehicles* (Tech. Rep.). University of Michigan, Ann Arbor, Transportation Research Institute. https://hdl.handle.net/2027.42/111735.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of business research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039.

Starship Technologies. (n.d.). *Autonomous robots for industry 4.0.* Retrieved March 1, 2022, from https://www.starship.xyz/b2b/

Starship Technologies. (2022, January 25). *Starship Technologies agrees €50m funding partnership from the European Investment Bank.* Retrieved March 1, 2022, from https://www.starship.xyz/press_releases/starship-technologies-agrees-e50m-funding-partnership-from-the-european-investment-bank/

Storsæter, A. D. (2021). *Designing and maintaining roads to facilitate automated driving.* [PhD thesis]. NTNU. Retrieved from https://hdl.handle.net/11250/2767106

SWOV. (n.d.). *About SWOV.* Retrieved March 11, 2022, from https://www.swov.nl/en/about-swov

Synopsys. (2022). *What is an autonomous car?* Retrieved April 4, 2022, from https://www.synopsys.com/automotive/what-is-autonomous-car.html

Tabone, W., De Winter, J., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., . . . Hancock, P. A. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. *Transportation research interdisciplinary perspectives*, *9*, 100293. https://doi.org/10.1016/j.trip.2020.100293.

Theeuwes, J., & Hagenzieker, M. P. (1993). Visual search of traffic scenes: On the effect of location expectations. *Vision in vehicles*, *4*, 149–158.

Thorn, E., Kimmel, S. C., Chaka, M., & Hamilton, B. A. (2018). *A framework for automated driving system testable cases and scenarios.* (DOT HS 812 623). United States. Department of Transportation. National Highway Traffic Safety Administration.

TNO. (n.d.). *Autonomous vehicles & systems.* Retrieved April 4, 2022, from https://www.tno.nl/en/focus-areas/artificial-intelligence/application-areas/autonomous-vehicles-and-systems/

TU Delft. (n.d.). *Dr. F. (Filippo) Santoni de Sio.* Retrieved June 25, 2022, from https://www.tudelft.nl/tbm/over-de-faculteit/afdelingen/values-technology-and-innovation/people/associate-professors/dr-f-filippo-santoni-de-sio

TÜV SÜD. (2018). *Automated driving requires international regulations.* Retrieved March 4, 2022, from https://www.tuvsud.com/en-in/-/media/global/pdf-files/whitepaper-report-e-books/tuvsud-whitepaper-had-regulation.pdf

Udacity. (2021, March 3). *How self-driving cars work: sensor systems.* Retrieved April 4, 2022, from https://www.udacity.com/blog/2021/03/how-self-driving-cars-work-sensor-systems.html

Vaishnavi, V., & Kuechler, W. (2004). *Design research in information systems.* Retrieved from http://www.desrist.org/design-research-in-information-systems/

van Binsbergen, A. (2020). *Introduction to TIL Design.* [PowerPoint slides]. Delft University of Technology. Retrieved June 11, 2022, from https://brightspace.tudelft.nl/d2l/le/content/321711/viewContent/1953413/View

Van Cranenburgh, S., Chorus, C., & van Wee, B. (2014). Vacation behaviour under high travel cost conditions–a stated preference of revealed preference approach. *Tourism Management*, *43*, 105–118. https://doi.org/10.1016/j.tourman.2014.01.022.

van der Stoep, P. (2022, February 24). *RDW Applied innovation.* [PowerPoint slides]. RDW.

Vehicle Certification Agency. (2022, February 25). *What is vehicle type approval?* Retrieved February 17, 2022, from https://www.vehicle-certification-agency.gov.uk/vehicle-type-approval/what-is-vehicle-type-approval/

Velzen, V. (2021, September 17). *Het proces over het dieselschandaal bij volkswagen is begonnen, wat kunnen we verwachten?* Retrieved March 17, 2022, from https://www.trouw.nl/economie/het-proces-over-het-dieselschandaal-bij-volkswagen-is-begonnen-wat-kunnen-we-verwachten~bb65501d

Vleeshouwer, T., Rotterdam, H., & Verbraeck, A. (2017). Implementatie van autonome bezorgrobots voor een kleinschalige thuisbezorgdienst. *Vervoerslogistieke Werkdagen 2017*.

Vleugel, J. (2021). *TIL5050-20 Course manual.* [PowerPoint slides]. Delft University of Technology. Retrieved March 12, 2022, from https://brightspace.tudelft.nl/d2l/le/content/401885/Home

Wishart, J., Como, S., Elli, M., Russo, B., Weast, J., Altekar, N., ... Chen, Y. (2020). Driving safety performance assessment metrics for ads-equipped vehicles. *SAE Technical Paper*, *2*(2020-01-1206). https://doi.org/10.4271/2020-01-1206.

Appendices

A Scientific paper

In this appendix the summary in the form of a scientific paper will be presented. For layout purposes this scientific paper has been moved to the next page.

A dynamic assessment framework for the safe performance of Sidewalk Autonomous Delivery Robots on public sidewalks

Focus area: the Netherlands

S.J. Beekes, Prof.dr.ir. L.A. Tavasszy, Dr. J.A. Annema and Dr.ir. A. van Binsbergen

CORRESPONDING AUTHOR: S.J. Beekes (e-mail: sebastiaan.beekes@gmail.com).

ABSTRACT Testing Sidewalk Autonomous Delivery Robot (SADR) performance in real world conditions is important to prove whether the innovation is ready for large-scale adoption. Currently, testing SADRs in public is not allowed in the Netherlands, because little is known about the actual risks associated with delivery robots, and the risks are currently difficult to assess because the robots are continuously evolving. Since the adoption of SADRs can reduce the negative effects of last mile logistics, in this article a theoretical dynamic assessment framework for the safe and sustainable performance of delivery robots is proposed. It is explored that there exist different degrees of relative difficulty within the Operational Design Domain of SADRs and a set of adapted metrics to objectively and completely quantify and qualify factors that cause this relative difficulty is drawn up. On this basis, the dynamic assessment framework has been developed. Both the theory of relative Operational Design Domain Difficulty and the proposed dynamic assessment framework have been validated with experts. Delivery robots can be safely tested in public, because based on the dynamic assessment framework and a living lab approach, associated risks can be constantly minimised.

INDEX TERMS Automated driving, dynamic Operational Design Domain, intelligent vehicles, living lab approach, performance assessment, proof of concept, real world testing, safety, Sidewalk Autonomous Delivery Robots.

I. INTRODUCTION

SIDEWALK Autonomous Delivery Robots (SADRs) are a promising innovation to mitigate the negative externalities associated with the last mile logistics problem (Vleeshouwer et al., 2017; Jennings & Figliozzi, 2019; Boysen et al., 2018). However their potential is known, so far, no SADRs can be tested for their performance in public on sidewalks in the Netherlands. In fact: little is known about the actual risks involved in the deployment of SADRs, there are not yet any good procedures to evaluate the safe performance of SADRs because no applicable vehicle category exists, and there is a lack of objective and complete data from other geographical areas that substantiates that SADRs can also be safely deployed on sidewalks in the Netherlands.

According to Stadler et al. (2022) there is a gap between the performance of Automated Driving Systems (ADSs) in a simulation versus the real world, because simulation models can only represent reality to a limited extent. Therefore, simulated performance cannot be assumed to be safe with certainty (Shetty et al., 2021). Because SADRs might ultimately be deployed on public sidewalks, their safe performance should also be evaluated on public sidewalks.

In this article the authors propose an innovative digital dynamic assessment framework to assess the safe performance of SADRs in their objective context, which is based on the theory that there exist different degrees of difficulty within the Operational Design Domain (ODD). The assessment framework is dynamic in the way it determines the difficulty of operating domains over time, based on data and experience gathered from previous SADR performances. The idea is to first test SADR performance in the known easiest real world conditions, after which SADRs will be allowed to undergo progressively more difficult operating conditions when it is proven that in the easier conditions a safe performance is achieved. To do so, actual location circumstances are monitored and stored in a database along the performance of an SADR.

The contributions of this article are:

- The theory that there exist different degrees of difficulty within the ODD.

- A set of adapted metrics to objectively and completely quantify and qualify factors that cause ODD difficulty for SADRs.
- The theoretical development of a dynamic assessment framework for the safe and sustainable performance of SADRs on public sidewalks.

This article is organized as follows: Section II discusses the substantive methods used in the different research process steps, Section III outlines the theories on which the design suggestion is based. Section IV introduces the set of adapted metrics to map the ODD and the proposed dynamic assessment framework. Section V discusses expert perspectives on the development of the theory of relative ODD level difficulty and the proposed dynamic assessment framework. Section VI and Section VII present the discussion and conclusion of the results. Section VIII presents an outlook on future research directions.

II. METHODOLOGY

The article will follow the design science research (DSR) approach to study and create an artifact that is innovative and solves a real-world problem (Merwe et al., 2017). For this article specifically the DSR cycle approach of Vaishnavi & Kuechler (2004) is adopted. The process steps included in this article are: the problem awareness step, the design suggestion step, and the design validation step. The design goal in this article is to design a dynamic assessment framework for the safe performance of SADRs on public sidewalks in the Netherlands. The outcome of the executed process steps can be used as the basis for the remaining process steps to develop a fully functioning digital dynamic SADR safety assessment system. The article structure is visualised in Figure 1.

In the problem awareness step by means of a review of the literature and desk research the theory on ODD and ODD classification is discussed. The same methods, extended by an informal interview with Prof.dr. Marjan Hagenzieker, are used to identify the risks associated with Automated Driving Systems (ADSs). In the design suggestion process step the theory of ODD difficulty is explored and a set of adapted metrics to map the complete ODD is presented. Additionally, based on the theory of different levels of difficulty within the ODD, the dynamic assessment framework is elaborated upon. In the validation process step both the research process and the designed dynamic assessment framework are validated by executing semi-structured interviews with nine experts, both practitioners and academics in the field (Hevner, 2007).

III. PROBLEM AWARENESS

A. OPERATIONAL DESIGN DOMAIN

Per SAE standard J3016, published in 2021, the Operational Design Domain for a driving automation system is defined as “operating conditions under which a given driving automation system, or feature thereof, is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics” (SAE International, 2021, p. 32). The Operational Design Domain has been created to distinguish between the operating domains where the self-driving system is able to control the vehicle and where the driver should control the vehicle. The ODD is also used in the evaluation of self-driving systems. A major aspect of overall system validation is ensuring that autonomous cars will function satisfactorily in their intended operational context (Koopman & Fratrik, 2019). The greatest challenge in safety assessment is that road traffic is an open parameter space in which an infinite number of different traffic situations can occur (Riedmaier et al., 2020), and that it is not viable to test all possible combinatorial road traffic situations, for example by simulation or on test sites, to classify an autonomous driving system as safe. According to Koopman & Fratrik (2019) to ensure that training and testing of Autonomous Vehicles (AVs) is complete, this requires at least ensuring that all aspects of the ODD have been addressed. This can either be done by assuring safe system operation or by ensuring that the system is capable of detecting and mitigating deviations from the defined ODD (Koopman & Fratrik, 2019).

It should be mentioned that the current knowledge of describing and classifying ODDs or derivation thereof is limited. There is not yet an objectively accepted methodology to describe or classify factors and elements that make up ODDs. Current publications or practical tests with self-propelled systems often describe the domain high level and in a qualitative way, so that no relative comparison can be made with other domains and the knowledge gained cannot easily be translated to new geographical areas. By classifying the infrastructure and environment according to objective factors, a start can be made to structure the testing and learning of self-driving systems on public roads. Besides the fact that there are few publications on the ODD, there is no approach to classify the ODD of the sidewalk yet. Most knowledge of the Operational Design Domain for a driving automation system has been gained in the light of self-driving (passenger) vehicles. These vehicles logically operate on public roads, so nothing is known yet specifically about the operating domain of the sidewalk, where SADRs will operate.

B. 6-LAYER MODEL

In this article, the 6-Layer Model by Scholtes et al. (2021) will be used to map the complete ODD for SADRs that are intended to operate on the public sidewalk. This model is used because it is structured, provides specific guidelines on how to be used, has a holistic approach by including the digital connectivity layer. and has already proven to be useful to map urban environments by other scholars (Khatun et al., 2021).

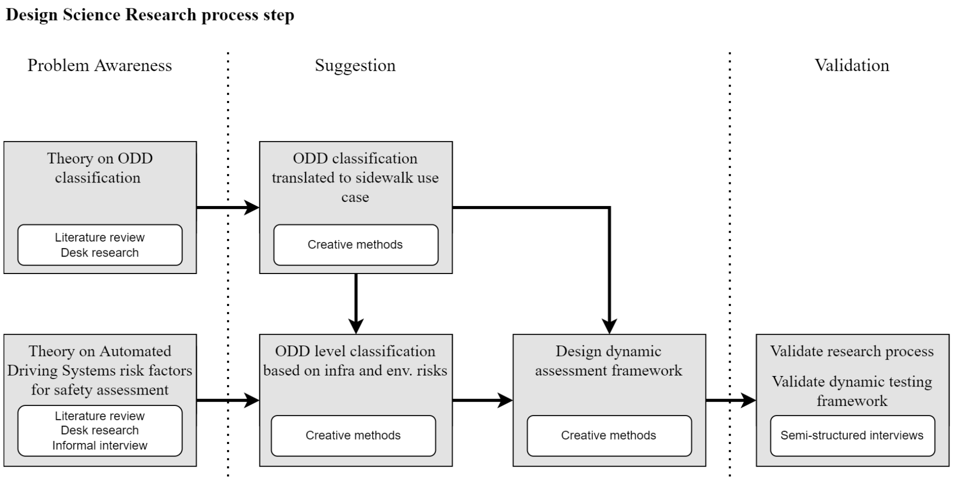


FIGURE 1: Article flow framework

Schuldt (2017) designed a 4-Layered model to structure environments for automated driving systems. The model was first extended by a fifth layer (Bagschik et al., 2018), and then by a sixth layer by Bock et al. (2018) to describe motorway scenario's. Scholtes et al. (2021) built on this work to describe urban scenario's with the 6-Layer Model. The 6-Layer Model is used in the concept as a tool to structure influencing factors on ADSs, allowing for the formation of scenario equivalence classes and, as a result, the selection of appropriate test methods (Scholtes et al., 2021). The six layers present in the model are shown in Figure 2. Layer 1 describes the road network and permanent traffic guidance objects. Layer 2 includes the roadside structures. Layer 3 is comprised of temporary modifications of elements of Layer 1 and Layer 2. Layer 4 includes the complete description of dynamic objects. Layer 5 contains environmental conditions and also includes road weather conditions. Layer 6 is defined to focus on all kinds of information exchange, communication, and cooperation on basis of digital data only (Scholtes et al., 2021, pp. 59135-59140).

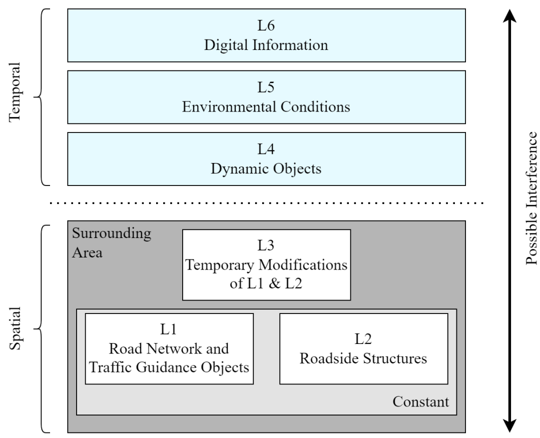


FIGURE 2: 6-Layer model overview of layers with spatial and temporal separation, adapted from Scholtes et al. (2021).

C. ADS PERFORMANCE RISKS

The scientific literature on risks associated with SADRs on public sidewalks is non-existent. Therefore, this section discusses the performance risks for ADSs, to be translated to the SADR case in Section IV.

1) INFRASTRUCTURE RISKS

The Operational Design Domain is the collection of operating conditions under which a given driving automation system is designed to function. When a given driving automation system finds itself in operating conditions for which it was not designed to function or detects a failure of the driving system, the driving task is handed over to the human driver (Boggs et al., 2020). According to Hillman & Capaldi (2020, p. 4) “disengagements are triggered when an AV cannot correctly match the perceived information with known datasets, due to the presentation of ambiguous or incomplete stimuli”. The neural networks that ADSs rely on for manoeuvring are trained on millions of photos and video frames to enable the correct recognition and identification of a stimulus, but there are few opportunities to train the system to recognize edge situations, because edge situations are so uncommon. In their research, Mihalj et al. (2022) found that static elements such as ambiguous traffic signs and deviations from standard road markings have a higher probability to lead to recognition failure by ADS, but do not mention what the critical limits for ambiguous traffic signs or deviations from standard road markings are to confuse ADS. Mihalj et al. argue that using a higher grade of retroreflective material or sheeting on traffic signs will improve the visibility of the signs under all environmental conditions and decreases sign degradation over time. Additionally, Storsæter (2021) concluded that colors, patterns, and textures can be utilized to improve the visibility of existing road infrastructure elements, including guardrails, dividers, and road markings to aid automated detection. Furthermore, for lane (departure) detection with cameras, the contrast between road markings and road surface was found to be more essential than retroreflectivity measures. Czarnecki (2018) argues that surface damage poses a direct risk for ADS. Research on unintended lane or road departure is mentioned more often in the evaluation of ADS, because lane departure is a clear indicator of unsafe driving. Next to the visibility of road infrastructure elements and the ability to detect them, road geometry, associated visibility distance and the speed of the vehicle are important factors that influence the probability to have a road or lane departure (Leroy et al., 2020; Farah et al., 2018; Czarnecki, 2018). Different road friction coefficients influence vehicle maneuverability and stability. The literature analysis by Farah et al. (2018) identifies the same infrastructure aspects that affect the ADS's driving ability as discussed in this section.

2) ENVIRONMENTAL RISKS

The influence of environmental factors on ADS safe driving performance is recognised by several scholars. Leroy et al. (2020) only includes weather conditions in their environment related risks concept group. Specific weather conditions have a direct impact on the road surface state and the visible distance for object detection. Rain, fog, sleet, snow and dense dust clouds will reduce the visibility distance and specific light conditions such as dusk or the setting sun can significantly influence the visibility level. Roadway conditions can change induced by the weather (Thorn et al., 2018; Rao et al., 2021). For example heavy rain or snow can flood the roadways or reduce the visibility of the road markings. Ren et al. (2019) studied environmental influences on the uncertainty of object detection, by a deep neural network methodology. By measuring the average precision of object detection the impact of dark, sunset, rain and motion blur were assessed in comparison to a base scenario. It was found that in comparison to the base scenario the average detection precision decreased at dark, is close to the base scenario at sunset and light snow conditions, seriously decreases at rain and is the worst in the sample for motion blur.

From practice, conclusions can also be drawn about the impact of environmental factors on the driving ability of AVs. Research on data from Waymo, one of the pioneers in ADS, has revealed that their vehicles have not been operated during inclement weather, such as heavy rain and dust storms (Schwall et al., 2020). While it is not stated explicitly, this is most reasonably linked to the safety risks associated with certain environmental conditions. The environmental characteristics that affect the safe movement of an AV also affect other road users. For example, limited object detection capabilities by ADS due to dense fog means that an AV can map less of its surroundings, but it also limits the visibility for other road users. This means that an AV can be poorly visible, which can lead to unsafe situations caused by other road users and should therefor also be taken into account regarding safety risk assessments.

Leroy et al. define obstacles as “objects present on the road surface and in the surrounding area of an ego-vehicle” (Leroy et al., 2020, p.213). The risks originating from obstacles depend on three different characteristics of the obstacle: the distance to the obstacle (in both the longitudinal and lateral direction), the dynamics of the obstacle (described by the yaw rate, longitudinal and lateral speed, and obstacle acceleration), and the type of obstacle. The distinction made in type of obstacle is between cars, pedestrians, cyclists, motorcyclists, busses, trucks and unidentified objects. These different obstacle types are of interest to know because different obstacle groups have different expected behaviours, characteristics, goals and, if possible, different traffic rules to abide by. The different obstacle types therefore pose different types of risks.

3) VULNERABLE ROAD USER INTERACTION

Vulnerable road users are “non-motorised road users, such as pedestrians and cyclists as well as motorcyclists and persons with disabilities or reduced mobility and orientation” (Tabone et al., 2021, p. 2). The interactions between vulnerable road users and ADS are not always logical or predictable, which poses additional risks that have to be addressed. For example, “the unpredictability of pedestrians makes it almost impossible for people or algorithms to avoid collisions” (Tabone et al., 2021, p. 8). In road traffic, scene dependent scanning behaviour is a phenomenon (Theeuwes & Hagenzieker, 1993). This means that people look for visible elements in places where they expect them, and consequently overlook elements in places where they do not expect them. The risk for ADS that gradually replace regular vehicles of being badly noticed by this phenomenon is small. Most AVs have a similar appearance to regular vehicles and are expected to show the same behaviour and follow the same traffic rules, but this phenomenon may not be overlooked for more futuristic and smaller sized autonomous systems, such as delivery robots.

Pokorny et al. (2021) investigated the interaction between vulnerable road users and an ADS equipped shuttle based on video footage of these interactions. The first observation is that the shuttle used had a very defensive driving style and the unnecessary triggering of hard stops caused risky situations for motorised vehicles that drove close to the shuttle. The low speed of ADS shuttles, which results from safe defensive driving style programming, is recognised in the literature (Heikoop et al., 2020; Hagenzieker et al., 2020). From practice, examples can be identified where the known defensive driving style of ADS is a reason for other road users to abuse it, for example by taking the right of way, cutting the vehicle off, blocking the vehicle or in some other way testing the braking capabilities (Heikoop et al., 2020; Tabone et al., 2021). According to Heikoop et al. (2020), ADSs are not yet able to deal properly with other road users who deviate from the formal rules in such manner.

4) SAFETY THROUGH DIGITAL CONNECTIVITY

Several researchers point out that the use of digital solutions can partly reduce safety risks in road traffic. Sensor fusion and road network digitisation via vehicular communication and digital maps provide ways to increase the overall road network's resilience by offering redundancy, according to Mihalj et al. (2022). For example, smart traffic sign technology and smart traffic light technology can ensure that in circumstances where the traffic signs or lights are poorly visible to machine vision systems, a vehicle can still know with certainty what the current speed limit is or whether, for example, an intersection can be crossed safely. Thorn et al. (2018) write that all infrastructural and environmental characteristics information can be digitised so that a vehicle can use this data to make or validate a safe driving choice. Digital dynamic maps can be established to guide connected vehicles through temporary lane closures or variable speed limits. The input from a static lidar sensor can be used to provide a connected vehicle with information that a connected vehicle cannot yet see on the basis of its own sensor information, for example about approaching vehicles that have priority (Bluecity, 2022).

IV. DESIGN SUGGESTION

In this section, the design basis and design suggestion will be presented. First, the set of adapted metrics to objectively and completely quantify and qualify factors that make up the ODD of SADRs will be presented, after which the effect of these factors on the actual SADR performance will be discussed. Based on this knowledge, the design of the dynamic test framework, associated living lab approach and performance indicators will be addressed.

A. COMPLETE AND MEASURABLE ODD OF SADRS

By applying the 6-Layer model the complete possible description of the ODD of SADRS has been drawn up. For each identified factor a metric has been determined to measure this factor, see Table 1. Why it is of interest to map and measure those specific factors is portrayed in Figure 7. In Figure 7, based on the infrastructural, environmental, road user interaction and digital factors identified from literature that impact ADS performance, it was determined how factors in the ODD of SADRs affect SADR performance. This influence is represented in the form of a conceptual causal model.

TABLE 1: Set of metrics to measure ODD difficulty factors

Factor	Sub-factor	Unit
Layer 1		
Sidewalk curvature		degree (°)
Non-sidewalk infrastructure at crossing[b]		One-way cyle path - Bi-directional multilane road with adjacent cycle path
Pedestrian crossing presence at crossing		binary (0,1)
Slope of sidewalk surface		degree (°)
Sidewalk surface friction coefficient		coefficient
Sidewalk entrance geometry		degree (°)
Sidewalk Width		meter (m)
Traffic light presence at crossing		binary (0,1)
Traffic signs guiding traffic participants		binary (0,1)
Layer 2		
Buildings	length	meter (m)
	width	meter (m)
	height	meter (m)
Static obstacles	length	meter (m)
	width	meter (m)
	height	meter (m)
Driveable obstacles		percentage (%)
Utility well cover or pothole		binary (0,1)
Physical separation with adjacent lane		binary (0,1)
Layer 3		
Pavement chalk		percentage (%)
Scaffolding construction	length	meter (m)
	width	meter (m)
	height	meter (m)
Road construction work	length	meter (m)
	width	meter (m)
	height	meter (m)
Fallen tree, pole, traffic light	length	meter (m)
	width	meter (m)
	height	meter (m)
Layer 4		
Traffic participants not on sidewalk	length	meter (m)
	width	meter (m)
	height	meter (m)
	speed	meter per second (m/s)
Traffic participants on SADRs route	length	meter (m)
	width	meter (m)
	height	meter (m)
	speed	meter per second (m/s)
Unintended use of the sidewalk	length	meter (m)
	width	meter (m)
	height	meter (m)
	speed	meter per second (m/s)
Animals	length	meter (m)
	width	meter (m)
	height	meter (m)
	speed	meter per second (m/s)
Miscellaneous objects	length	meter (m)
	width	meter (m)
	height	meter (m)
	speed	meter per second (m/s)
Vulnerable road users	length	meter (m)
	width	meter (m)
	height	meter (m)
	speed	meter per second (m/s)
Layer 5		
Lux		lux (lx)
Precipitation		millimeters per hour (mm/h)
Fog, visibility distance		meter (m)
Temperature		degrees Celsius (°C)
Wind		kilometer per hour (km/h)
Layer 6		
V2P		percentage (%)
V2I		percentage (%)
V2N[b]		No connectivity - 5G connectivity
V2V		percentage (%)
V2C		binary (0,1)
Other		
SADR component quality[b]		Low - High
SADR operating speed		meter per second (m/s)
SADR size	length	meter (m)
	width	meter (m)
	height	meter (m)
	weight	kilogram (kg)
Known interaction rules[b]		No traffic rules - Fully developed traffic rules

B. CAUSAL CONCEPTUAL MODEL

The successful functioning of an SADR can be divided into three parts: the degree to which the SADR is able to correctly detect and recognise environmental elements, the degree to which the SADR is able to plan a route in real-time and the degree to which the SADR is able to actually manoeuvre in the physical environment (Synopsys, 2022).

[b]Factor measured at qualitative scale, portrayed with lower and upper bound of scale.

From Figure 7 it turns out that these three driving tasks are mainly affected by intermediate constructs. First, a low visibility distance and edge case occurrence prevent SADRs from perceiving or planning correctly. Second, reduced sidewalk conditions, specific sidewalk geometries and the presence of static and dynamic objects make SADRs technically unable to plan or drive a path. Third, a reduced visibility distance and a low visibility of an SADR cause an increased risk in the interaction with other road users. More in depth, the exact impact can be examined from the overview per layer of the 6-Layer model.

From Layer 1 can be concluded that infrastructural elements such as a pedestrian crossing, traffic signs and traffic lights at intersections help to ensure that an SADR is able to plan a route and that other road users show safe driving behaviour at these locations. The sidewalk geometry and friction coefficient of the sidewalk surface material determine whether an SADR is technically able to manoeuvre. The curvature of the sidewalk impacts the visibility distance of an SADR. From Layer 2, it follows that the presence of roadside structures reduces the visibility distance. In addition, roadside structures influence SADR manoeuvrability. Temporary modifications, as described with Layer 3, result in reduced visibility and a lower manoeuvrability due to their presence, but also make it more difficult for an SADR to find landmarks in the surrounding area. This increases the occurrence of edge cases. From Layer 4 follows that all different dynamic objects that can be identified on sidewalks result in a reduced visibility distance and a lower manoeuvrability due to their presence. From Layer 5, it can be noted that, in general, the better, clear and calm weather conditions provide the best visibility distance and road conditions. This also ensures that an SADR is most visible to other road users, which improves overall traffic system safety. Layer 6 covers all the digital means that can assist an SADR in perceiving and planning by artificially increasing the visibility distance and reduce the number of edge cases that an SADR has to deal with. From the last added layer with ego vehicle characteristics it becomes clear that SADR characteristics can affect the ability to perceive, plan and manoeuvre, and that established rules for the driving behaviour of SADRs and for the interaction with SADRs benefit the (unintended) unsafe behaviour of other road users.

C. DYNAMIC ASSESSMENT FRAMEWORK

Acknowledging different levels of difficulty within the ODD enables a more nuanced evaluation of the performance of SADRs in their objective context. The relative difficulty associated with a specific sidewalk section can vary over time due to the dynamic nature of the urban environment, and SADR performance evaluation must take this into account. To be able to evaluate this performance objectively, the exact location characteristics are therefore monitored and stored in a database alongside the actual SADR performance metrics. When it is actually known in which contexts an SADR can most easily drive safely, a secure proof-of-concept method to actually test SADRs on public sidewalks, where SADRs are allowed more as their driving technology is proven to be safe (Humblet, 2021), can be rolled out in phases in the Netherlands.

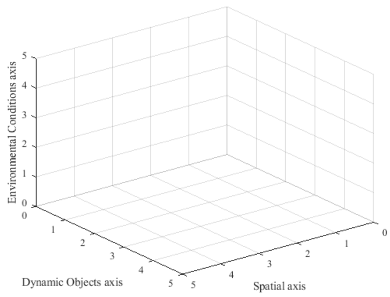


FIGURE 3: 3D ODD classification space

The analysed layers of the 6-Layer Model can be divided into a spatial axis (Layer 1, 2 and 3), a dynamic objects axis (Layer 4) and an environmental characteristics axis (Layer 5). The sixth layer, dependent on the digital functionalities the sixth layer includes for a specific geographic location, can affect all three axes. This resulting matrix can form the basis for the proof-of-concept methodology (Figure 3). This matrix is the complete abstract space in which every geographic location can be classified in relative difficulty regarding spatial elements, dynamic objects and environmental characteristics. For illustrative purposes it has been chosen to present the three axes on a scale from 0 to 5. To date these scores cannot be benchmarked against one another since emperical data is missing. By distinguishing between these three axes, the relative difficulty along each axis can be determined separately and then combined into a unified ODD difficulty classification. Based on a unified ODD difficulty per link in a sidewalk network, a real-time dynamic assessment framework map on ODD difficulty can be plotted (see Figure 4). In Figure 5 the underlying link node representation of the dynamic assessment framework map used in this article is portrayed. Based on the performance and location data that is accumulated in the database over time, it becomes possible to determine the relative difficulty of specific locations with increasing accuracy. Lessons can be learned from the relationship between these environmental factors and the performance of SADRs, and the assessment framework can be updated accordingly. Additionally, because operating conditions are accurately monitored, knowledge is gained that can be translated to other geographical locations.

The data present in the database system can be used to add functionalities to the digital dynamic assessment framework. For example, by clicking on a specific link in the network map in Figure 4, an explanation can be displayed of the elements in the operating environment that have the greatest impact on the ultimate ODD difficulty level for a delivery

FIGURE 4: Dynamic assessment framework map - illustrative

FIGURE 5: Node-link representation applied to dynamic assessment framework map

robot. This representation can give stakeholders insight into why a location is or is not passable for a delivery robot. A possible incident reporting function would be useful, because incident reporting is mandatory in leading countries on autonomous driving innovations and also contributes to system transparency (NHTSA, 2021; Canada Transport, 2021). In addition to making the system understandable to human stakeholders, the dynamic assessment framework can be linked to the navigational component of delivery robots (Google, n.d.). By means of geofencing, a delivery robot can be denied access to locations with difficult ODD levels if it has not yet proven its ability to perform safely in such circumstances.

Whether a sidewalk is passable for an SADR depends on the position and composition of surrounding elements (Coppola & Marshall, 2021). For the spatial axis therefor the ‘Sidewalk Minimum Clear Width’, defined as the narrowest passage point within a sidewalk polygon, and the ‘Sidewalk Minimum Clear Width Accounting for Static Obstructions’, which additionally takes the impact of obstacles on the clear width for pedestrian access into account, established by (Coppola & Marshall, 2021), can be used to evaluate SADR performance against. For the dynamic objects axis the ‘average busyness by VRUs’ and the percentage of ‘total occupied sidewalk area’ are proposed.

As mentioned in the introduction of this article it is currently unknown what location characteristics are truly difficult for SADRs to cope with. It is therefore argued that the impact of the spatial and environmental characteristics axis on SADR performance should first be determined on a closed test site. Thereafter, the safest approach is to allow an SADR at the easiest locations first. By making safe kilometers in this specific, monitored setting, more difficult settings can then be gradually allowed.

D. LIVING LAB APPROACH

According to Bergvall-Kareborn et al. (2009, p. 1) “a living lab is a gathering of public–private partnerships in which businesses, researchers, authorities, and citizens work together for the creation, validation, and test of new services, business ideas, markets, and technologies in real-life contexts” and is a living lab “an environment in which people and technology are gathered and in which the everyday context and user needs stimulate and challenge both research and development, since authorities and citizens take active part in the innovation process”. Quak & Nesterova (2021) argue that the focus of a living lab is on practical implementation, learning and improvement, which is in line with what is proposed in this article. Because to date very little is known about the actual performance of SADRs in real world conditions, and simulated performance cannot be assumed to be safe with certainty (Shetty et al., 2021), structured sidewalk testing can overcome this challenge. Not only about the SADR performance, but about numerous aspects of SADRs in the public environment in the Netherlands little or nothing is known. A living lab is more sustainable, more educational and more adaptive to changes due to the involvement of different stakeholders, the predefined goals, the longer duration and the iterative development approach of living lab projects compared to standalone pilots (Quak & Nesterova, 2021).

The advantage of the dynamic assessment framework proposed in this article is that it focuses on learning about SADR performance and the collection of data, which — combined with the living lab approach — allows to learn about the complete traffic system that includes SADRs. The third and perhaps most important focus of living labs is the possibility for improvement. Improvements to the dynamic assessment framework can be made as more is learned about the performance of the robot in real world environments. SADRs are also far from being fully developed and improvements need to be properly evaluated in public environments. In addition, from a technical perspective SADRs can perform safely, but there are social developments which could make the system requirements stricter or less strict. The ongoing development of the innovation, the assessment framework and the system requirements fit in well with the iterative approach used in living labs. The presence of a learning and development approach is also an important requirement for the Netherlands Vehicle Authority to test automated functions on public roads (van der Stoep, 2022). Adopting a living lab approach ensures this learning and development approach. A schematic representation of the functioning of dynamic assessment framework and the associated living lab approach is given in Figure 6.

E. SADR PERFORMANCE INDICATORS

Wishart et al. (2020) created a comprehensive list of driving safety performance metrics for AVs based on an evaluation of 50 references that included one or more performance metrics. In this section, the performance metrics deemed relevant during an initial field test with SADRs and the dynamic assessment framework are summarized:

- Minimum Safe Distance Violation (MSDV), "an instance in which the actions of the ego vehicle result in encroaching upon its safe boundaries with another (safety-relevant) entity within the scenario environment, as defined by current velocities and acceleration capabilities of both entities" (Wishart et al., 2020, p. 3).
- Proper Response Action (PRA), "an instance of an action (longitudinal and/or lateral acceleration) taken by the ego vehicle to restore itself to its calculated safety boundaries after a safe distance violation has occurred." (Wishart et al., 2020, p. 4).
- Minimum Safe Distance Calculation Error (MSDCE), "the accuracy of the ADS calculation of its safety boundaries with respect to other safety-relevant entities in comparison to ground truth" (Wishart et al., 2020, p. 5).
- Collision Incident (CI), "an instance of the ego vehicle being at fault in a collision, as determined by an examination of the data from the on-board and/or off-board sensors, potentially including the vehicle event data recorder, along with the police report" (Wishart et al., 2020, p. 6).
- Rules-of-the-Road Violation (RRV), "an instance of the ego vehicle violating a traffic regulation that would result in an infraction or citation" (Wishart et al., 2020, p. 6).
- ADS Active (ADSA), "a confirmation that the ADS is active while executing the behavioral competency" (Wishart et al., 2020, p. 7).
- Modified Time-to-Collision (MTTC), "the time until a collision between two entities in the scenario environment would occur if both continue with the present velocities and accelerations" (Wishart et al., 2020, p. 9).

The ADSA metric is important, because in this way the actual driving behaviour of an SADR is assessed and control by a teleoperator is not wrongly used to classify the autonomous driving system as safe. The MSDCE metric provides insight into the average deviation in the perceived and actual geographic location of an SADR and can be used to determine applicable safety margins. The RRV metric indicates the need for (basic) traffic rules for SADRs. Violations and obeying to these rules can be objectively measured. The MTTC can provide insight into how often unsafe situations occur that do not immediately lead to a CI. All of the above performance metrics can be calculated. For the exact formulas is referred to the work of Wishart et al. (2020). It must be noted that data regarding the velocity, speed and exact location of individual SADRs are to be stored in the database system as well. Although their parameter values on its own do not give any information about SADR performance, the complete set of data can be analysed to investigate incidents.

V. VALIDATION

Concerning the development of different levels of difficulty within the ODD experts are predominantly very positive. the methodological approach is rated as complete. In terms of exact ODD difficulty content, most experts mention that there is a need for empirical evidence that quantitatively supports the hypothesis of relative levels of difficulty within the ODD. About whether the presented method is going to contribute to standardization regarding the mapping of the ODD there are two different perspectives: the first is that the proposed method is a good start now and will develop over time, the second is that the method will need further development and more detail before it can be used. Several experts mention additionally that the human interaction component regarding the risks that exist for SADRs is currently under researched.

There are two perspectives among experts regarding the applicability of the dynamic assessment framework designed in this article. These perspectives are in line with the conclusion of the establishment of different levels of difficulty within the ODD. The majority, seven out of nine, of the experts state that they like the risk minimising and learning approach that is proposed with the living lab and proof of concept methodologies that go together with the dynamic assessment framework. Over time ODD difficulty will be mapped as accurately as possible according to the performance data gathered on public sidewalks. On the other hand, the experts who felt that the risks, on which ODD difficulty in this article is theoretically based, were not specific enough also felt that the development of the theoretical assessment framework on the basis of these factors was not good enough yet to be brought to practice. Again, empirical evidence that quantitatively supports the hypothesis of relative difficulty within the ODD will bridge the gap between the theoretical framework and its further development in practice.

VI. DISCUSSION

The main points that will be discussed are the accuracy of the risk factors used to explain ODD difficulty and the dif-

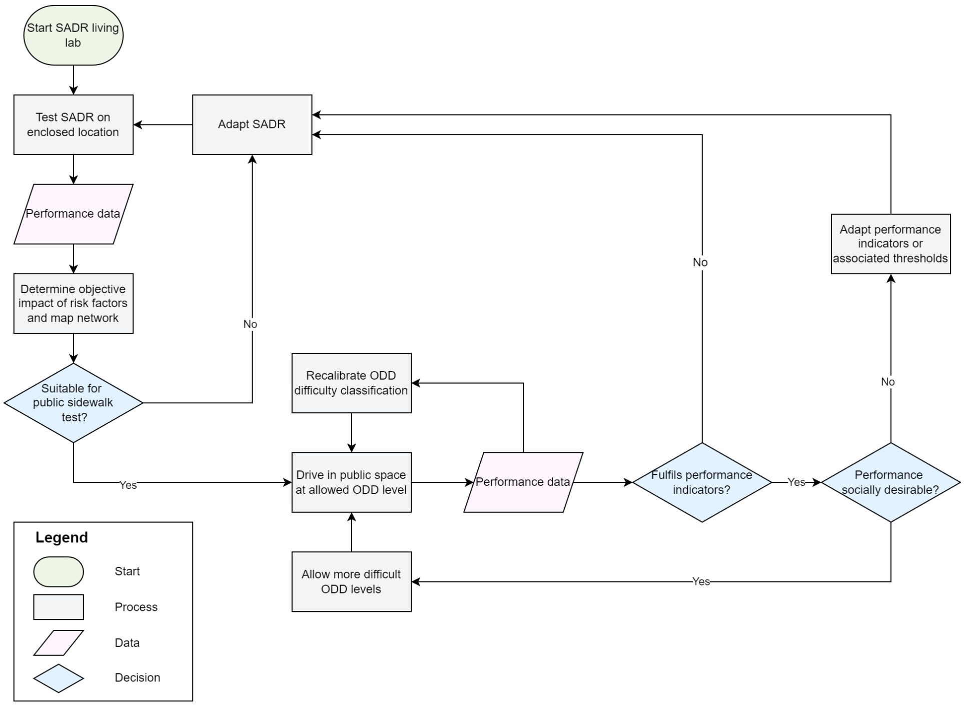


FIGURE 6: Process map dynamic assessment framework with living lab approach

ficulties that arise when the dynamic assessment framework is taken from theory to practice.

A. ACCURACY OF ODD DIFFICULTY RISK FACTORS

In this article, the theory is explored in which the ODD is no longer approached as a binary concept, but where, through the presence of environmental elements, a relative level of difficulty can be found within the ODD. This theory is backed up by theoretical evidence and validated by experts in the field. During this research the line of reasoning is used that risks that deteriorate the performance of an AV can pose a similar, not necessarily equal, risk to deteriorate the performance of SADRs. This line of reasoning is followed because AVs and SADRs use comparable hardware and software components to execute the autonomous driving tasks. The presence or absence of these risk factors leads to relative difficulty of ODDs. However, as several interviewees in the validation process of this article noted, this insight must be substantiated with empirical data. Because of the lower operating speed of SADRs in comparison to AVs it might be that risk factors that do deteriorate the AV performance do not necessarily deteriorate the performance of an SADR. In addition, because the operating environment of an SADR differs from the operating environment of an AV, there might be risk factors of influence on the SADR performance which have not been identified in the complete literature study on ADS risk factors. Also, there are no (scientific) publications as of yet that specify the risk factors associated with SADRs on public sidewalks. Therefore, the risk factors for SADRs identified in this article might not be complete or risk factors might be included that do not affect the SADR performance at all. According to the interviewees this does not invalidate the theory of relative difficulty levels existing within the ODD. Risk factors may be adjusted or added to the presented overview over time. This research contributes to partially removing the encountered issues for follow-up research.

The overview of ODD risk factors and their causal impact on the performance of an SADR are identified and displayed per layer of the 6-Layer Model. In practice, this separation between the six layers is not so pronounced and the different layers may influence each other. For example Layers 1, 2 and 3 in the design suggestion have been aggregated or the sixth layer with digital communication means that can possibly reduce the impact of other road user's presence (Layer 4) on the SADR performance. Because the different layers have been deliberately taken apart and depicted separately, Figure 7 is a simplified representation of reality. Again, it must be noted that empirical research is the next step to find out how and which factors contribute to relative ODD difficulty.

B. DYNAMIC ASSESSMENT FRAMEWORK TO PRACTICE

A current limitation of the proposed dynamic assessment framework is that it assumes that there are indeed different levels of difficulty within the ODD and that this difficulty can be objectively determined. The experts interviewed for the validation of this research agreed that the theory of existence of different ODD levels is correct, eight out of nine experts agreed and one stated to have no opinion. However, most of the experts also stressed the importance of empirical data to support this theory and stated that it is not possible to actually develop this assessment framework on a theory until it has been proven. According to some of the interviewees, what is proposed is correct and feasible from a theoretical perspective. Especially since the steps that are followed are logically connected and reproducible. However, from a practical perspective, there are still many barriers to be overcome before the proposed dynamic assessment framework can be used to guide SADR tests on public sidewalks. Once empirical relationships between risk factors and SADR performance have been found, an important consideration is how to digitise these risk factors to be able to use them in the dynamic assessment framework. If the impact of monitored risk factors on ODD difficulty changes over time, this does not pose a problem because the framework can be altered, but when the impact of new, not yet network-wide monitored risk factors on ODD difficulty become statistically significant over time, these risk factors should also be digitized. How best to approach this needs further research.

A strength of the dynamic assessment framework is its associated living lab approach. This recognises that the traffic system is dynamic, but also that system requirements and user preferences can change over time. The current knowledge on autonomous innovations is insufficient to develop a rigid assessment system for autonomous innovations. The living lab approach contributes to the assessment framework and the deployment of SADRs developing in a sustainable way and to ensuring that the assessment method does not become obsolete in the short term.

VII. CONCLUSION

This article has focused on a dynamic assessment framework design to assess SADRs on safe performance on the public sidewalks in the Netherlands. This framework design is based on the theory that there exist different degrees of difficulty within the Operational Design Domain, caused by the presence of specific characteristics of the environment. The development of this theory of relative ODD difficulty is the first conclusion of this article. The associated level of difficulty for a specific geographic location can vary over time due to the dynamic nature of the public environment. A dynamic operating domain that changes in difficulty is much better suited to the development and assessment of SADRs than the traditional binary approach to the concept of the Operational Design Domain and the rigid safety testing procedures of vehicles. By using different levels of difficulty within the ODD, and having insight into what actually causes this relative difficulty objectively, it is possible to steer SADR development in a much more structured way towards the critical parts of an SADR that actually need further development. At the same time, by applying this nuance within the ODD, knowledge can be gained about the performance of SADRs in the public domain in a structured and risk-minimising way. The proposed assessment framework is also innovative in the way it not only investigates whether SADRs perform technically safe, but also works towards the most socially acceptable integration of SADRs into the existing traffic system.

The proposed dynamic assessment framework can be of value because it provides guidance to integrating SADRs into the Dutch traffic system and society over time. This integration is important because SADRs are a means to mitigate the negative externalities associated with last-mile delivery. According to experts interviewed the theory of different levels of difficulty within the Operational Design Domain is promising and worthy of further elaboration. The actual development of the designed digital dynamic assessment framework can contribute to tests with SADRs on public sidewalks in the Netherlands. As emerges from the expert validations as well, empirical evidence is the missing link between the theoretically established ODD difficulty levels and the envisaged dynamic assessment framework. Development of the dynamic assessment framework should be driven by factors that actually impact the performance of SADRs. This and related follow-up questions arising from this article are discussed in detail in the recommendations section in Section VIII.

VIII. RECOMMENDATIONS

The first recommended future research direction is to verify whether there is indeed an empirical relationship between the location characteristics identified in this article and the performance of SADRs. These empirical relationship(s) are not necessarily linear and self-contained. Thorough research needs to be done on the best models to explain ODD difficulty. When the empirical relation between ODD factors and SADR performance is confirmed, and testing on public sidewalks is allowed based on the proposed proof of concept methodology, a subsequent study could focus on the usefulness of the proposed performance indicators for SADRs. Additionally it will then be possible to investigate which elements, including digital Layer 6 solutions, actually impact the performance of SADRs and how these elements should be digitally captured as efficiently as possible to actually realise the digital dynamic assessment framework proposed in this research.

A second research direction should focus on the interaction between humans and SADRs. Finding out about user acceptance is essential to successfully deploy SADRs in public. Research can focus on the appearance of SADRs to increase their acceptance. There are studies in this direction (e.g. de Groot (2019)), but this has not yet been widely studied for SADRs. The other side of this research direction can focus on: what is the most desirable behaviour of a robot for other road users? The results of this research can on the one hand be used by an SADR manufacturer or operator to increase the acceptance by adjusting the behaviour, on the other hand the results can provide policy makers with tools to define initial traffic rules and guidelines for SADRs (and similar innovations). This research direction can be explored in the envisaged living lab.

A third research direction should focus on the broader development theory of different difficulty levels within the Operational Design Domain. First of all, it would be scientifically useful if this theory is also confirmed by other studies and researchers. Currently, this theory has only been examined for the ODD of the sidewalk and the innovation of an SADR. It is scientifically relevant whether this theory also applies within other operating domains and for other innovations. The follow-up question could be whether the notion of different ODD difficulty degrees is also a useful insight for the performance assessment of other vehicles, or whether this is only true for SADRs because of their unique characteristics.

REFERENCES

Bagschik, G., Menzel, T., & Maurer, M. (2018). Ontology based scene creation for the development of automated vehicles. In *2018 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1813–1820). https://doi.org/10.1109/IVS.2018.8500632.

Bergvall-Kareborn, B., Hoist, M., & Stahlbrost, A. (2009). Concept design with a living lab approach. In *2009 42nd Hawaii International Conference on System Sciences* (pp. 1–10). https://doi.org/10.1109/HICSS.2009.123.

Bluecity. (2022). *Bluecity AI^{TM} Enhanced Mobility Data & traffic Actuation.* Retrieved July 23, 2022, from https://bluecity.ai/traffic-monitoring-solutions/

Bock, J., Krajewski, R., Eckstein, L., Klimke, J., Sauerbier, J., & Zlocki, A. (2018). Data basis for scenario-based validation of had on highways. In *27th Aachen colloquium automobile and engine technology* (pp. 8–10). https://doi.org/10.1080/15389588.2019.1630827.

Boggs, A. M., Arvin, R., & Khattak, A. J. (2020). Exploring the who, what, when, where, and why of automated vehicle disengagements. *Accident Analysis & Prevention*, *136*, 105406. https://doi.org/10.1016/j.aap.2019.105406.

Boysen, N., Schwerdfeger, S., & Weidinger, F. (2018). Scheduling last-mile deliveries with truck-based autonomous robots. *European Journal of Operational Research*, *271*(3), 1085–1099. https://doi.org/10.1016/j.ejor.2018.05.058.

Canada Transport. (2021, August 6). *Guidelines for testing automated driving systems in canada.* Retrieved May 31, 2022, from https://tc.canada.ca/en/road-transportation/innovative-technologies/connected-automated-vehicles/guidelines-testing-automated-driving-systems-canada

Coppola, N. A., & Marshall, W. E. (2021). Sidewalk static obstructions and their impact on clear width. *Transportation research record*, *2675*(6), 200–212. https://doi.org/10.1177%2F0361198121991833.

Czarnecki, K. (2018). Operational world model ontology for automated driving systems–part 1: Road structure. *Waterloo Intelligent Systems Engineering Lab (WISE) Report, University of Waterloo*.

de Groot, S. (2019, August). *Pedestrian acceptance of delivery robots.* [MSc Thesis]. Delft University of Technology. Retrieved February 24, 2022, from http://resolver.tudelft.nl/uuid:f9e8c003-c8fc-4075-bff3-0d54e0f0fecb

Farah, H., Erkens, S. M., Alkim, T., & Arem, B. v. (2018). Infrastructure for automated and connected driving: State of the art and future research directions. *Road vehicle automation 4*, 187–197. https://doi.org/10.1007/978-3-319-60934-8_16.

Google. (n.d.). *Provide useful information to your users when they are near an area of interest.* Retrieved June 4, 2022, from https://developers.google.com/location-context/geofencing

Hagenzieker, M., Boersma, R., Velasco, J., Ozturker, M., Zubin, I., & Heikoop, D. (2020). Automated buses in europe. *An Inventory of Pilots ", Version 0.5. TU Delft*.

Heikoop, D. D., Velasco, J. P. N., Boersma, R., Bjørnskau, T., & Hagenzieker, M. P. (2020). Automated bus systems in europe: A systematic review of passenger experience and road user interaction. *Advances in Transport Policy and Planning*, *5*, 51–71.

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems*, *19*(2), 4.

Hillman, R., & Capaldi, R. (2020). *Test methods for interrogating autonomous vehicle behaviour.* [Video]. HORIBA MIRA. Retrieved March 14, 2022, from https://www.horiba-mira.com/webinars/

Humblet, M. (2021, November 17). *What is proof of concept & how to actually use it.* [Video]. YouTube. Retrieved June 1, 2022, from https://www.youtube.com/watch?v=3BRrxjHGy6U

Jennings, D., & Figliozzi, M. (2019). Study of sidewalk autonomous delivery robots and their potential impacts on freight efficiency and travel. *Transportation Research Record*, *2673*(6), 317–326. https://doi.org/10.1177%2F0361198119849398.

Khatun, M., Glaß, M., & Jung, R. (2021). A systematic approach of reduced scenario-based safety analysis for highly automated driving function. In *Proceedings of the 7th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2021)* (pp. 301–308). http://dx.doi.org/10.5220/0010397403010308.

Koopman, P., & Fratrik, F. (2019). How many operational design domains, objects, and events? In *Safeai@ aaai.*

Leroy, J., Gruyer, D., Orfila, O., & El Faouzi, N.-E. (2020). Five key components based risk indicators ontology for the modelling and identification of critical interaction between human driven and automated vehicles. *IFAC-PapersOnLine*, *53*(5), 212–217. https://doi.org/10.1016/j.ifacol.2021.04.141.

Merwe, A. v. d., Gerber, A., & Smuts, H. (2017). Mapping a design science research cycle to the postgraduate research report. In *Annual Conference of the Southern African Computer Lecturers' Association* (pp. 293–308). https://doi.org/10.1007/978-3-319-69670-6_21.

Mihalj, T., Li, H., Babić, D., Lex, C., Jeudy, M., Zovak, G., ... Eichberger, A. (2022). Road infrastructure challenges faced by automated driving: A review. *Applied Sciences*, *12*(7), 3477. https://doi.org/10.3390/app12073477.

NHTSA. (2021, June 29). *NHTSA Orders Crash Reporting for Vehicles Equipped with Advanced Driver Assistance Systems and Automated Driving Systems.* Retrieved May 31, 2022, from https://www.nhtsa.gov/press-releases/nhtsa-orders-crash-reporting-vehicles-equipped-advanced-driver-assistance-systems

Pokorny, P., Skender, B., Bjørnskau, T., & Hagenzieker, M. P. (2021). Video observation of encounters between the automated shuttles and other traffic participants along an approach to right-hand priority t-intersection. *European Transport Research Review*, *13*(1), 1–13. https://doi.org/10.1186/s12544-021-00518-x.

Quak, H., & Nesterova, N. (2021). *Living labs for transitions in urban freight transport systems.* (Final draft paper)

Rao, S. J., Deosthale, E., Barickman, F. S., Elsasser, D., & Schnelle, S. C. (2021). *An approach for the selection and description of elements used to define driving scenarios.* (Report No. DOT HS 813 073). United States. Department of Transportation. National Highway Traffic Safety Administration.

Ren, L., Yin, H., Ge, W., & Meng, Q. (2019). Environment influences on uncertainty of object detection for automated driving systems. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 1–5). https://doi.org/10.1109/CISP-BMEI48845.2019.8965948.

Riedmaier, S., Ponn, T., Ludwig, D., Schick, B., & Diermeyer, F. (2020). Survey on scenario-based safety assessment of automated vehicles. *IEEE Access*, *8*, 87456–87477. https://doi.org/10.1109/ACCESS.2020.2993730.

SAE International. (2021, July 15). *Taxonomy & Definitions for Operational Design Domain (ODD) for Driving Automation Systems J3259.* Retrieved March 14, 2022, from https://www.sae.org/standards/content/j3259/e

Scholtes, M., Westhofen, L., Turner, L. R., Lotto, K., Schuldes, M., Weber, H., ... Körtke, F. (2021). 6-layer model for a structured description and categorization of urban traffic and environment. *IEEE Access*, *9*, 59131–59147. https://doi.org/10.1109/ACCESS.2021.3072739.

Schuldt, F. (2017). *Ein beitrag für den methodischen test von automatisierten fahrfunktionen mit hilfe von virtuellen umgebungen* (Doctoral dissertation). https://doi.org/10.24355/dbbs.084-201704241210.

Schwall, M., Daniel, T., Victor, T., Favaro, F., & Hohnhold, H. (2020). Waymo public road safety performance data. *arXiv preprint arXiv:2011.00038*.

Shetty, A., Tavafoghi, H., Kurzhanskiy, A., Poolla, K., & Varaiya, P. (2021). Risk assessment of autonomous vehicles across diverse driving contexts. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)* (pp. 712–719). https://doi.org/10.1109/ITSC48978.2021.9564744.

Stadler, C., Montanari, F., Baron, W., Sippl, C., & Djanatliev, A. (2022). A credibility assessment approach for scenario-based virtual testing of automated driving functions. *IEEE Open Journal of Intelligent Transportation Systems*, *3*, 45–60.

Storsæter, A. D. (2021). *Designing and maintaining roads to facilitate automated driving.* [PhD thesis]. NTNU. Retrieved from https://hdl.handle.net/11250/2767106

Synopsys. (2022). *What is an autonomous car?* Retrieved April 4, 2022, from https://www.synopsys.com/automotive/what-is-autonomous-car.html

Tabone, W., De Winter, J., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., ... Hancock, P. A. (2021). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. *Transportation research interdisciplinary perspectives*, *9*, 100293. https://doi.org/10.1016/j.trip.2020.100293.

Theeuwes, J., & Hagenzieker, M. P. (1993). Visual search of traffic scenes: On the effect of location expectations. *Vision in vehicles*, *4*, 149–158.

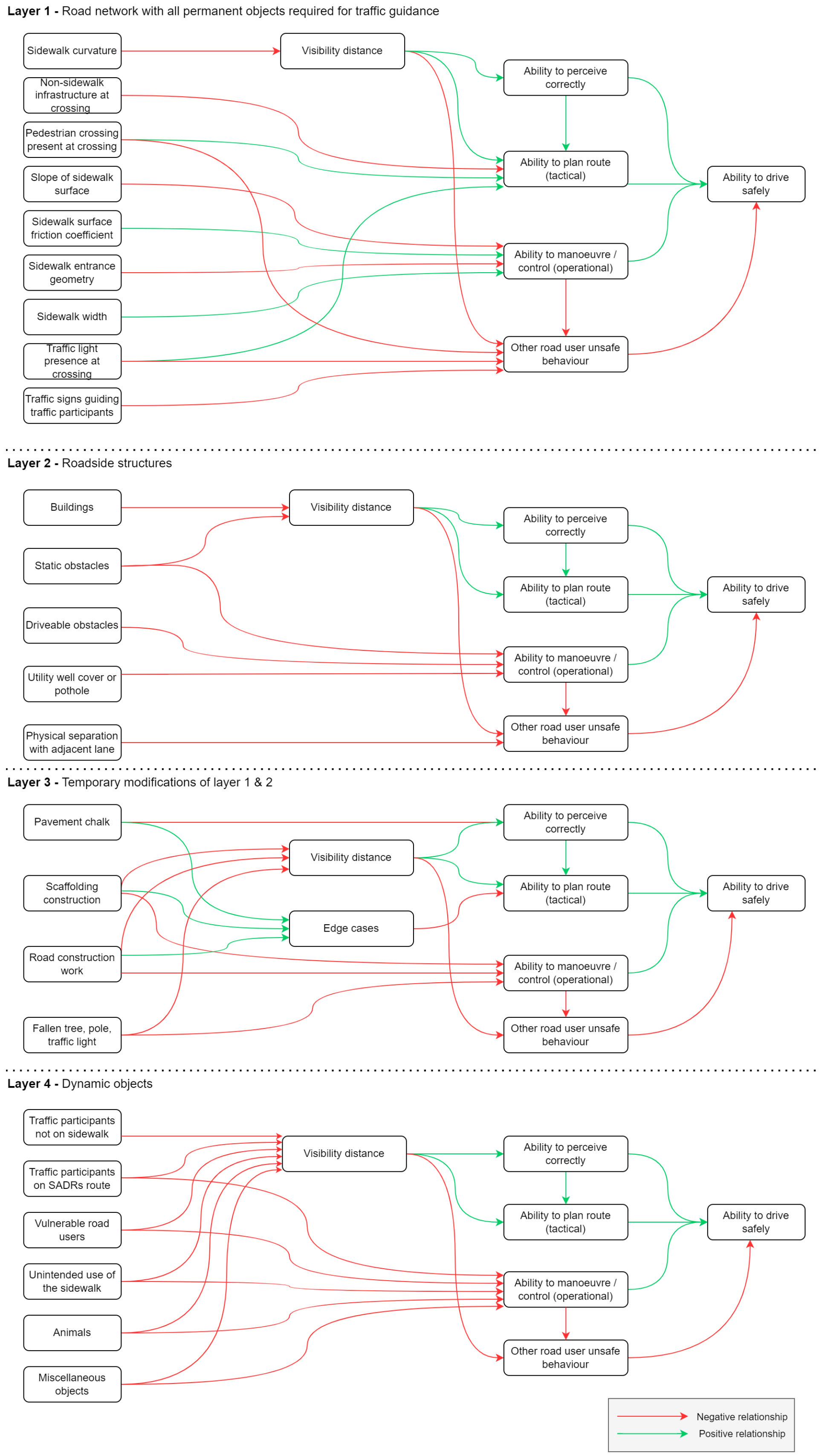
Thorn, E., Kimmel, S. C., Chaka, M., & Hamilton, B. A. (2018). *A framework for automated driving system testable cases and scenarios.* (DOT HS 812 623). United States. Department of Transportation. National Highway Traffic Safety Administration.

Vaishnavi, V., & Kuechler, W. (2004). *Design research in information systems.* Retrieved from http://www.desrist.org/design-research-in-information-systems/

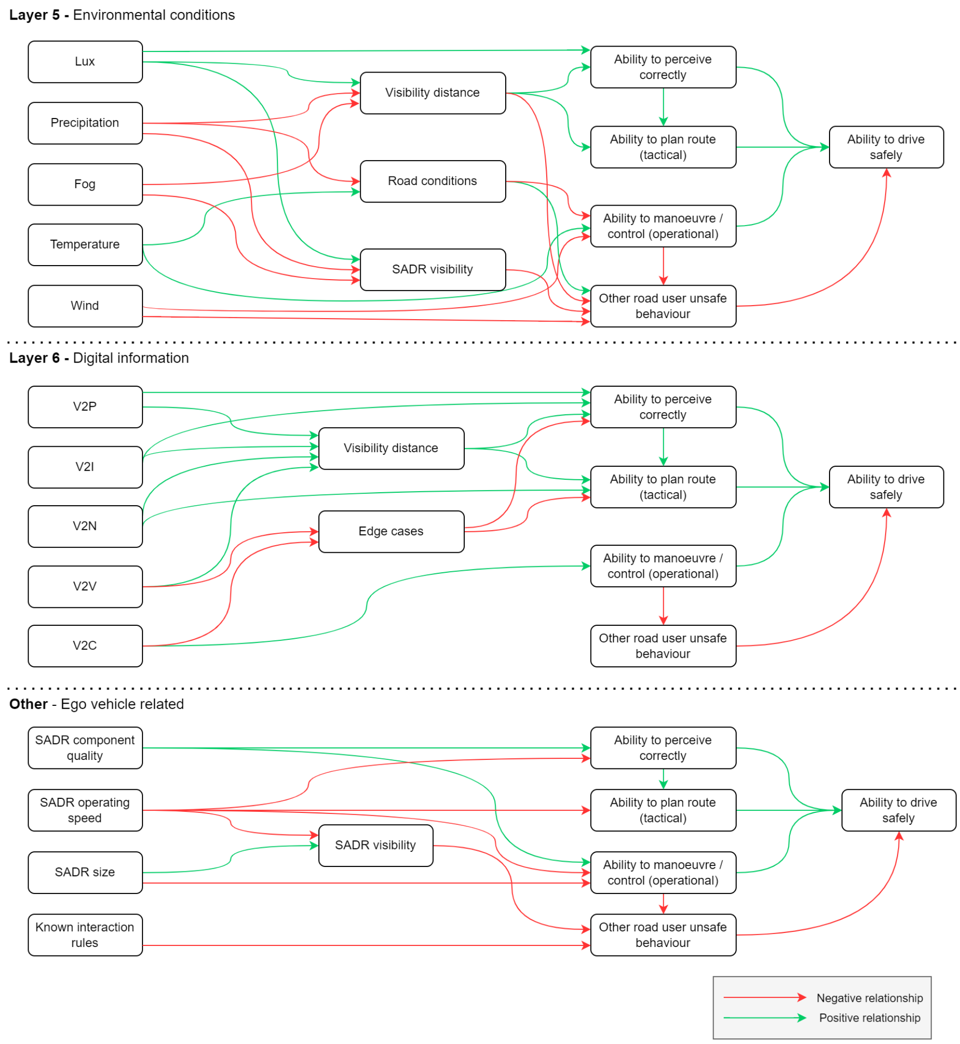
van der Stoep, P. (2022, February 24). *RDW Applied innovation.* [PowerPoint slides]. RDW.

Vleeshouwer, T., Rotterdam, H., & Verbraeck, A. (2017). Implementatie van autonome bezorgrobots voor een kleinschalige thuisbezorgdienst. *Vervoerslogistieke Werkdagen 2017*.

Wishart, J., Como, S., Elli, M., Russo, B., Weast, J., Altekar, N., ... Chen, Y. (2020). Driving safety performance assessment metrics for ads-equipped vehicles. *SAE Technical Paper*, *2*(2020-01-1206). https://doi.org/10.4271/2020-01-1206.



(a) SADR performance risks guided from the 6-Layer Model: layers 1, 2, 3 and 4



(b) SADR performance risks guided from the 6-Layer Model: layers 5, 6 and extended with 'other'

FIGURE 7: SADR performance risks guided from the 6-Layer Model

B ODD literature study journal ranks

In this appendix the final publications used in the literature study on structured ODD descriptive methods are presented, with the number of citations, SCImago Journal Rank and impact score per publication, if possible.

Table 13: Number of citations, SCImago Journal Rank and Impact Score for ODD classification literature study publications

Publication	**Citations**	**SCImago Journal Rank (SJR)**	**Impact Score**	**Comment**
Koopman & Fratrik (2019)	53			Conference proceedings
Riedmaier et al. (2020)	80	0.587	4.48	
Roh & Im (2020)	5	0.612	3.48	
Griffor et al. (2021)	1	0.202	1.33	
Geyer et al. (2014)	126	0.579	3.21	
Czarnecki (2018a)	38			Waterloo Intelligent Systems Engineering Lab publication
Czarnecki (2018b)	18			Waterloo Intelligent Systems Engineering Lab publication
Czarnecki (2018c)	12			Waterloo Intelligent Systems Engineering Lab publication
Cho (2020)	0			PhD thesis at Massachusetts Institute of Technology
Scholtes et al. (2021)	23	0.587	4.48	
Bagschik et al. (2018)	182	0.241	2.91	Conference proceedings
Bock et al. (2018)	16			Colloquium
Schuldt (2017)	63			PhD thesis at University of Braunschweig
Khatun et al. (2021)	2			Conference proceedings

C ADS performance risk factor literature study journal ranks

In this appendix the final publications used in the literature study on ADS risks and risk factors are presented, with the number of citations, SCImago Journal Rank and impact score per publication, if possible.

Table 14: Number of citations, SCImago Journal Rank and Impact Score for ADS risk factors literature study publications

Publication	Citations	SCImago Journal Rank (SJR)	Impact Score	Comment
Roh & Im (2020)	5	0.612	3.48	
Czarnecki (2018a)	38			Waterloo Intelligent Systems Engineering Lab publication
Czarnecki (2018b)	18			Waterloo Intelligent Systems Engineering Lab publication
Boggs et al. (2020)	39	1.816	5.55	
Feng et al. (2020)	0			3rd World Conference on Mechanical Engineering and Intelligent Manufacturing
Storsæter (2021)	0			Doctoral Thesis at Norwegian University of Science and Technology
Mihalj et al. (2022)	0	0.435	3.02	
Ren et al. (2019)	4	0.108	0.21	2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics
Leroy et al. (2020)	1	0.308	1.13	
Thorn et al. (2018)	113			National Highway Traffic Safety Administration publication
Rao et al. (2021)	0			National Highway Traffic Safety Administration publication
Hillman & Capaldi (2020)	0			Findings from HumanDrive project
Sivak & Schoettle (2015)	109			Tech report
Farah et al. (2018)	50			Book based on Automated Vehicles Symposium 2016
Schwall et al. (2020)	27			arXiv publication, arXiv is maintained by Cornell University
Theeuwes & Hagenzieker (1993)	78			Older publication, no ranks available for journal
Pokorny et al. (2021)	0	0.741	2.98	
Tabone et al. (2021)	34	0.383	1.78	
Heikoop et al. (2020)	10			No ranks available for journal
M. Hagenzieker et al. (2020)	3			TU Delft publication
M. P. Hagenzieker et al. (2020)	44	0.504	2.02	

D Ontology Czarnecki

In this appendix an overview of the ontology elements established by Czarnecki (2018b,c) is presented in full detail.

Road structure - Part 1 (Czarnecki, 2018b)

Road structure

- Road type and capacity
  - General road classification
  - Road capacity
  - Road classification criteria
  - Road location by zone
- Road surface type and quality
  - Road surface type
  - Road surface friction
  - Road surface roughness
  - Road surface damage
- Road geometry
  - Horizontal alignment
    - Circular curves
    - Spiral curves
  - Vertical alignment
- Cross section design
  - Lane structure
    - Lane types
    - Lane widths
  - Roadside structure
- Road traffic control devices
  - Traffic signs
    - Regulatory traffic signs
    - Warning traffic signs
    - Guide and information signs
  - Traffic signals
  - Roadway pavement markings
  - Vertical deflections
- Pedestrian crossing facilities
- Cycling facilities
- Junctions
  - Intersections
    - Intersection forms
    - Intersection maneuvers and traffic conflicts
    - Intersection designs
    - Intersection traffic control
  - Interchanges
- Railroad level crossings
- Bridges
- Tunnels
- Driveways and driver access points
- Temporary road structure

Figure 37: Road structure elements by Czarnecki (2018b)

Road users, animals, other obstacles and environmental conditions - Part 2 (Czarnecki, 2018c)

Road users

- Road user classification
  - Ground vehicles
    - Road vehicles
      - Motor vehicles
        - Passenger cars
        - Trucks
        - Busses
        - Motorcycles
        - Small and low-speed vehicles
        - Emergency vehicles
      - Pedalcycles
      - Trailers
    - Off road vehicles
      - Motorized off-road vehicles
      - Animal-drawn vehicles
    - Railed vehicles
    - Rail-road vehicles
  - Animal riders
  - Pedestrians
  - Traffic control persons

Road user behavior

- Behavioral factors
  - Traffic rules
    - Traffic laws
    - Informal best practice rules
  - Social norms
  - Individual behavior
- Behavioral models
  - Vehicle behavior models
    - Vehicle models
    - Driver behavior models
      - Human driver behavior models
        - Operational behavior
        - Tactical behavior
      - Automated driving system models
  - Pedestrian models

Animals

Other obstacles

Environmental conditions

- Atmospheric conditions
  - Visibility
  - Wind
  - Cloud conditions
  - Precipitation
  - Atmospheric obscuration
- Lighting conditions
- Road surface conditions

Figure 38: Road structure elements by Czarnecki (2018c)

E SADR of focus specifications

In this appendix an overview of the technical characteristics of the SADR of focus in this study, named Rosie, is presented. Rosie is manufactured by Cartken Inc. and illustrated in Figure 39. In Table 15 it is summarized by what technical characteristics Rosie is described. Table 16 summarizes the ODD characteristics in which the manufacturer of Rosie claims the SADR can operate safely.

For reasons of confidentiality contents have been removed from this page.

Figure 39: SADR of focus: Rosie (photo by Jonathan van Rijn)

Table 15: Known SADR (Rosie) Characteristics

SADR characteristic	**Specification**
Length	71 cm
Width	46 cm
Height	60 cm
Weight	40 kg

Table 16: Known safe ODD characteristics

Known safe ODD characteristics	**Specification**
Precipitation conditions	Light rain and snow
Temperature conditions	Temperature between 4 and 40 degrees Celsius
Light conditions	Dark and light
Dynamic objects	Traffic participants that are moving up to 15 km/h

F SADR risk explanation

In this appendix the complete risk overview visualised in Figure 18a, 18b & 19 will be discussed.
The visualisations are all focused on the ability of an SADR to drive safely. This variable is directly dependent on three variables that reflect the core functions an SADR has to be able to perform: to perceive correctly, to plan a route (tactically) and to actually manoeuvre. The fourth variable effecting the ability of an SADR to drive safely is other road user unsafe behaviour. This variable represents intended and unintended unsafe behaviour by other road users, e.g. because they make mistakes in traffic, violate traffic rules or, in theory, make unauthorised manoeuvres to avoid mistakes by third parties. If an SADR can perceive with 100% certainty this will positively effect the ability to plan a route, because an SADR will have the most complete overview of a traffic situation. If an SADR is able to manoeuvre well, this will decrease the unsafe behaviour of other road users because the SADR will not make safety stops or come to a halt unintentionally, where on its turn other road users have to cope with.

F.1 Layer 1 risk factors

An increase in sidewalk curvature reduces the visibility distance for an SADR. A reduced visibility distance results in an SADR having less distance to observe correctly and less distance to plan a route. Finally, a reduced visibility distance also applies to other road users, because they also have less distance to make a choice in traffic and this leads to an increase in other road user unsafe behaviour. More difficult non-sidewalk infrastructure at crossing points makes it more difficult for an SADR to plan a route. Within this factor it is about how many lanes and for what kind of mobility those lanes are intended. The dynamic traffic participants, if present, are discussed in Layer 4. The presence of a pedestrian crossing at an intersection makes it easier for an SADR to plan a route because it can be assumed that there are traffic rules in force. In addition, other road users are familiar with pedestrian crossings and will show safer behaviour than in the absence of a regulated pedestrian crossing. An increased slope of the sidewalk complicates an SADR's manoeuvrability. From APPENDIX D it follows that the robot central to this study can handle slopes of up to 20The higher the sidewalk surface friction coefficient the better an SADR can manoeuvre. Driving on an asphalt strip is easier than driving in loose sand. The sidewalk entrance geometry affects how easily an SADR is able to drive up the sidewalk. A gradual slope is easier than a straight curb with a specific height. When the curb is too high, it can even prevent an SADR from entering the sidewalk at all. A wider sidewalk will ensure that an SADR has more opportunity to plan a route and also to execute it operationally. The presence of traffic lights at intersections ensures that an SADR is guided in planning a route and that other road users are guided in their behaviour. The presence of traffic signs also steers the behaviour of other road users and informs them about possible upcoming traffic situations and the applicable traffic rules.

F.2 Layer 2 risk factors

The presence of buildings next to a sidewalk reduces the visibility distance for an SADR. The same applies to static obstacles on the sidewalk, and these obstacles also reduce the manoeuvrability of an SADR because they reduce the effective width of the sidewalk. Driveable obstacles, such as speed bumps, utility well covers and potholes, make the sidewalk less passable and reduce the manoeuvrability of an SADR. If a physical separation with adjacent carriageways is present, the chance that other road users will intentionally or unintentionally enter the sidewalk is reduced.

F.3 Layer 3 risk factors

The presence of pavement chalk or other forms of art on the sidewalk can cause the image recognition of an SADR camera system to identify objects that are not actually there. A RADAR or Lidar system will not perceive these objects either, which disturbs the ability to perceive correctly. The presence of a scaffolding construction or a road construction work section will, firstly, reduce the visibility distance because of its size and, secondly, increase the number of edge cases because recognition points from the surroundings can no longer be detected (at least temporarily). Thirdly, such work will reduce the ability to manoeuvre of an SADR, which may be caused by a degradation in road quality, a reduction in the available infrastructure, blockages or the temporary positioning of static and driveable obstacles. Finally, fallen trees and bollards can reduce the visibility distance (depending on their size) and also affect the manoeuvrability.

F.4 Layer 4 risk factors

Although in the overview all dynamic objects have been separated and described per category, it can be seen from the figure that all dynamic objects, by their presence, influence the visibility distance and the extent to which an SADR can move.

F.5 Layer 5 risk factors

In bright conditions, the image recognition of an SADR camera system can detect better. There is no effect of light on the RADAR or LIDAR systems. In addition, on a clear day the visibility distance is greater, and other road users can perceive an SADR better, which results in safe driving behaviour. Precipitation on the other hand reduces the visibility distance, worsens the road conditions and makes an SADR less visible. The worsened road conditions make it more difficult for an SADR to move and for other road users to move as well. In itself, fog has no direct effect on road conditions, but it does greatly reduce the visibility distance and the visibility of the SADR, which (unintentionally) increases the risk of unsafe situations. From Appendix D it follows that the robot at the centre of this study can operate between temperatures of 4 and 40 degrees Celsius. Temperatures outside this range, lower than 0 degrees for example, worsen the road conditions on the one hand, and on the other hand the documentation shows that a robot can no longer manoeuvre properly. Wind makes it more difficult for the robot to follow its path and sometimes it even has to correct itself. This effect is also felt by other road users, which can lead to unsafe situations.

F.6 Layer 6 risk factors

The digital communication solutions described in the sixth layer all improve the performance of the robot. When pedestrians digitally communicate signals about their location and surroundings, this leads to an increased visibility distance because they can be seen outside the range of the camera and radar systems of an SADR. In addition, it becomes easier for a robot to perceive correctly because an SADR can compare the information from camera and RADAR systems with the digital signals it receives. The same applies to infrastructural features and signals. If the state of traffic lights is shared with a robot, it can again perceive outside the range of its own resources and its ability to perceive correctly increases. Vehicle to Network solutions are all online communication means that an SADR can use to gather data from external information sources and for example use external systems to plan a route. This increases the visibility distance and the ability to plan a route. Communication with other vehicles also leads to an increase in visibility distance. In addition, precise information about the position, speed and intended path of a vehicle reduces the number of edge cases. The Vehicle to Control room communication reduces the number of edge cases because a teleoperator can look along and make a safe decision about the action to be performed by the robot. In addition, a teleoperator can ensure that the robot is still controlled if the control software fails in some way.

F.7 SADR vehicle risk factors

When the quality of the components and systems that make up an SADR is better, the ability to perceive correctly and the manoeuvrability of an SADR increase. A higher operating speed of an SADR means that an SADR has to perceive in less time, has less time to plan a route and is less manoeuvrable. In addition, a higher speed can result in other road users not noticing the SADR because they were paying attention to other features in the environment, resulting in safety risks. The size of an SADR on the other hand increases this visibility for other road users, but makes an SADR less manoeuvrable. Finally, the development of complete traffic rules for SADRs and for the interaction with SADRs can reduce the (unintended) unsafe driving behaviour of other road users.

G Regression analysis

This appendix describes how a regression analysis can contribute to actually objectively determining the ODD difficulty of a geographical location on the basis of risk factors identified using the 6-Layer-Model. To describe the concept of a regression analysis, linear relationships between the risk factors and the ODD difficulty score are assumed. At the end of this appendix, the justification for the spatial axis, merging Layers 1, 2 and 3, as mentioned in Section 6.5.1 will be discussed.

G.1 Theory on regression analysis

Regression analysis is used to examine the effects of predictor variables on the dependent variable. For example, the dependent variable X depends on the predictor variables A, B and C (Figure 40).

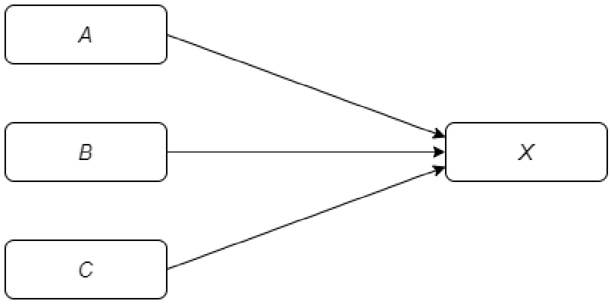


Figure 40: Regression analysis

The standardized equation to determine the value for the dependent variable based on the predictor variables is denoted by:

$$Z_X = \beta_A * Z_A + \beta_B * Z_B + \beta_C * Z_C + \varepsilon \tag{8}$$

By taking the standardized Beta coefficients for the predictor variables, the impact of the variables that are measured on different scales can be compared to each other. The standardized coefficients can be interpreted as weights: a larger Beta indicates a better predictor variable (Molin, 2019). ε represents a random error component to account for unobserved heterogeneity in the measurements.

G.2 Regression analysis applied to ODD of SADRs

To illustrate what a regression analysis would look like to determine the difficulty of different geographical locations at the level of manoeuvrability, an example is given. In Table 17 the variable notation for the spatial risk factors are arranged by layer.

Table 17: Overview of Layer 1, 2 and 3 risk factor variables for regression

Layer	Risk Factor	Variable Notation
1	Sidewalk curvature	f_{11}
	Non-sidewalk infrastructure at crossing	f_{12}
	Pedestrian crossing presence at crossing	f_{13}
	Slope of sidewalk surface	f_{14}
	Sidewalk surface friction coefficient	f_{15}
	Sidewalk entrance geometry	f_{16}
	Sidewalk width	f_{17}
	Traffic light presence at crossing	f_{18}
	Traffic signs guiding traffic participants	f_{19}
2	Sidewalk Minimum Clear Width Accounting for Static Obstacles	f_{21}
	Sidewalk Second Minimum Clear Width Accounting for Static Obstacles	f_{22}
	Sidewalk Third Minimum Clear Width Accounting for Static Obstacles	f_{23}
	Sidewalk Fourth Minimum Clear Width Accounting for Static Obstacles	f_{24}
	Driveable obstacles	f_{25}
	Utility well cover or pothole	f_{26}
	Physical separation with adjacent lane	f_{27}
3	Temporary modification	f_{31}
	Pavement chalk	f_{32}

For the impact of infrastructure elements by means of experiment can be assessed what the effect of spatial risk factors is on the ability of an SADR to cope with infrastructural characteristics. For the risk factors from Layer 1 direct relationships between the risk factors and the manoeuvrability are assumed. As discussed in Section 6.5.7 it is more difficult to determine the actual impact of obstacles on manoeuvrability because the extent to which an obstacle hinders the SADR in its movement depends on the context in which the obstacle is placed. Therefore surrogate measures are necessary, such as the Sidewalk Minimum Clear Width Accounting for Static Obstacles. Temporary modifications to the operating environment on the one hand change the values for the parameters in Layer 1 and Layer 2, on the other hand has to be accounted for the fact that operating environment differs from what was known beforehand, which is why the 'Temporary modification' risk factor is included in Table 17.

$$Z_{Spatial} = \textit{Component Layer 1} + \textit{Component Layer 2} + \textit{Component Layer 3} + \varepsilon \tag{9}$$

$$\begin{aligned}\textit{Component Layer 1} = {} & \beta_{11}Z_{f_{11}} + \beta_{12}Z_{f_{12}} + \beta_{13}Z_{f_{13}} + \beta_{14}Z_{f_{14}} + \beta_{15}Z_{f_{15}} + \beta_{16}Z_{f_{16}} \\ & + \beta_{17}Z_{f_{17}} + \beta_{18}Z_{f_{18}} + \beta_{19}Z_{f_{19}}\end{aligned} \tag{10}$$

$$\textit{Component Layer 2} = \beta_{21}Z_{f_{21}} + \beta_{22}Z_{f_{22}} + \beta_{23}Z_{f_{23}} + \beta_{24}Z_{f_{24}} + \beta_{25}Z_{f_{25}} + \beta_{26}Z_{f_{26}} + \beta_{27}Z_{f_{27}} \tag{11}$$

$$\textit{Component Layer 3} = \beta_{31}Z_{f_{31}} + \beta_{32}Z_{f_{32}} \tag{12}$$

$$\begin{aligned}Z_{Spatial} = {} & \beta_{11}Z_{f_{11}} + \beta_{12}Z_{f_{12}} + \beta_{13}Z_{f_{13}} + \beta_{14}Z_{f_{14}} + \beta_{15}Z_{f_{15}} + \beta_{16}Z_{f_{16}} + \beta_{17}Z_{f_{17}} + \beta_{18}Z_{f_{18}} \\ & + \beta_{19}Z_{f_{19}} + \beta_{21}Z_{f_{21}} + \beta_{22}Z_{f_{22}} + \beta_{23}Z_{f_{23}} + \beta_{24}Z_{f_{24}} + \beta_{25}Z_{f_{25}} + \beta_{26}Z_{f_{26}} + \beta_{27}Z_{f_{27}} \\ & + \beta_{31}Z_{f_{31}} + \beta_{32}Z_{f_{32}} + \varepsilon\end{aligned} \tag{13}$$

For the ODD difficulty resulting from the dynamic objects present, the risk factor variables in Table 18 can be used for regression. The corresponding equation is given with Equation 14.

Table 18: Overview of Layer 4 risk factor variables for regression

Layer	Risk Factor	Variable Notation
4	Sidewalk Minimum Clear Width Accounting for Static Obstacles	f_{41}
	Average busyness by VRUs	f_{42}
	Total occupied sidewalk area	f_{43}

$$Z_{Dynamic} = \beta_{41}Z_{f_{41}} + \beta_{42}Z_{f_{42}} + \beta_{43}Z_{f_{43}} + \varepsilon \tag{14}$$

The environmental difficulty can be determined on the risk factor variables in Table 19 and corresponding Equation 15

Table 19: Overview of Layer 5 risk factor variables for regression

Layer	Risk Factor	Variable Notation
5	Lux	f_{51}
	Precipitation	f_{52}
	Fog	f_{53}
	Temperature	f_{54}
	Wind	f_{55}

$$Z_{Environment} = \beta_{51}Z_{f_{51}} + \beta_{52}Z_{f_{52}} + \beta_{53}Z_{f_{53}} + \beta_{54}Z_{f_{54}} + \beta_{55}Z_{f_{55}} + \varepsilon \tag{15}$$

As can be noticed from the Equations 9 to 15, the possibility that digital means will make the ODD level easier has not been taken into account. This is a simplification of the reality and should be included in a follow-up study.

G.3 ODD constraints

The representation of the formulas implies that the different location characteristics can compensate each other. For example, a high parameter value for f_{11} can be compensated by a low parameter value for f_{14}. However, there are a number of hard system constraints which ensure that, no matter how favourable the other system conditions are, a sidewalk segment is by definition not passable for an SADR. The currently known system constraints are listed in Table 20 with their associated values.

Table 20: Overview of system driveablity constraints

Layer	Risk Factor	Variable Notation
1	Slope of sidewalk surface[2]	>20°
	Sidewalk width	<46 cm
2	Sidewalk Minimum Clear Width Accounting for Static Obstacles	<46 cm
5	Temperature	<4° Celsius >40° Celsius

Additional system limits can be found by running SADRs under many different conditions. Furthermore, the currently mentioned system boundaries may shift due to modifications to SADRs.

G.4 Justification spatial axis

Layer 1 describes the road (sidewalk) network elements with all permanent objects needed for traffic guidance. Layer 2 describes all roadside structures on and next to the sidewalk, but which are not necessarily needed for traffic guidance. The consequence of modifications to Layer 1 and 2, that are thus described in Layer 3, is that edge cases arise. These are moments when an SADR, for example, has difficulty finding landmarks in the environment that the robot normally uses to verify its geographical position. Also, temporary modifications could cause an SADR to have more difficulty in manoeuvring due to the presence of structures caused by work activities, due to the effective width of the sidewalk being reduced because the sidewalk is broken up for the work activities in question, or due to the fact that instead of a stone sidewalk, there is now a temporary dirt path which makes it difficult for the robot to manoeuvre the sidewalk.

There are two reasons why it is justified to take these three layers together in the spatial axis. The first is that the elements in Layer 3 are of exactly the same type as Layer 1 and 2, the only difference being that their presence is temporary. Layer 1 and 2 can be merged because the roadside structures that apply to ADS are primarily adjacent to the road and therefore on the sidewalk. For SADRs, although these elements do not contribute to guidance, they do affect an SADR's ability to move forward. As a result, theoretically, all of the elements that an SADR faces from an infrastructural perspective are captured in the first three layers of the 6-Layer Model. The second reason why these three layers can be merged is that although the impact of elements in the different layers may differ, this will follow from the regression analysis. Merging the three layers does not necessarily mean lumping all factors together and estimating their impact as equal.

G.5 Discussion

The traffic system is a complex environment. It is not likely that the relative ODD difficulty depends only on linear relationships with the identified risk factors. Some relationships may be non-linear and it is also likely that (some of the) predictor variables correlate with each other. The latter often happens with a multiple of predictor variables, which is known as multicollinearity (Kroesen, 2021). Although science has found several solutions for this problem, it will not be discussed further in this study. As mentioned before, an important follow-up research after this study is to carry out a complete study into the actual effects of the spatial factors, dynamic objects and environmental conditions on ODD difficulty. Due to a lack of public data, among other things, there is currently no scientific ground for taking non-linear and correlating variables as hypotheses. Estimating different models after data collection in a follow-up study could further investigate this matter.

[2]Note that the slope of the sidewalk depends on the direction of travel of the robot.

H Performance Metrics for SADRs

In this appendix an overview of the driving safety performance metrics for ADS-equipped vehicles by Wishart et al. (2020) will be given (see Table 21) and will be motivated which metrics are not deemed relevant during an initial field test with the dynamic assessment framework.

Table 21: Performance Metrics by Wishart et al. (2020)

Performance metric	Abbreviation	Useful for dynamic assessment framework
Minimum Safe Distance Violation	MSDV	✓
Proper Response Action	PRA	✓
Minimum Safe Distance Factor	MSDF	No, derived from MSDV
Minimum Safe Distance Calculation Error	MSDCE	✓
Collision Incident	CI	✓
Rules-of-the-road Violation	RRV	✓
Achieved Behavioral Competency	ABC	No, designed to be used in specific testing
ADS Active	ADSA	✓
Human Traffic Control Detection Error Rate	HTCDER	No, as long as teleoperator can help not necessary
Human Traffic Control Violation Rate	HTCVR	No, as long as teleoperator can help not necessary
Time-to-Collision	TTC	No, MTTC is a more robust metric
Modified Time-to-Collision	MTTC	✓
Aggressive Driving	AD	No, operating speed of SADR too low

The MSDF metric is directly derived from the MSDV metric and will therefore not provide additional insight in the SADR performance. The ABC metric is designed to evaluate if a vehicle executed a specific behavioral competency in a testing scenario. The intended field test resulting from this research does not suit the testing scenario. The HTCDER and HTCVR metrics are both established to evaluate if a vehicle is capable of detecting and executing (illegal) manoeuvres according to the instructions of a traffic controller. Both metrics are currently not (yet) relevant because a situation with a traffic controller in a sidewalk environment will most likely lead to an edge case, in which the help of a teleoperator will be called upon to achieve a safe manoeuvre. Both metrics are too advanced for what we now know of an SADR and what we expect an SADR to be able to do. The TTC will not provide additional insight in the SADR performance because the MTTC is a more robust metric. Note that the only difference between the two metrics is that the MTTC does not rely on the assumption that traffic participants will maintain constant speeds.

I Semi-structured interview approach

In this appendix an overview is given of the interview questions that have been used to guide the semi-structured expert interviews in the validation step of the research process. Next to a few background questions, the questions are divided in a subset of questions that focus on the process that led to the determination of different levels of difficulty within the Operational Design Domain and a subset of questions that focuses on the design suggestion based on these different levels of difficulty within the Operational Design Domain.

I.1 Operational Design Domain difficulty

At the beginning of the interview, the purpose of the validation is explained, terms of confidentiality are addressed, the format of the semi-structured interview is explained and the interviewee is asked if she is comfortable with recording the audio of the conversation.

Question 1: "Are you familiar with delivery robots driving on the pavement? If so, in what way are you familiar with them?"

Question 2: "Have you had the opportunity to watch the introductory video I sent you? Did you have any questions about it? Were there any parts that you would like to elaborate on later in the interview?"
Topics that come up at that time are noted

Interviewees are reminded of the arguments in the introductory video that sum up why delivery robots cannot currently be tested on the sidewalk, namely: that robots currently cannot be classified in an established vehicle category, that safe kilometres in foreign countries are no argument to also expect safe kilometres in the Netherlands because the circumstances of those safe kilometres cannot be ascertained, and that there is currently no unequivocal way to test the black-box operating systems of self-driving delivery robots for safety.

Question 3: "Do you recognise the barriers identified?"

Question 4: "Do you think there are any other barriers preventing delivery robots from being tested on public sidewalks in the Netherlands?"

Interviewees is explained that the study found that there is no uniform way to identify the ODD of delivery robots and that therefore there is no agreement among stakeholders as to what are the actual risks of using a delivery robot. In this study, the concept of the Operational Design Domain is not approached as purely binary, but it is hypothesized that within the Operational Design Domain there are different degrees of difficulty because there are situations that a delivery robot can handle more easily and less easily.

Question 5: "What do you think about the theory/hypothesis that there are different levels of difficulty within the Operational Design Domain?"

Interviewees is explained that in order to identify the risks associated with the deployment of delivery robots, a literature review was conducted on the safety risks associated with Autonomous Vehicles and its components. It is explained that the hardware and software components used in both innovations are to some extent similar. It is explained that this line of reasoning was followed because there is no literature on the risks of delivery robots.

Question 6: "Do you agree that risks that apply to AVs could also apply to delivery robots? Do you have any additional reasons than the reason explained?"

It is explained to interviewees that, based on a literature review, it was decided to identify all risk factors in the ODD of a delivery robot using the 6-Layer Model. A link was then made with the risks identified for Autonomous Vehicles and these were translated to the operating environment of delivery robots. The conceptual overview with risk factors per layer of the 6-Layer Model is shown and explained further.

Question 7: "What is your impression of these risk factors and the way they are represented?"

Following on from the risk factors for delivery robots and their causal relationship, the interviewees are shown an overview of how the individual risk factors can be quantified and qualified one by one, in order to then be able to compare geographical locations on measured factual characteristics.

Question 8: "Do you think that mapping and measuring the risk factors in such a way can contribute to a more standardised description of the Operational Design Domain?"

I.2 Dynamic assessment framework

It is explained that based on the relative levels of difficulty, in line with the intended outcome of the Design Science Research cycle, a digital dynamic assessment framework for delivery robots on the public sidewalk has been developed that functions from a risk minimisation perspective. The introductory video is quoted to once again name the functionalities of the dynamic assessment framework, which include: the network link difficulty explanation, incident reporting and the geographic location comparability means. The intended living lab approach is also discussed.

Question 9: "What is your first impression of the dynamic assessment framework designed?"

Question 10: "Do you miss functionalities in the proposed dynamic assessment framework? Are there elements not-included that you had expected to be included in the dynamic assessment framework?"

Question 11: "Do you foresee any difficulties in developing this dynamic assessment framework?"

Question 12: "Do you think that the actual development of the proposed dynamic assessment framework could contribute to allowing public sidewalk tests of delivery robots in the Netherlands?"

After the last interview question is discussed and is made sure that interviewees do not have further questions, the interview is concluded.