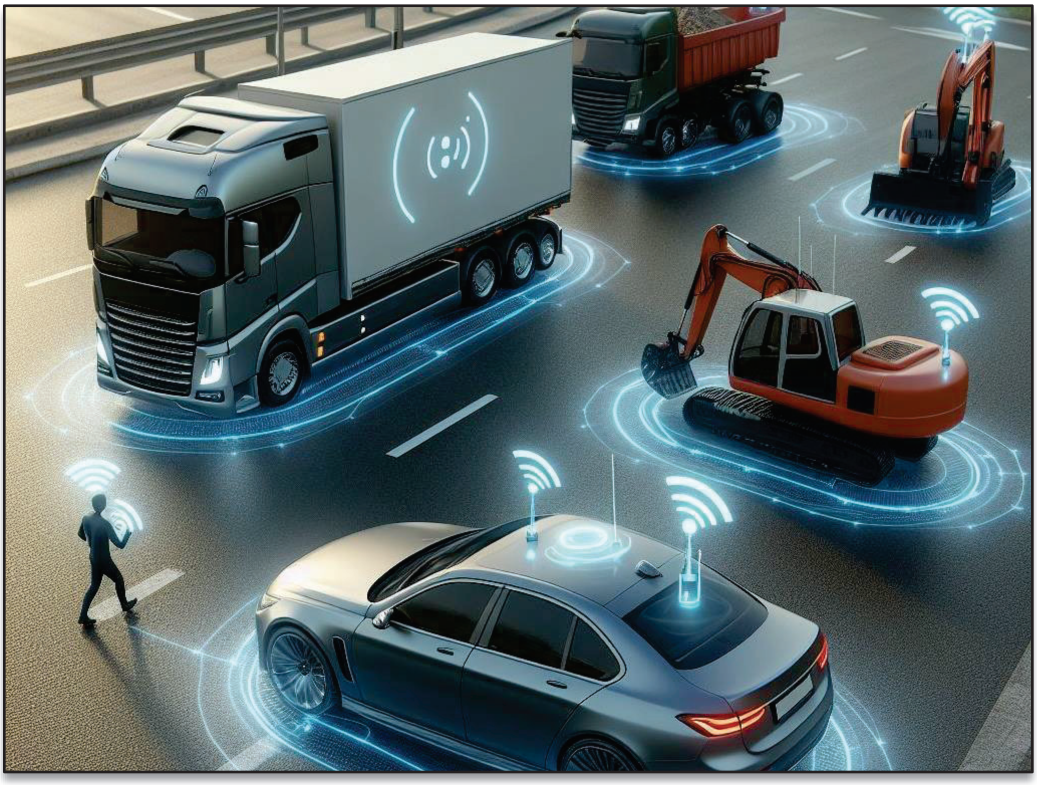
SALIENCE4CAV

Safety Lifecycle Enabling Continuous Deployment for Connected Automated Vehicles

Public report



Project within	Trafiksäkerhet och automatiserade fordon (Road safety and automated vehicles)
Author	Fredrik Warg, Anders Thorsén, DeJiu Chen, Jens Henriksson, Gabriel Rodrigues de Campos
Date	2023-01-29

FFI Fordonsstrategisk Forskning och Innovation

Content

FFI in short

FFI is a partnership between the Swedish government and automotive industry for joint funding of research, innovation and development concentrating on Climate & Environment and Safety. FFI has R&D activities worth approx. €100 million per year, of which about €40 is governmental funding.

For more information: www.vinnova.se/ffi

Cover picture generated by Bing/DALL•E 3

1 Summary

Connected automated vehicles (CAVs) are—compared conventional vehicles—expected to provide more efficient, accessible, and safer transport solutions in on-road use cases as well as confined areas such as mines, construction sites or harbours. As development of such vehicles has proved more difficult than anticipated, especially when it comes to ensuring safety, more cautious strategies for introduction are now being pursued. An approach where new automated features are initially released with more basic performance to enable successful safety assurance, followed by gradual expansion of performance and number of use-cases using an iterative development process as the confidence in the solution increases, e.g., due to more available field data, improved machine learning algorithms, or improved verification, is highly interesting. Hence a key research question targeted by the SALIENCE4CAV project was: *How to ensure the safety of CAVs while enabling frequent updates for automated driving systems with their comprising elements?* Today, many of the used methods and practices for safety analysis and safety assurance are not adequate for continuous deployment. In addition, the project has investigated several open questions raised by the predecessor project ESPLANADE and from needs identified by the industry partners; this includes how to handle safety assurance for machine learning components, use of quantitative risk acceptance criteria as a key part of the safety argument, safety for collaborative CAVs including use in mixed traffic environments, the role of minimal risk manoeuvres, and interaction with human operators.

Some key results are: investigation of safety assurance methods and gaps with regards to frequent updates and other challenges for CAV safety assurance; use of safety contracts as an enabler for continuous integration, continuous deployment and DevOps; a method for human interaction safety analysis; application of the principle of precautionary safety for meeting a quantitative risk norm and using field data for continuous improvements; definition of classes of cooperative and collaborative vehicles and their respective characteristics and definition of minimal risk manoeuvre and minimal risk condition strategies for individual, cooperative and collaborative vehicles; use of out-of-distribution detection for safety of machine learning; a simulation-aided approach for evaluating machine learning components; and methods for variational safety using high-dimensional safety contracts.

The SALIENCE4CAV project ran from January 2021 to December 2023 with the partners Agreat, Comentor, Epiroc Rock Drills, KTH Royal Institute of Technology, Qamcom Research and Technology, RISE Research Institutes of Sweden, Semcon Sweden, Veoneer (during the project acquired by Magna) and Zenseact. Coordination was done by RISE.

This final report is a summary of the project results and contains summaries of content from the project deliverables and publications.

2 Sammanfattning på svenska

Uppkopplade och automatiserade fordon förväntas ge transportlösningar som är mer effektiva, tillgängliga och säkrare jämfört med konventionella fordon, både i användningsfall på vägar och i avgränsade områden som gruvor, byggarbetsplatser eller hamnar. Att utveckla automatiserade fordon har dock visat sig vara svårare än förväntat, särskilt när det gäller säkerhetsaspekten. Därför har många tillverkare börjat utforska strategier för stegvis introduktion av automatisering, så att nya funktioner först släpps med enklare prestanda för att möjliggöra säkerhetsbevisningen, följt av gradvis utökning av prestanda och antal användningsfall med hjälp av en iterativ utvecklingsprocess. Sådana utökningar kan ske allteftersom förtroendet för produktens säkerhet ökar genom tillgång till mer fältdata, bättre prestanda för maskininlärnings-komponenter, bättre verifiering o.s.v. En viktig forskningsfråga som SALIENCE4CAV-projektet ställde var därför: *Hur kan man säkerställa att ett automatiserat fordon hela tiden förblir säkert samtidigt som man möjliggör frekventa uppdateringar av det automatiserade körsystemet med dess ingående komponenter?* De flesta metoder och standarder för säkerhetsanalys och säkerhetsbevisning som används idag är inte anpassade för kontinuerliga uppdateringar. Projektet har också undersökt flera relaterade öppna frågor som dels framkom under detta projekts föregångare, ESPLANADE, dels kommer från behov som identifierats av projektets industripartners. Det inkluderar hur man hanterar säkerhetsbevisning för maskininlärningskomponenter, användning av kvantitativa riskacceptanskriterier som en viktig del av säkerhetsargumentet, säkerhet för samarbetande automatiserade fordon, inklusive användning i blandad trafikmiljö, vilken roll säkra manövrar har i säkerhetsstrategin, och interaktion med mänskliga operatörer.

Projektet handlade alltså om säkerhet för automatiserade fordon (Automated Vehicle—AV), eller uppkopplade automatiserade fordon (Connected Automated Vehicle—CAV), och system-av-system av sådana fordon, i projektet även benämnt samarbetande och samverkande fordon. Själva fordonssystemet som utför automationen kallas för ett automatiserat körsystem (Automated Driving System—ADS). Utmaningen har flera delar. En AV måste fungera säkert när automatiseringen är aktiv. För att begränsa design- och verifieringsuppgiften specificeras en operativ designdomän (Operational Design Domain—ODD) som definierar de driftsförhållanden under vilka en ADS-funktion är avsedd att användas, t.ex. väg- och miljöförhållanden. Det innebär att funktionen måste vara säker inom sin ODD, och att den inte tillåts vara aktiv utanför sin ODD. Dessutom måste prestandakritiska systemfel, eller om fordonet riskerar att lämna sin ODD, hanteras säkert. Om det händer måste fordonet utföra en säker manöver (minimal risk manoeuver—MRM) och uppnå ett stabilt avstannat tillstånd (minimal risk condition—MRC). Till sist måste interaktionen mellan människor (t.ex. AV-operatörer, passagerare och andra trafikanter) och ADS (human machine interaction—HMI) vara säker.

Huvudmålet för projektet var att utveckla metoder och arbetssätt som kan användas av OEM:er, underleverantörer och tjänsteleverantörer, och som kan bli möjliggörare för kontinuerlig integration (continuous integration—CI) och kontinuerliga uppdateringar (continuous delivery—CD) för automatiserade fordon. Forskningsmetoden baserades på analys av användningsfall som tillhandahölls av partners, och som är relevanta för de definierade forskningsfrågorna. För varje ämne utfördes en state-of-the-art-analys, och identifiering av luckor där vi såg ett behov och en möjlighet att bidra. Forskningsfrågorna har förfinats iterativt under projektets gång. Metodutveckling gjordes främst på en teoretisk nivå, i vissa fall med stöd av simuleringar. Resultaten har validerats genom vetenskapliga publikationer som genomgått peer-review, presentationer i relevanta branschforum, och återkoppling från projektets industripartners.

Tabellen nedan listar de förfinade forskningsfrågorna och vilka resultat från projektet som ger ett bidrag till att besvara var och en av forskningsfrågorna. I resultaten hänvisas också till avsnitt i den engelska delen av slutrapporten där resultaten förklaras mer utförligt tillsammans med referenser till de publikationer projektet gjort inom området.

Forskningsfråga	**Resultat**
Vilka metoder finns redan tillgängliga och vilka utmaningar återstår för att uppnå CI/CD och kontinuerlig säkerhetsbevisning för ADS-utveckling?	Kartläggning och gap-analys av existerande metoder (se avsnitt 6.1.1) och undersökning av användning av säkerhetskontrakt i praktiken och deras potentiella relevans (se avsnitt 6.1.2).
Hur man härleder säkerhetskontrakt på olika abstraktionsnivåer för användning i kontinuerlig säkerhetsbevisning?	Säkerhetskontrakt för flera abstraktionsnivåer samt för komponent-hierarkier (se avsnitt 6.1.3).
Hur man väljer (kvantitativa och/eller kvalitativa) riskacceptanskriterier (risk acceptance criteria—RAC) för säkerhetsbevisning för en ADS?	Undersökning av designmål och deras relation till RAC, och en föreslagen uppsättning RAC som kan användas för en ADS (se avsnitt 6.1.4).
Hur man analyserar HMI, inklusive ändringar vid en produktuppdatering, ägarbyte, förarbyte etc.?	Process för säkerhetsanalys av HMI inklusive överenskommelser mellan ADS och olika aktörer (se avsnitt 6.1.4).
Hur man designar för uppfyllandet av en risknorm, och säkerställer detta under drift?	Design med en försiktighetsprincip för säker körning—precautionary safety (se avsnitt 6.2.1).
Vilken roll MRM/MRC spelar i en säkerhetbevisning, och dess relation till risk, ODD och taktiska beslut för individuella såväl som samarbetande eller samverkande AV:er?	Definition av MRC:s roll och strategier för att balansera taktiska beslut för att undvika MRC för olika kategorier och klasser av AV:er (se avsnitt 6.2.3).
Hur man definierar beslutshierarkier och interaktionsstrategier för samarbetande och samverkande fordon?	Genomgång/strukturering av möjliga strategier samt skapande av en enhetlig taxonomi för AV:er (se avsnitt 6.2.2).
Hur kan ett säkerhetskoncept med ML i en ADS utvärderas?	Ett tillvägagångssätt med simuleringsstöd för ML-säkerhetsanalys har föreslagits och utvärderats (se avsnitt 6.3.3). Effekten av ovanliga händelser (edge cases) utvärderades (se avsnitt 6.3.2).

Hur kan säkerhetsegenskaperna hos en ML-komponent övervakas?	Undersökning av möjliga användningar av tekniken Out-of-distribution detection (se avsnitt 6.3.1)
Vilka parametrar - inklusive viktiga systemparametrar och beroenden – karaktäriserar säkerhetskoncept med varianthantering över produktlinje och livscykel?	Undersökning av hur man utför systematisk specifikation av variationspunkter och systemparametrar med hjälp av verktyg, komponent-baserad design och säkerhetskontrakt (se avsnitt 6.4.2).
Hur kan modellen för varianthantering genereras automatiskt, och hur kan känsligheten hos parametrar effektivt analyseras från systemarkitektur och produktlinjemodeller?	Undersökning av verktyg för funktionell modellering kombinerat med variabilitetskonfiguration, och statistiska och datadrivna tekniker för känslighetsanalys (se avsnitt 6.4.3).
Hur kan ett säkerhetskoncept översättas till kontrakt på funktionell och teknisk abstraktionsnivå med varianthantering?	Undersökning av säkerhetskontrakt och förslag på högdimensionella kontrakt (se avsnitt 6.4.3 och avsnitt 6.1.3).

Rapporter har levererats för vart och ett av de fyra huvudämnena, *säkerhetsbevisning*, *design för säkerhet*, *säkerhet för maskininlärning*, och *säkerhet med varianthantering*. Huvudresultaten har beskrivits i femton vetenskapliga artiklar, varav 11 hittills är peer-granskade och publicerade, samt i 12 olika presentationer eller workshops. Forskningen har bidragit till en doktorsavhandling, en halvvägs-presentation, samt en studie för ytterligare en doktorand, och använts som underlag för flera projektmedlemmar som är med i standardiseringsgrupper inom ämnet. Positiva avvikelse från den ursprungliga planen är fler publikationer, bidrag till fler doktorander, samt tidig nytta i standardisering. Negativa avvikelser är att vissa forskningsfrågor övergivits eller inte kommit så långt som vi hoppades, både på grund av att det är utmanande problem (speciellt arbetet med säkerhetskontrakt) och på grund av förändringar i projekt och partners som medfört att vi inte kunnat fortsätta med alla ursprungliga forskningsfrågor. Bidraget till FFI-programmet[1] är främst till målet om minskade trafikskador (genom fokus på säkerhet) och ökad konkurrenskraft (genom forskning i ett avancerat område med förväntad tillväxt, genom samarbete mellan akademi, forskningsinstitut, tjänsteleverantörer, underleverantörer, och fordonstillverkare, samt genom kunskapsöverföring mellan olika domäner—vägtrafik och gruvor). Arbetet bidrar även till globala målen (3.6 minska antalet dödsfall och skador i vägtrafiken, 9.5 öka forskningsinsatser och teknisk kapacitet inom industrisektorn, samt 11.2 tillgängliggör transportsystem för alla).

SALIENCE4CAV-projektet pågick från januari 2021 till december 2023 med Agreat, Comentor, Epiroc Rock Drills, KTH Royal Institute of Technology, Qamcom Research and Technology, RISE Research Institutes of Sweden, Semcon Sweden, Veoneer (som under projektet förvärvades av Magna) och Zenseact som partners. Koordinering utfördes av RISE.

Denna slutrapport är en sammanfattning av projektresultaten och innehåller sammanfattningar av innehållet från projektleveranser och publikationer.

[1] Eftersom projektet antogs under 2020 gäller detta FFI:s färdplan från 2019.

3 Background

During the last decade, the automotive industry has gone through structural transformations powered by new breakthroughs in electrification, connectivity, and automation technologies. Regarding automation, recent advances in perception and computing technologies, as well as on active safety and advanced cruising features, have led to high expectations on a rapid development of automated vehicles (AVs). However, the development of AVs has proven to be more difficult and take longer than anticipated just a few years ago, partly due to the challenges in safety assurance[2]. Several national and international projects have focused on this issue including previous FFI projects such as FUSE, ESPLANADE, ASETS and ARCHER. Although progress has been made, several challenges remain.

This project focuses on safety for AVs, including both road vehicles and vehicles in confined areas, as well as AVs operating independently or a system-of-systems (SoS) of AVs, defined as a collection of vehicles that collaborate to yield positive effects not achievable by a single system. An example of the former is a privately owned road vehicle, while an example of the latter can be multiple vehicles of different types in a mine, collaborating to extract and transport ore. An AV is a vehicle equipped with one or more automation features that can perform the driving task on a sustained basis without human intervention[3]. The system providing automation is called an automated driving system (ADS[4]). If the AV also relies on connectivity to other vehicles, infrastructure, or cloud services, it is also called connected automated vehicle (CAV).

The safety assurance challenge consists of several parts. First, the AV must operate safely when the automation is enabled. To confine the design and verification task, an operational design domain (ODD) is specified; the ODD defines the operating conditions under which an ADS feature is designed to function, e.g., road and environmental conditions. The first task thus means making sure the ADS operates safely within its ODD, and that it cannot be active outside the ODD. Secondly, the AV must have a safe fallback to execute in case of performance-critical system failure, or if the AV risks exiting the ODD; if this happens the AV must be able to execute a minimal risk manoeuvre (MRM) and achieve a stable stopped state called a minimal risk condition (MRC). Thirdly, the human machine interaction (HMI) between AVs and various human stakeholders (e.g., AV operators, passengers, and

[2] https://www.theguardian.com/technology/2022/mar/27/how-self-driving-cars-got-stuck-in-the-slow-lane

[3] The terminology around vehicle automation varies between sources which can make it difficult to understand the exact meaning in various contexts. For instance, terms such as fully automated, highly automated, and autonomous are used in various standards, legislative documents, and research papers for vehicles able to operate without relying on a human operator or supervisor. Assisted, driving automation, or semi-autonomous are used for types of automation where some form of human involvement or supervision is still necessary. We mainly use the terms defined in the standard SAE J3016, however, due to the common use and for the sake of brevity, we also use the terms AV and CAV to refer to a vehicle equipped with an ADS, able to operate at least for some period of time without constant human supervision.

[4] SAE. "J3016:2021 - Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles." Retrieved from https://www.sae.org/standards/content/j3016_202104/.

other road users) must be safe, taking both human and AV capabilities and limitations into account.

This project includes research questions related to all of these parts. For defining risk acceptance criteria for the safety assurance, previous work in the ESPLANADE project proposed a quantitative risk norm (QRN) based on accident statistics as a base for safety requirements rather than a classical hazard analysis[5], due to limitations in the existing methods when applied to ADS features. This project explores the use of such a QRN when designing a safe ADS driving policy. The role of the MRC in safety assurance and safety analysis of human interactions are also investigated, including MRM/MRC in the context of SoS.

Another safety assurance challenge relates to machine learning (ML), which is often used in ADSs, especially deep neural networks (DNN) used as part of the perception system, e.g., for object classification. However, current ML models have weaknesses when it comes to robustness, susceptibility to adversarial attacks, and handling natural perturbations that occur in open-world scenarios[6]. The project investigates some ML issues and especially investigates the potential role of out-of-distribution (OoD) detection[7] in safety assurance.

Due to the safety assurance challenges, we are now starting to see actors adopting a more cautious approach where ADSs are developed progressively, rather than aiming for the more aggressive rollout originally envisioned when vehicle OEMs first started investing heavily in AVs[8]. That is, to be able to release the systems to the market in reasonable time, they will be conservative when first released in the sense that they only take on simpler tasks where the safety can be assured, e.g., only be available on certain roads or cities, or in certain weather conditions, or only allowed to operate in certain access-restricted zones within a confined area such as a mine or a harbour. The ADS software can then be continuously updated to allow the system to handle more challenging driving tasks or perform existing tasks with higher performance as confidence in the solution increases, e.g. due to better ML algorithms and training data, improved verification, or more field data to validate real-world performance and tune the ADS driving policy.

In other domains, agile and iterative development has long been a common development methodology to allow for new product versions at a predictable and high pace[9]. This is typically done by adopting agile ways-of-working and the implementation of continuous integration (CI)—which implies frequent and

[5] Warg, F., Skoglund, M., Thorsén, A., Johansson, R., Brännström, M., Gyllenhammar, M., & Sanfridson, M. (2020, June). The quantitative risk norm-a proposed tailoring of HARA for ADS. In *2020 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)* (pp. 86-93). IEEE. (Paper from the ESPLANADE project).
[6] Mohseni, S., Wang, H., Xiao, C., Yu, Z., Wang, Z., & Yadawa, J. (2022). Taxonomy of machine learning safety: A survey and primer. *ACM Computing Surveys*, *55*(8), 1-38.
[7] Yang, J., Zhou, K., Li, Y., & Liu, Z. (2021). Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334*.
[8] https://techxplore.com/news/2023-09-self-driving-car-revolution-slowly.html
[9] https://www.infoworld.com/article/3655646/a-brief-history-of-the-agile-methodology.html

automated tests to reduce the effort of integrating product components into a complete and functional system—and continuous deployment (CD)—which is the practice of deploying every update directly to the customer, or even DevOps—integration of development and operations by using direct feedback from deployed systems to facilitate rapid development of new improved versions. However, iterative development is more challenging for safety-critical systems as they must have a safety case proving that the product is safe, and this safety case must be complete and consistent for every product release. Lifecycles described in standards for safety-critical products are not yet adapted for iterative development, neither are many methods and tools. One way-of-working proposed in the predecessor project ESPLANADE, is continuous assurance cases[10], i.e., assurance (or safety) cases designed for iterative development using a combination of safety contracts[11], modular assurance cases[12] and reusable assurance patterns. Some of the work in this project builds on these principles, including work on safety contracts and on formal reasoning of variability in operational safety for effective conformity assessment and change management.

Speed of development is of essence for ensuring the competitiveness of companies, especially when it comes to software-based systems, and software is increasingly important in new vehicle features. Therefore, the automotive domain is already moving in this direction. Enabling iterative development of safety-critical functions is imperative for securing competitiveness for development of advanced functions such as ADSs.

[10] Warg, F., Blom, H., Borg, J., & Johansson, R. (2019, October). Continuous deployment for dependable systems with continuous assurance cases. In *2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)* (pp. 318-325). IEEE. (Paper from the ESPLANADE project).

[11] Graydon, P., & Bate, I. (2014, November). The nature and content of safety contracts: Challenges and suggestions for a way forward. In *2014 IEEE 20th Pacific Rim International Symposium on Dependable Computing* (pp. 135-144). IEEE.

[12] Fenn, L., Hawkins, R. D., Williams, P. J., Kelly, T. P., Banner, M. G., & Oakshott, Y. (2007, October). The who, where, how, why and when of modular and incremental certification. In *2007 2nd Institution of Engineering and Technology International Conference on System Safety* (pp. 135-140). IET.

4 Purpose, research questions and method

CAVs are expected to provide more efficient, accessible, and safer transport solutions. However, there remain open questions when it comes to developing a safe design and how to conduct the safety assurance. The purpose of the project is to contribute with methods towards enabling the safe introduction and continuous evolution of CAVs for both on-road and confined area use cases. The main expected results of the project were therefore *methods and practices to be used by OEMs, suppliers and service providers alike*, which will work as *enablers for market introduction followed by continuous deployment for CAVs*. This includes tackling challenges in several areas identified by the industry partners and in the predecessor project ESPLANADE, which is reflected in the research questions and objectives discussed below.

The research method was based on analysis of use cases provided by the partners, which are relevant to the defined research topics. For each topic, state-of-the-art analysis was performed, followed by identification of research gaps where we saw a need and opportunity to contribute. The research questions were iteratively refined during the project. Method development was mainly done on a theoretical level, in some cases supported by simulation experiments. Physical/prototype tests were not performed or planned to be part of the project. The main method of validation has been submitting our results to peer-review through scientific publications and talks. The OEM/Tier1 representatives in the group has also provided feedback regarding the relevance of the examples and methods.

Due to the rapid changes in the CAV landscape during the project, both from a scientific perspective — several concurrent international research projects and standardization efforts have been active in the same areas — and from a business perspective — the last few years have seen a consolidation of actors and revision of goals and business plans for CAVs — there has been a need to conduct the project in an agile fashion and be prepared to update the research questions and objectives.

The project application defined a gross list of research questions under the main topics of *safety assurance, safety design, machine learning, and variational safety* summarized below[13]:

Continuous assurance case[14] is a proposed way-of-working with safety contracts to handle component-based design and evolving the safety case for iterative development:

- What safety argumentation strategies enables a continuous safety case?
- How to efficiently express semi-formal safety contracts for use in requirements refinement and continuous safety cases?

[13] The list of research questions is from the project application, however edited for the sake of brevity.

[14] Warg, F., Blom, H., Borg, J., & Johansson, R. (2019, October). Continuous deployment for dependable systems with continuous assurance cases. In *2019 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)* (pp. 318-325). IEEE. (Paper from the ESPLANADE project).

The QRN approach[15] is a way to elicit safety goals with quantitative targets, which can avoid some problems with, e.g., ISO 26262[16] ASIL rules:

- How to build up a quantitative safety case structure?
- How to integrate qualitative arguments in a quantitative safety case?
- How to design for fulfilment of the risk norm and ensure fulfilment during run-time (e.g., tactical/operational decisions)?
- Can big data and automated testing assist in achieving CD, and can it involve verification of requirements on component level formulated as part of a QRN?
- How can a frequency-based ODD (as part of a QRN) be formulated and updated to support CD?

To facilitate updating features, HMI agreements[17] must be clearly specified, and analysed to make sure they are handled safely, considering the potential human errors:

- How to analyse a proposed way of handling an agreement change for a product update, ownership change, driver change etc?
- How to ensure that agreements are understood and entered in good faith?
- If there is a sudden change of context necessitating an immediate revision, how to ensure that the revised agreement is communicated in a clear, timely fashion?

After market introduction, a new challenge for system design is a need for continuously updating ADSs to maintain safety due to changes in real-world operating conditions:

- How can the safety concept of an ADS be designed to support CD?

MRMs must be performed such that the safety of CAVs is maintained. We aim to bridge the gap between ODDs and MRMs, to suggest a strategy on how to approach this matter:

- How to approach MRMs in relation to ODDs?

Safety assurance of system of system (SoS) implies considering emergent properties that are not visible at the level of a single vehicle, issues to investigate include:

- Decision hierarchies and decision allocation in SoS.
- Cooperative strategies for guaranteeing not to exit ODD.
- Balancing centralized and decentralized decision with respect to safety.
- Communication requirements in assuring safety.
- Operational Fallback and Minimal Risk Conditions for SoS.

[15] Warg, F., Skoglund, M., Thorsén, A., Johansson, R., Brännström, M., Gyllenhammar, M., & Sanfridson, M. (2020, June). The quantitative risk norm-a proposed tailoring of HARA for ADS. In *2020 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)* (pp. 86-93). IEEE. (Paper from the ESPLANADE project).

[16] International Organization for Standardization. (2018). *Road vehicles Functional safety* (ISO Standard No. 26262:2018).

[17] Skoglund, M., Warg, F., & Sangchoolie, B. (2018, September). Agreements of an automated driving system. In *37th International Conference on Computer Safety, Reliability, & Security SAFECOMP2018 SAFECOMP2018*. (Paper from the ESPLANADE project).

ML use in CAVs imply many challenges from a safety perspective due to the inherent uncertainty mainly originating from the lack of formal explainability:

- How can a safety concept using ML in an AD application be designed?
- Criteria for safety monitoring of ML?

DNNs are typically trained to classify objects. From a safety perspective, however, ensuring the absence of certain objects in the path of the vehicle is more relevant:

- What is the difference between object and absence detection networks?
- Are absence detection networks more usable for safety?

Description of variational safety concerns across both product-line and product lifecycle is a challenge considering variabilities in e.g., ODD and safety requirements.

- What are the variation points characterizing variational safety concepts across product-line and product lifecycle?
- What are the key system parameters, constraints and utilities charactering such variational safety concepts, their interdependencies, and binding criteria?
- How can the variability model be automatically synthesized in alignment with the corresponding system architecture and product line models?
- How can the sensitivity of variational safety concepts be revealed effectively?
- How can the variational safety concepts be bound with the contextual assumptions of system operation?
- How can the variational safety concepts be translated to the functional and technical contracts of components?

During the project the research questions have been refined due to the mentioned changes in the CAV landscape and our developing understanding of the relevance of different questions. The research question on frequency-based ODD was abandoned due to evolving standardization on ODDs prompting us to rethink and abandon this line of work, and some of the original research questions regarding HMI, ML, communication requirements, and big data, were removed due to changes in the group of project participants prompting us to focus on questions we had better expertise to pursue. In some cases, the questions have simply been combined and/or rephrased for clarity as we have gained more understanding of the topic. An additional question regarding available methods and gaps for achieving CI/CD was instead added. The updated research questions and their mapping to project results is shown in Table 1.

Table 1 Research questions and mapping to project results.

Research question	Results
Which methods are available and what challenges remain for achieving CI/CD and continuous assurance for ADS development?	Survey and gap analysis of current methods (see Section 6.1.1) and state-of-practice/safety contract relevance investigation (see Section 6.1.2).
How to derive safety contracts on different abstraction levels for use in continuous safety cases?	Safety contracts for multiple abstraction levels and compositional hierarchies (see Section 6.1.3).

How to select (quantitative and/or qualitative) risk acceptance criteria (RAC), such as the QRN, for an ADS safety case?	Investigation of design goals and their relation to RAC, and a proposed set of RACs for an ADS (see Section 6.1.4).
How to analyse HMI agreements including agreement changes for a product update, ownership change, driver change etc?	Human interaction safety analysis process including agreement analysis (see Section 6.1.4).
How to design for fulfilment of the risk norm and ensure fulfilment during run-time?	Design using the principle of precautionary safety (see Section 6.2.1).
The role of MRM/MRC in safety assurance and its relation to risk, ODD and tactical decisions for individual as well as cooperative and collaborative AVs?	Defining the role of MRC and strategies for balancing tactical decisions to avoid MRC for different categories and classes of AVs (see Section 6.2.3).
How to define decision hierarchies and interaction strategies in cooperative and collaborative vehicles[18]?	Investigation in possible strategies and creation of unified taxonomy for AVs (see Section 6.2.2).
How can a safety concept using ML in an ADS be evaluated?	A simulation-aided approach to ML safety analysis was proposed and evaluated (see Section 6.3.3) and the impact of edge cases evaluated (see Section 6.3.2).
How can the safety properties of an ML component be monitored?	Investigation into possible uses of one technique for ML monitoring—Out-of-distribution detection (see Section 6.3.1)
What are the variation points—including their key system parameters and interdependencies—characterizing variational safety concepts across product-line and product lifecycle?	Investigation into systematic specification of variation points and the key system parameters using tools, component-based design, and safety contracts (see Section 6.4.2).
How can the variability model be automatically synthesized, and the sensitivity of variational safety concepts be revealed effectively, in alignment with the system architecture and product line models?	Investigation of tools for feature modelling combined with variability configuration, and statistical and data-driven techniques for sensitivity analysis (see Section 6.4.3).
How can the variational safety concepts be translated to the functional and technical contracts of components?	Investigation of safety contract refinement and proposing high-dimensional contracts (see Section 6.4.3 and Section 6.1.3).

[18] To better define the type of interaction between multiple CAVs we have defined categories and classes of cooperative and collaborative vehicles, further described in Section 6.2.2,and henceforth use this terminology to describe our work instead of SoS.

5 Objective

The project focus is methodology for the design and assurance of safety-critical systems, in particular automated driving systems. Thus, the main expected results are methods enabling iterative development and eventually continuous deployment of such systems. The objective is to develop methods in the following four topic areas:

- Methods for safety assurance, including continuous safety cases, quantitative safety cases, and agreements between human and ADS.
- Methods for safety design including a safety concept enabling CD, and defining MRCs and decisions hierarchies in individual, cooperative, and collaborative vehicles[19].
- Methods for ensuring safety when components include machine learning.
- Methods for operational safety including formal specification of operational safety for product lines and managing quantitative operating conditions in run-time.

In addition to the concrete deliverables and methods, project objectives include:

- Contribute to the research for an industrial Ph.D. student on the topic of ML.
- Dissemination of project results mainly through scientific publications and talks.
- It is expected that some of the project results may be relevant for standardization.

[19] This objective has been changed compared to the application reflecting: (1) the removal of the research question on frequency-based ODD as described in Section 4, and instead adding MRC and decision hierarchies for individual, cooperative, and collaborative vehicles, which were listed in the original research questions but missing from the objectives.

6 Results and deliverables

In this section, the main results of the project are described, with one subsection for each of the four main objectives described in Section 5. The last subsection reflects on the goal fulfilment and contribution to the FFI program.

6.1 Safety assurance

6.1.1 Methods, gaps, and directions

Two surveys on ADS safety assurance have been conducted. The first[20] focused on analysing some assurance methods proposed in research based on of seven criteria deemed critical based on the goals of the SALIENCE4CAV project:

1. Support for ways-of-working with frequent updates.
2. Ability to make use of operational data in development (e.g., DevOps).
3. Support monitoring for changes in operational context to ensure that the safety case remains valid.
4. Management of multiple variants
5. Support for modularity and inclusion of parts from different suppliers
6. Support assurance for self-adaptive systems (perpetual assurance)
7. Support for quantitative safety cases (e.g., based on a QRN)

In Table 2 below, the name of each method is stated together with our evaluation of how they, as currently described, support each of the criteria (X in parentheses means partly supporting). We note that in several cases the methods could be extended or combined to fulfil additional criteria. In particular, we compare other methods to the ideas around *continuous assurance cases* from the ESPLANADE project that SALIENCE4CAV aimed to build upon.

Table 2 Evaluation of safety assurance methods.

Assurance method	**Assurance method criteria**						
	1	2	3	4	5	6	7
Safety-contract based					X		
Conditional safety certificates (ConSert)			(X)	X	X	X	X
Dynamic assurance cases			X			(X)	X
Product-line contract-based design				X	X		
Continuous assurance cases	X	X			X		

[20] Gyllenhammar, M., Bergenhem, C., & Warg, F. (2021). ADS Safety Assurance–Future Directions. In *CARS 2021: 6th International Workshop on Critical Automotive Applications: Robustness & Safety*. (Paper from the SALIENCE4CAV project).

In the second study[21], the scope was expanded to investigate a larger number of techniques related to ADS safety assurance, divided into four main categories: *design techniques, verification and validation techniques, run-time risk assessment, and run-time adaptation*. The techniques were investigated considering eight identified challenges for ADS safety assurance:

1. Uncertainties associated with the operational environment.
2. Uncertainties associated with the interaction with other traffic participants.
3. Responsibility of the ADS for strategic, tactical, and operational decisions.
4. Complex functions and sub-systems.
5. Self-adaption capabilities, e.g., to cope with permanent or temporary degradations.
6. High dependability requirements
7. Validation of black-box components, e.g., containing machine learning.
8. Frequent releases and continuous learning.

The study identified which of the challenges each of the 17 investigated methods supports in solving and identified many research gaps related to the methods. In conclusion, existing methods are found to provide a good base for provision of safety evidence, however some open questions remain, and there is also a need to combine methods to bridge some of the gaps.

6.1.2 CI/CD – challenges and state of practice

The path for an organization to attain the ability of frequent releases/updates and the use of operational data as feedback to development has been illustrated as a "stairway to heaven"[22], illustrated in Figure 1, where each step introduces new abilities, and represents maturity steps many organizations go through when transitioning from a traditional software development model to CD or DevOps. The first step is to introduce an agile way-of-working, followed by automating tests and implementing a continuous integration machinery. When fully implemented, this enables continuous deployment, which in turn simplifies the further integration of field data collection as part of the development lifecycle.

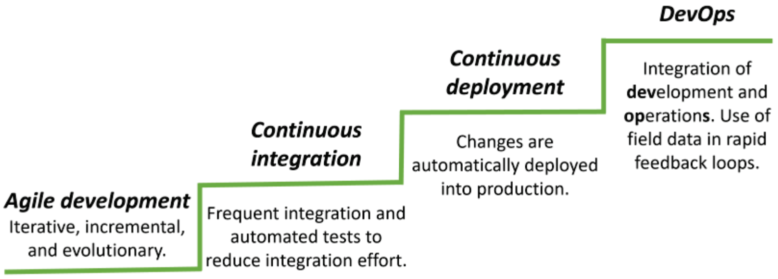

DevOps
Integration of development and operations. Use of field data in rapid feedback loops.
Continuous deployment
Changes are automatically deployed into production.
Continuous integration
Frequent integration and automated tests to reduce integration effort.
Agile development
Iterative, incremental, and evolutionary.

Figure 1 The 'stairway to heaven' model of maturity towards continuous updates.

[21] Gyllenhammar, Magnus; Rodrigues de Campos, Gabriel; Törngren, Martin (2022): Holistic Perspectives on Safety of Automated Driving Systems - Methods for Provision of Evidence. TechRxiv. Preprint.
https://doi.org/10.36227/techrxiv.20331243.v1. (Paper from the SALIENCE4CAV project).

[22] Olsson, H. H., Alahyari, H., & Bosch, J. (2012, September). Climbing the" Stairway to Heaven"--A Mulitiple-Case Study Exploring Barriers in the Transition from Agile Development towards Continuous Deployment of Software. In *2012 38th euromicro conference on software engineering and advanced applications* (pp. 392-399). IEEE.

At SCSSS 2022[23], around 60 practitioners gathered in a workshop arranged by the SALIENCE4CAV project to discuss state-of-the-art for CI/CD and DevOps for safety-critical systems and the use of safety-contract-based design as one potential enabler. Most of the participants had safety or development roles in domains developing safety-critical products. Figure 2 shows that the represented organizations are at different levels of the "stairway to heaven", with most having adopted agile ways-of-working, a majority also using CI, but a minority having implemented CD or DevOps.

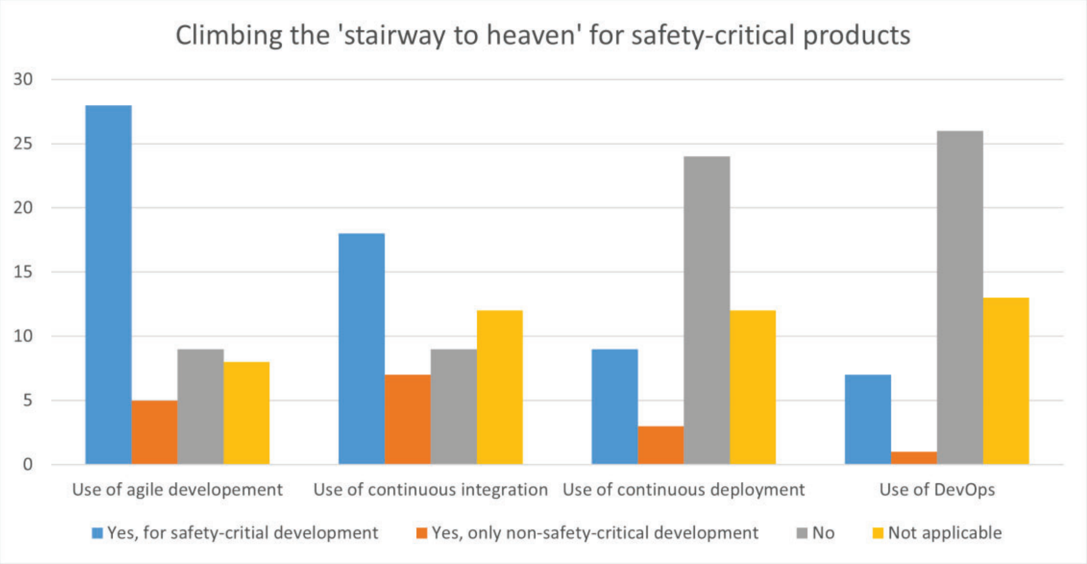


Figure 2 Results of workshop participant poll. State of practice for CI/CD/DevOps.

After a presentation of safety-contract-based design, the participants were asked whether they had heard about this technique before and if they believed it could be a useful tool. Figure 3 shows that it was a new concept to many, despite being a relatively old topic in research, but that a majority believed it would be, or has the potential to be useful given proper tool support. We believe this workshop indicates the relevance of the research topics in the project related to CI/CD and safety contracts.

The presentations of CI/CD challenges and safety contracts as well as feedback and answers to further questions to the participants have been published in a report from the project[24].

[23] 10th Scandinavian Conference on System & Software Safety. Göteborg, November 22-23, 2022, http://safety.addalot.se/2022

[24] Warg, F., Thorsén, A., Cassel, A., Jaradat, O., Nejad, N., Chen, D., & Ursing, S. (2022). Managing Continuous Assurance of Complex Dependable Systems : Report from a workshop held at the Scandinavian Conference on System and Software Safety (SCSSS) 2022. Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:ri:diva-67730. (Report from the SALIENCE4CAV project).

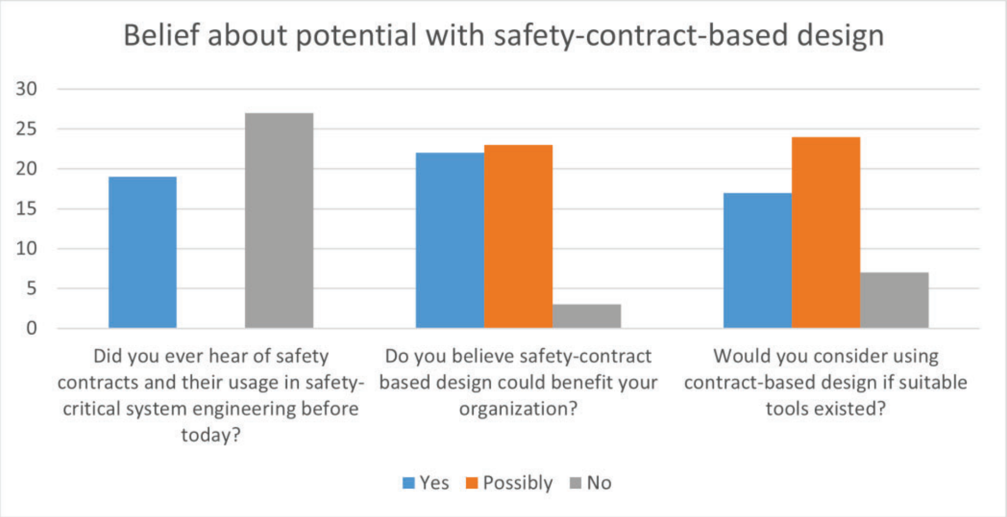


Figure 3 Results of workshop participant poll. Belief of usefulness of safety-contract-based design.

6.1.3 Managing continuous assurance with safety contracts

An automotive electrical/electronic (E/E) system consists of a variety of functions and components for features relating to vehicle control and infotainment. Each function or component can have its own requirements for functionality, performance, reliability, safety, extensibility, etc. During the system development, each functional feature of a vehicle, often referred to as item[25], is gradually realized through decomposition and mapping decisions across multiple abstraction levels. As shown in Figure 4, these abstraction levels normally range from highly abstract FSC (Functional Safety Concept) to TSC (Technical Safety Concept), and then down to HW and SW at the lowest level. Compliance with safety standards and regulations is also a critical aspect. The decisions involve not only considerations of functional effectiveness and efficiency but also concerns of faults and their potential safety consequences. One of the most challenging tasks is related to safety assurance and safe planning. The goal is to at least guarantee that, if other road users act according to certain assumed behaviours, the vehicle will not crash and will mitigate collisions for unforeseen behaviours of other road-users[26].

[25] International Organization for Standardization. (2018). *Road vehicles Functional safety* (ISO Standard No. 26262-3:2018).
[26] SAE. "J3016:2021 - Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles." Retrieved from https://www.sae.org/standards/content/j3016_202104/.

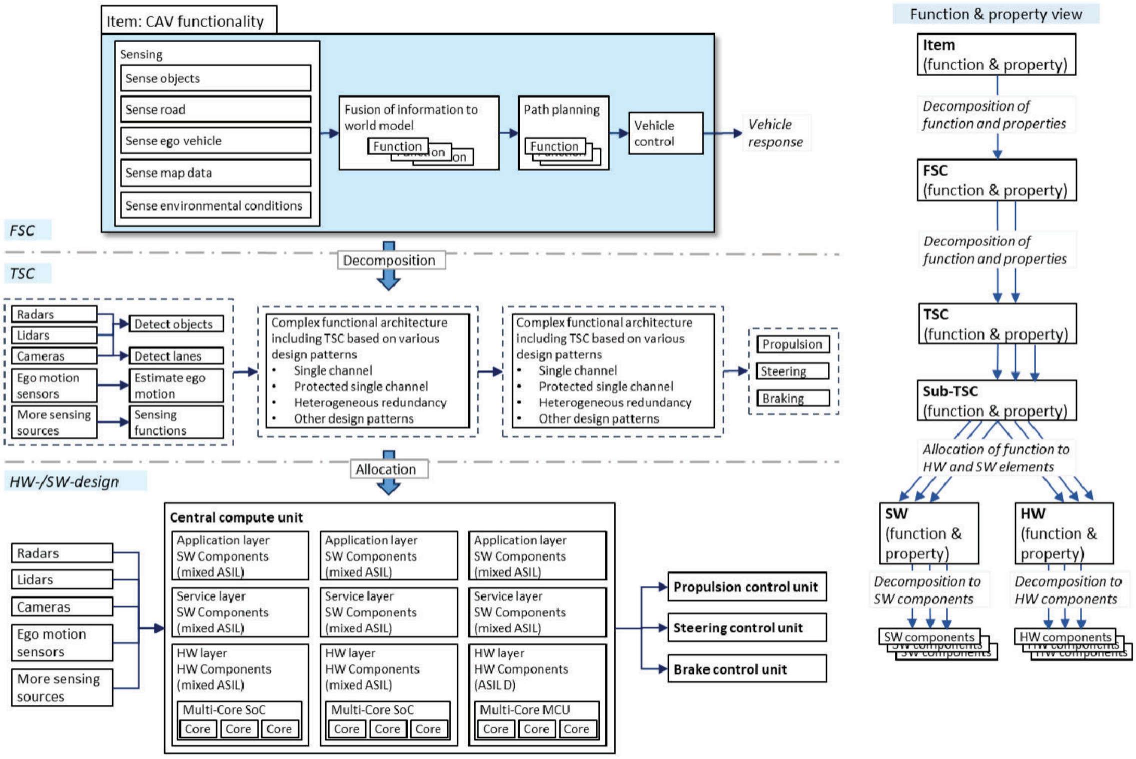


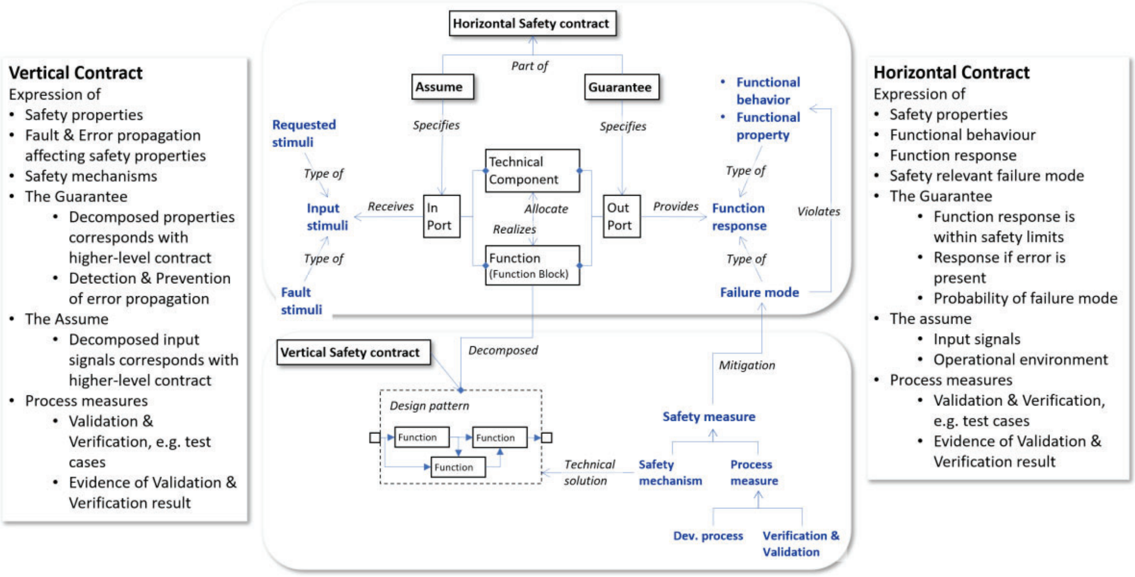
Figure 4. Target system and system realisation from abstract functions to software and hardware components.

To this end, modularization plays a key role for the success as the design decisions involve communication and cooperation across engineering teams, OEMs, and suppliers. Over the years, researchers and practitioners have explored ways to support modularization. For software design, the notion of *contract* that follows the object-oriented design principle has been proposed for improving software quality and reducing errors by explicitly specifying the various computational predicates in terms of *preconditions*, *postconditions* and *invariants*. For a formal reasoning of component composability and compositionality, such contracts have then been extended with formal semantics regarding system behaviours based on operational traces. In recent years, the notion of contract has been extended to cover the development of E/E systems with multiple abstraction layers, heterogeneous models of computation, and multidimensional contractual concerns such as timing, reliability, power, etc. One well-known concept is *Rich Components* (RC), which extends classical component models with both functional and technical dependencies in a multilayered design hierarchy for electronic systems. The goal is to support speculative design where multiple engineering teams work concurrently without awaiting synchronization. *Heterogeneous Rich Components* (HRC) is a component model that extends the RC model and Tagged-Signal model for heterogeneous synchronization and communication mechanisms, timing, and other non-functional properties.

This work is focused on the development of a structured methodology for holistically specifying component compliance concerns in multiple dimensions while enforcing the safety-related requirements and constraints throughout continuous integration and development cycles of automotive E/E systems[27]. The key concept, referred to as

[27] This work has not been published at the time of writing this report; however, a paper on the subject is ongoing and we aim at publication after the end of the project.

Safety Contracts (SCs), aims to serve as a mechanism for Design by Contract (DbC) for safety-critical systems across multiple abstraction levels and compositional hierarchies. See Figure 5 for an overview of the design. By formally specifying the expectations and obligations regarding functionality, performance, robustness, as well as related design rationale, the contract mechanism allows a component to be developed, maintained, and evolved independently with varying details as long as the specified properties are satisfied. Such a contract mechanism offers several benefits as it is centred on the actual design commitments and compliance agreements using directly the functional and technical parameters of components. This contrasts with conventional requirement specifications, which are centred on the desired conditions or capabilities that must be met or possessed, often described without explicit information about the actual design commitments and compliance agreements. Moreover, requirements are often given in natural or other high-level language. Essentially, for component specification, a requirement-based approach tends to become disconnected from component changes and implementations over time in the context of continuous integration and development (CI/CD).



Vertical Contract
Expression of
- Safety properties
- Fault & Error propagation affecting safety properties
- Safety mechanisms
- The Guarantee
  - Decomposed properties corresponds with higher-level contract
  - Detection & Prevention of error propagation
- The Assume
  - Decomposed input signals corresponds with higher-level contract
- Process measures
  - Validation & Verification, e.g. test cases
  - Evidence of Validation & Verification result

Horizontal Contract
Expression of
- Safety properties
- Functional behaviour
- Function response
- Safety relevant failure mode
- The Guarantee
  - Function response is within safety limits
  - Response if error is present
  - Probability of failure mode
- The assume
  - Input signals
  - Operational environment
- Process measures
  - Validation & Verification, e.g. test cases
  - Evidence of Validation & Verification result

Figure 5. Safety Contracts (SCs), extending Design by Contract (DbC) for multiple abstraction levels and compositional hierarchies.

Current development of automotive systems follows the ISO 26262 V-cycle across Tier-1 sub-suppliers, where *Development Interface Agreements* (DIA) are used for the sharing of safety requirements, test results, and compliance with ISO 26262. Such interface mechanisms create bottlenecks delaying simultaneous development and feedback loops regarding the items and related safety assurance as they are shown ineffective for continuous information exchange due to sequential dependencies, sequential verification, and validation (V&V) cycle and manual compilation of safety case, and assessment of completeness and correctness of safety argumentation and supporting evidence. A CI/CD based approach to safety-critical functionality requires also that safety assurances activities are part of the CI/CD loop, as shown in Figure 6.

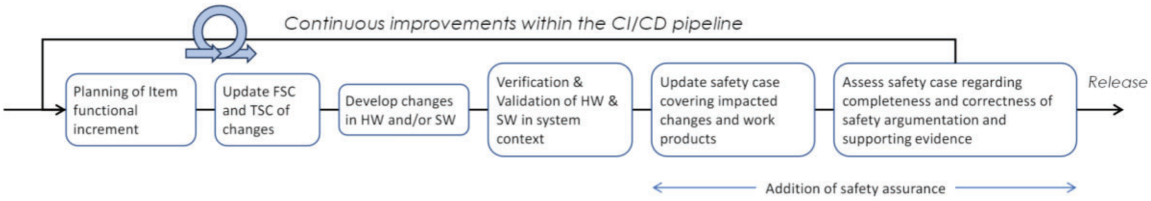

Continuous improvements within the CI/CD pipeline
Planning of Item functional increment
Update FSC and TSC of changes
Develop changes in HW and/or SW
Verification & Validation of HW & SW in system context
Update safety case covering impacted changes and work products
Assess safety case regarding completeness and correctness of safety argumentation and supporting evidence
Release
Addition of safety assurance

Figure 6. Integration of safety assurance in CI/CD.

To support a structured approach to managing the complex interdependencies across continuous integration and development cycles through a contract mechanism, a novel concept, referred to as *High-Dimensional Safety Contracts (SCs)*, for a modularized specification of expected compliance beyond basic functional operation across the lifecycle stages has been proposed by the project and paved a basis for future work. In particular, for safety assurance over various maintenance and evolution cycles, the contract mechanism contains an extension for additional dimensions of compliance, including failure modes, safety mechanisms, and expected V&V measures. This offers several advantages over conventional component specifications. The proposed methodology is still to be validated through case studies involving real-world automotive systems.

6.1.4 Selecting risk acceptance criteria for an ADS

It is easy to agree that an ADS shall be safe, but it is an on-going discussion what safe means. Several Risk Acceptance Criteria (RAC) candidates have been suggested, e.g., in standards and papers, but a closer analysis indicates that not all of them are related to risk in a traffic safety sense and that perhaps they are better described as properties that an ADS should be designed to exhibit for other reasons. This work[28] analysed safety aspects of ADS features and different design goals for safe automated driving and puts forward a combination of Risk Acceptance Criteria (RAC) for limiting the risk of harm, as a design and analysis tool. These criteria are analysed in terms of what they achieve and articulates the consequences and implications on the design, implementation, and validation of ADS features, and on why such a combination of RAC is suitable. Furthermore, it is also shown why run-time risk transfer is unavoidable in any system that makes tactical decisions under uncertainty and why this motivates avoiding thought-examples such as the trolley problem as basis for ADS design.

6.1.5 Human interaction safety analysis

A vital part of ADS safety is the interaction between the machine and humans. Different functions of a CAV may involve various stakeholders like drivers, passengers, and other road users leading to numerous safety-related interactions, including driver interface issues, communication with other road users, and changes in function behaviour due to over-the-air updates. In the context of HMIs, we use the

[28] Sandblom, F., De Campos, G. R., Hardå, P., Warg, F., Beckman, F. (2024) Choosing Risk Acceptance Criteria for Safe Automated Driving. *Submitted for review.*

term *'agreement'*[29] to denote the mutual understanding required for successful interaction. Ensuring HMI safety involves identifying all relevant agreements and ensuring that the necessary interactions for each agreement are properly designed and implemented in the CAV to maintain a risk that is acceptable for all stakeholders.

Building on results from the ESPLANADE project, where a safety analysis method for individual interactions was developed[30], the SALIENCE4CAV project defined the Human Interaction Safety Analysis (HISA)[31] process illustrated in Figure 7. The process includes agreement analysis, interaction analysis, risk assessment, HMI V&V, and impact analysis of HMI changes. The process is iterative including improvements in the HMI function if the interaction cannot be proven sufficiently safe, in a similar manner as the SOTIF[32] standard that also includes HMI safety in its scope. Hence, we believe HISA is suitable to use as part of a safety case according to this standard.

In particular, the agreement analysis step was elaborated in the project. The agreement analysis is a method to systematically consider different aspects of an Automated Function Under Analysis (AFUA), aiming to ensure all relevant agreements are included in the safety analysis. The process involves:

1. *Defining Concerns*: Listing quality attributes and their acceptance criteria, including safety and related attributes like security and legal considerations.
2. *Defining lifecycle phases and events*: Listing events or phase transitions in a CAV lifecycle that may affect agreements with human stakeholders. This includes detailing the vehicle lifecycle to capture less obvious, safety-relevant agreements. A visual example of such a lifecycle analysis is shown in Figure 8.
3. *Defining Stakeholders*: Listing stakeholders that may be part of agreements, divided into users (of the CAV), proximal stakeholders (persons in the vicinity of the vehicle), and distal stakeholders (persons or entities with a more indirect relation to the AFUA).
4. *Listing Functional Agreements*: Eliciting all applicable agreements by considering combinations of lifecycle/stakeholder that will constitute an agreement for the AFUA. This is part of the concept stage of development and the agreements can then be further refined into an implementation proposal to be analysed in the interaction analysis step.

[29] Skoglund, M., Warg, F., & Sangchoolie, B. (2018, September). Agreements of an automated driving system. In *37th International Conference on Computer Safety, Reliability, & Security SAFECOMP2018 SAFECOMP2018*. (Paper from the ESPLANADE project).

[30] Warg, F., Ursing, S., Kaalhus, M., & Wiik, R. (2020). Towards Safety Analysis of Interactions Between Human Users and Automated Driving Systems. In *10th European Congress of Embedded Real Time Systems (ERTS 2020)*. (Paper from the ESPLANADE project).

[31] Warg, F., Skoglund, M., & Sassman, M. (2021, September). Human Interaction Safety Analysis Method for Agreements with Connected Automated Vehicles. In *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)* (pp. 01-07). IEEE. (Paper from the SALIENCE4CAV project).

[32] International Organization for Standardization. (2022). *Road vehicles - Safety of the intended functionality* (ISO Standard No. 21448:2022).

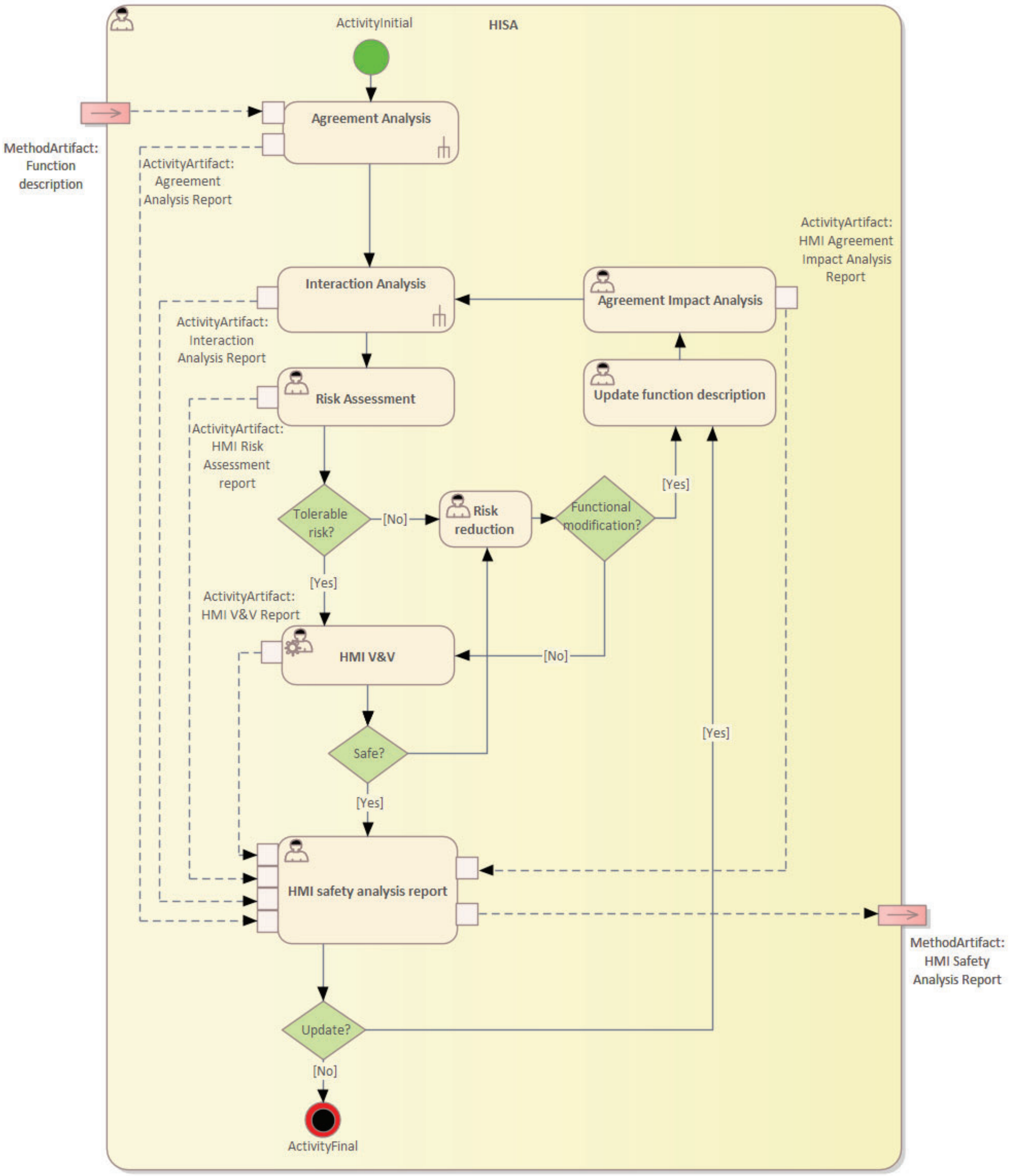


Figure 7 Human interaction safety analysis process.

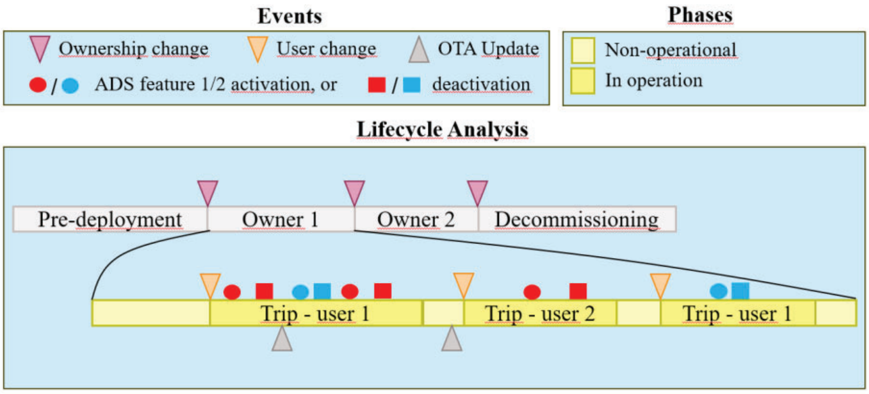


Figure 8 Visual example of lifecycle analysis.

Updating safety-related functionality involves planning updates and performing safety activities as required by the functional changes. For an AFUA update, the function description is revised, and an impact analysis is conducted to identify any altered or new agreements. The process is repeated for these agreements, which is simplified if HISA is integrated into a traceable development process, e.g., using safety contracts as discussed in Section 6.1.3.

6.2 Safety design

6.2.1 Precautionary safety

As mentioned, safe planning remains one of the most challenging tasks for ADS development. There are several industrial driven solutions that aimed at providing holistic safety approaches. For instance, NVIDIA's Safety Force Field concept[33], as a safety layer for obstacle avoidance which guarantees that the AD vehicle does not expose other road users to dangerous behaviours. Similarly, Mobileye has proposed a white-box, interpretable, mathematical model for safety assurance, denoted as Responsibility-Sensitive Safety[34].

A crucial element of an ADS is the underlying notion of safety, which is often constructed from multiple approaches and concepts. For instance, some researchers define safe driving as legal safety, i.e., in the sense that ADSs are considered safe if they always obey to a set of rules[35]. However, the underlying assumption that other road users always follow rules, is questionable. In fact, many people violate traffic rules, either on purpose or by mistake: driving faster than the speed limits, getting distracted, taking way when changing lanes or when driving through intersections. Fortunately, the infrastructure is built to be resilient to human errors, and other (human) road users are fairly good at counteracting other's mistakes by using a combination of proactive and reactive actions. It is therefore important to acknowledge that people do make mistakes and to design ADSs that are resilient to human errors.

Within this scope, the notion of Precautionary Safety (PCS)[36] is introduced and proposes a methodology such that AVs can adjust their trajectory planning to their capabilities, external conditions, and knowledge on human mistakes in order to satisfy overall requirements on accident-, injury- and fatality rates, as summarized in Figure 9. Instead of using the legal safety concept alone, an alternative definition of safe

[33] Nistér, D., Lee, H. L., Ng, J., & Wang, Y. (2019). The safety force field. NVIDIA White Paper.

[34] Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2017). On a formal model of safe and scalable self-driving cars. *arXiv preprint arXiv:1708.06374*.

[35] Pek, C., Manzinger, S., Koschi, M., & Althoff, M. (2020). Using online verification to prevent autonomous vehicles from causing accidents. *Nature Machine Intelligence*, *2*(9), 518-528.

[36] De Campos, G. R., Kianfar, R., & Brännström, M. (2021, September). Precautionary safety for autonomous driving systems: Adapting driving policies to satisfy quantitative risk norms. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)* (pp. 645-652). IEEE. (Paper from the SALIENCE4CAV project).

driving has been introduced as a low accident rate with low severity, no matter whose fault is it. The advantages with the proposed methodology are:

- any existing driving policy can be used as a base;
- any emergency manoeuvre algorithm can be utilized;
- perception capabilities, evasive ability, and knowledge on exposure to risky situations are jointly assessed to identify how the driving policy needs be adjusted to stay safe, and/or when it is safe to activate the AD system.

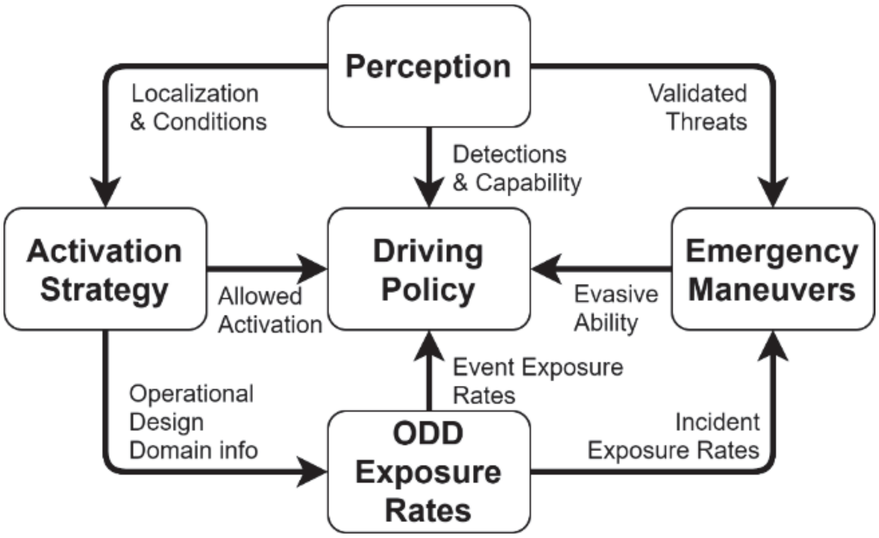


Figure 9 Precautionary Safety Driving Policy for Autonomous Driving, adapting the trajectory planning to the ability to perform evasive manoeuvres.

In this work, as a complement to existing work, the authors propose a structured way to adapt ADS driving policies to satisfy quantitative safety requirements. Simply put, ADSs are proposed to adapt their driving policies to their abilities, such that they can satisfy any given quantitative safety requirement, denoted hereafter as Quantitative Risk Norm (QRN)[37]. The QRN can advantageously be split into QRNs for different accident types and their severity, to ensure that real-world safety is achieved for all types of road users, in any given ODD.

[37] Warg, F., Skoglund, M., Thorsén, A., Johansson, R., Brännström, M., Gyllenhammar, M., & Sanfridson, M. (2020, June). The quantitative risk norm-a proposed tailoring of HARA for ADS. In *2020 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)* (pp. 86-93). IEEE. (Paper from the ESPLANADE project).

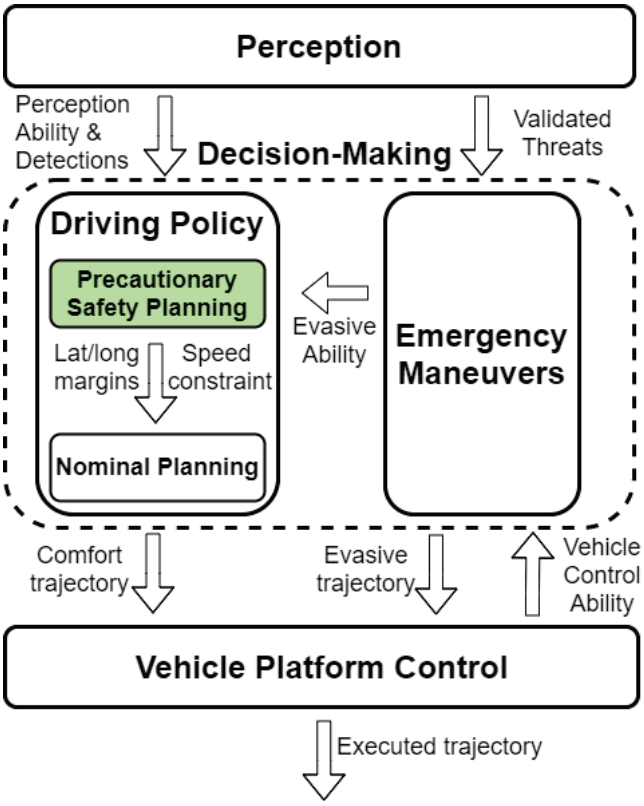


Figure 10 ADS architecture. The Precautionary Safety Planning module is added to guide the Nominal Planning module with additional speed constraints and/or longitudinal/lateral margins to ensure that the Emergency Module is given the prerequisites it needs to satisfy a given QRN.

A high-level illustration of the functional architecture of the proposed Precautionary Safety concept is provided in Figure 10, where each of the operational blocks play a particular role:

Advisor/precautionary safety planning:
To cope with the shortcomings of reactive only systems (many of the existing ADAS and cruising systems), precautionary measures need to be taken by the decision-making module. While precautionary measures are easy to introduce for anticipated human behaviours, there is a need for combined precautionary and reactive measures when handling jaywalkers and other unexpected situations.

The core idea behind the PCS module is to drive with precaution to facilitate collision avoidance/mitigation by emergency manoeuvres in case of unexpected events. Precautionary measures can be defined as a set of advisory inputs to the trajectory planner. In particular, a set of speed constraints and lateral/longitudinal margins can be provided to the planner to reduce the risk for accidents due to unexpected events. But the amount of measures taken by the PCS planner should depend on the ability of the ADS to detect and react to critical situations, and the prescribed QRN requirements. Specifically, the AVs ability to detect and react can be categorized into three main sources:

- perception limitations;
- planning and prediction limitations;
- vehicle control limitations.

Nominal/precautionary planning:
The nominal planner utilizes the surrounding information from perception and advised inputs from PCS planner to determine a smooth, comfortable, and legal trajectory. For instance, in a vehicle following scenario in a highway, the nominal planner should receive information about the road and surrounding objects, together with advised inputs, such as the advised time gap, from the PCS planner. It should then output a trajectory that maintains a sufficiently large distance to the lead vehicle and does not violate the posted speed and ensures that the QRN for rear-end accidents is satisfied.

For example, an adequate driving policy should keep sufficient distance to both ensure that it doesn't collide with the lead vehicle, and only rarely needs to use hard braking, to minimize the risk of being rear-ended.

Emergency/reactive planning:
The emergency or reactive planning module is responsible to exploit the full capability of the vehicle platform to deal with conflicts and to contribute to the fulfilment of the QRN. This is similar to how traditional collision avoidance/mitigation systems, such as AEB, operate, even if ADSs need to handle many more scenarios than traditional AEB systems. It is important to highlight that the emergency planning module is designed to only act in conflict scenarios, whereas the nominal planning module is responsible for interacting with other road users in the first place and is designed to avoid as many conflicts as possible. For the design of a PCS driving policy, it is therefore proposed that the ability of the emergency module to detect and react to unexpected events should be analysed using simulations and directed testing at test tracks, and the outcome used to put precautionary constraints on the nominal planning to ensure that the QRN is satisfied.

The Precautionary Safety concept was later extended in a follow-up paper[38], where a design and monitoring methodology for determining safe driving policies while accounting for perception failure and event exposure rates is presented.

[38] Gyllenhammar, M., de Campos, G. R., Sandblom, F., Törngren, M., & Sivencrona, H. (2022, June). Uncertainty Aware Data Driven Precautionary Safety for Automated Driving Systems Considering Perception Failures and Event Exposure. In *2022 IEEE Intelligent Vehicles Symposium (IV)* (pp. 607-615). IEEE. (Paper from the SALIENCE4CAV project).

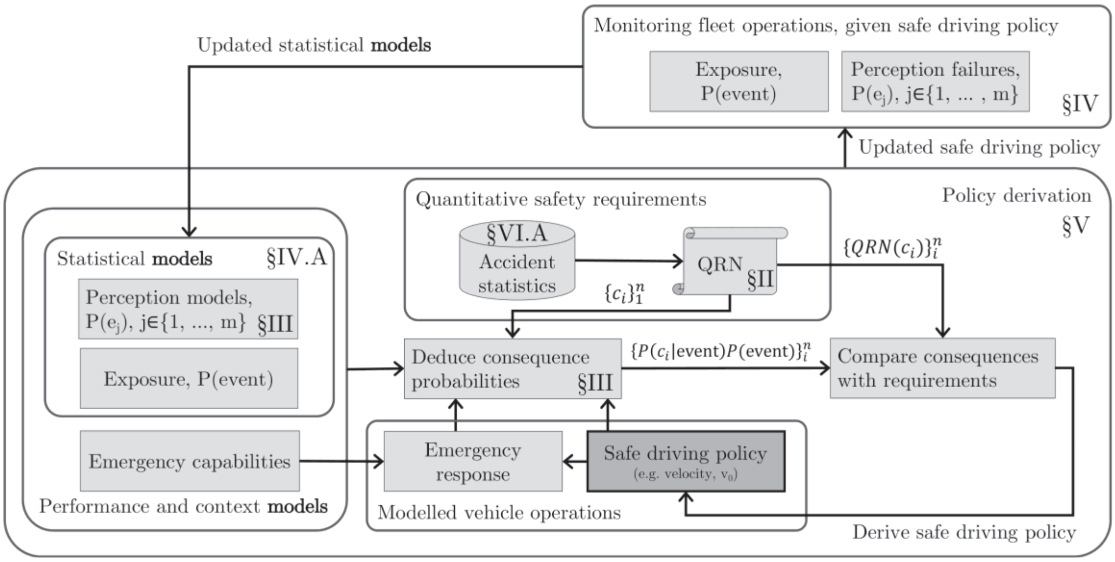

Updated statistical models
Monitoring fleet operations, given safe driving policy
Exposure, P(event)
Perception failures, P(e_j), j∈{1, ..., m}
§IV
Updated safe driving policy
Quantitative safety requirements
Policy derivation
§V
§VI.A
Accident statistics
QRN
§II
{c_i}_1^n
{QRN(c_i)}_i^n
Statistical models
§IV.A
Perception models, P(e_j), j∈{1, ..., m}
§III
Exposure, P(event)
Deduce consequence probabilities
§III
{P(c_i|event)P(event)}_i^n
Compare consequences with requirements
Emergency capabilities
Emergency response
Safe driving policy
(e.g. velocity, v_0)
Performance and context models
Modelled vehicle operations
Derive safe driving policy

Figure 11 Illustration of the proposed methodology for deriving a safe driving policy. The objective is the maximization of the safe driving velocity with maintained confidence that the QRN is met.

In particular, this paper presents a methodology for designing a safe driving policy where uncertainties from underlying statistical models are considered to achieve an uncertainty aware system, see illustrated in Figure 11. The underlying question is how continuously updated estimates of these rates may impact the ADS's safe driving policy. This exploration is achieved by deriving a driving velocity, such that the prescribed safety requirements are fulfilled given the field data and prior knowledge. As before, it is assumed that the safety requirements are provided in the form of a QRN. The safe driving velocity for a given road segment is derived using the estimated arrival rate of an adverse event, in combination with the estimations of the failure rates of the perception system at different range. The rates are modelled as random variables and estimates are calculated based on the posterior distribution of the rates given (simulated) field data. In this way, estimates change as a consequence of receiving more data, and so does the allowable velocity. However, the confidence that the QRN is met remains constant. The contributions of this research are detailed as follows:

- A methodology enabling continuous updates of the ADS's safe driving policy, given field data,
- An uncertainty-aware formulation of the ADS functionality providing a basis for realizing such functions, where i) perception system failures (e.g., false positive detection of free space) and ii) exposure rates with respect to an adverse event, are considered for deriving a safe driving policy fulfilling the safety claims,
- An analysis of the impact on the safe driving policy of the ADS, from the continuous incorporation of field data to update the statistical models underpinning the uncertainty awareness, and
- A numerical evaluation of the proposed methodology on an end-to-end use case, including the derivation of quantitative safety requirements from accident statistics.

For illustration purposes, an example was used considering statistical models of the exposure to an adverse event, as well as failures related to the system's perception system in a traffic scenario including the presence of wild animals. Estimations from these models, using statistical confidence limits, are used to derive a safe driving policy.

The results of this research highlight the importance of leveraging field data to improve the system's abilities and performance. In particular, it exemplifies the use of evidence produced at run-time for design-time activities. The results also stress the importance of incorporating field data into the design and development process of an ADS. It is worth mentioning that while the proposed methodology only considers design-time updates to an ADS, i.e., where all ADSs are implicitly assumed to drive with the same policy until the next update, the methodology could be extended to also consider run-time aspects. For example, the proposed methodology could support dynamic risk assessment, or be used to perform safe risk-aware, tactical decisions. In such a context, the considered statistical models could be extended to include the probabilities conditioned on different operational conditions.

6.2.2 Cooperative and collaborative vehicles

The project has investigated the terminology and taxonomies related to CAVs, comparing road vehicle use cases with cooperative and collaborative AVs and off-road use cases, e.g., using standards such as SAE J3016[39] for terms related to individual automated vehicles, SAE J3216[40] for cooperative driving automation (CDA), ISO 17757[41] for automated earth-moving machinery and mining vehicles, and ISO 18497[42] for automated agricultural machinery and tractors. We propose a unified taxonomy that includes different types of AV interaction: individual, cooperative, and collaborative as well as a simple definition of levels of automation aimed to be more generally applicable to land-based vehicles regardless of domain; it includes manual or assisted operation (non-AV), and supervised or unsupervised automation (AV). The purpose has been to simplify cross-domain knowledge transfer.

Cooperative AVs are defined as multiple vehicles interacting for mutual benefit, each retaining its own individual strategic goal. An example given is automated cars on public roads coordinating passage through an intersection to improve traffic flow. *Collaborative AVs*, on the other hand, are defined as multiple AVs with a common strategic goal, collaborating to complete a joint task. An example is an automated digger loading an automated truck with gravel, where the truck then transports the material to another location. In this case, two collaborating AVs are necessary to complete the task. Table 3 shows the definition of different categories and classes of AVs (the cooperative classes are from J3216, the other definitions proposed by us). The work also discusses cooperative and collaborative ODDs, extends the concept of dynamic driving task (DDT) to the dynamic manoeuvring task (DMT) and

[39] SAE. "J3016:2021 - Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles." https://www.sae.org/standards/content/j3016_202104/.
[40] SAE. "J3216:2021 – Taxonomy and Definitions for Terms Related to Cooperative Driving Automation for On-Road Motor Vehicles". https://www.sae.org/standards/content/j3216_202107/
[41] ISO, "ISO 17757:2019 Earth-moving machinery and mining Autonomous and semi-autonomous machine system safety". https://www.iso.org/standard/76126.html
[42] ISO, "ISO 18497:2018 Agricultural machinery and tractors Safety of highly automated agricultural machines". https://www.iso.org/standard/62659.html

investigates the major ecosystem parts for AVs: the manoeuvring system, the infrastructure, and supervisory/management systems[43].

Table 3 Categories and classes of AVs.

Category	**Class**	**Description**
Individual	Ego-sensing	Depends solely on its own sensors and the ability to make decisions. This may encompass information from static digital infrastructure.
	Connected	Leverages linked services, such as cloud services, dynamic digital infrastructure data, or infrastructure perception, to improve sensing and/or decision-making capabilities.
Cooperative	Status-sharing	AVs disseminate information, such as location, sensor data, or world model, to assist other AVs in their decision-making process (J3216 Class A).
	Intent-sharing	AVs communicate their planned future actions (operational, tactical, strategic) to aid other AVs in their decision-making process (J3216 Class B).
	Agreement-seeking	AVs engage in communication to establish (voluntary) consensus with other AVs, aiming to optimize certain parameter(s) for shared advantage (J3216 Class C).
	Prescriptive	AVs typically operate independently but can adopt specific temporary directive measures to attain an objective set by another entity, such as the road operator (J3216 Class D).
Collaborative	Coordinated	AVs engage in communication to establish consensus on how to collaboratively act to accomplish a shared strategic goal.
	Choreographed	AVs operate independently but are engineered to adhere to a shared global scenario or goal. Unlike coordinated vehicles, they do not depend on communication to execute the collaborative task.
	Orchestrated	AVs are guided by a single entity that acts to accomplish the strategic goal. The guiding entity could be one of the AVs, or an external system, such a traffic management system (TMS).

6.2.3 Minimal risk manoeuvres and minimal risk conditions

Ensuring the safety of CAVs before they hit public roads involves reducing residual risk through design, implementation, and verification. The Operational Design Domain (ODD) is a key tool for confining the V&V effort to the use cases relevant for a particular ADS feature. It also means if an ADS is near exiting its ODD, or if there is a failure in the CAV that makes it unable to continue automated operation, there must be a fallback in place, taking the CAV to a safe state before disengaging. In such situations the ADS would perform a minimal risk manoeuvre (MRM) to

[43] Warg, F., Thorsén, A., Vu, V., & Ebrahimi, H. (2023). A Unified Taxonomy for Automated Vehicles: Individual, Cooperative, Collaborative, On-Road, and Off-Road. *arXiv preprint arXiv:2304.02705*. (Paper from the SALIENCE4CAV project).

achieve a minimal risk condition (MRC). However, when analysing existing literature, some gaps in the definitions and handling of MRCs were found. Figure 12 illustrates how there are two possible top-level strategic goals, either the user defined, which is executed by the nominal DDT, or MRC, which is executed by an MRM. The tactical and operational decisions are made within the constraints set forth by the currently active strategic goal.

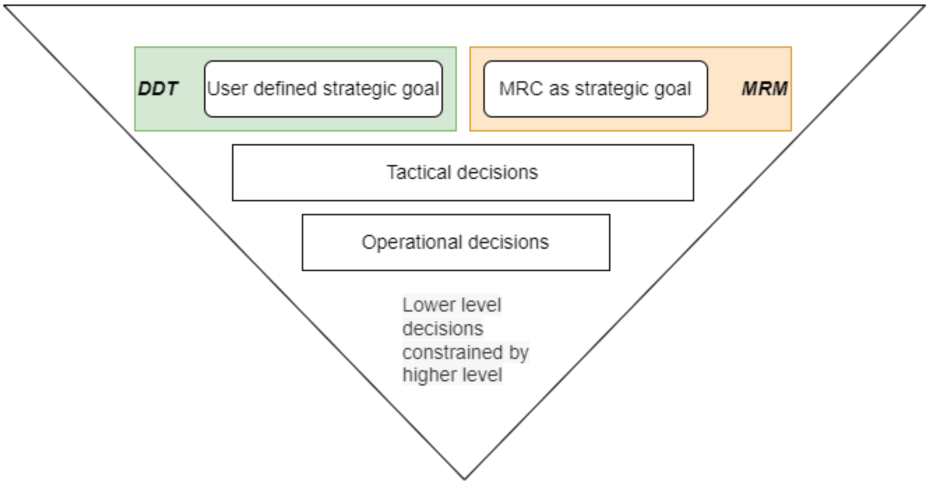

DDT
User defined strategic goal
MRC as strategic goal
MRM
Tactical decisions
Operational decisions
Lower level decisions constrained by higher level

Figure 12 DDT or MRC as strategic goal. Tactical and operational decisions constrained by strategical.

The role of the MRC was investigated in the context of the overall safety argumentation, together with the related subject of managing a loss of capabilities through tactical decisions, such as limiting ADS actions or re-routing to avoid unsuitable conditions. For use cases involving an individually acting CAV, a refined definition (compared to the standard definition from SAE J3016[44]) of MRC was proposed[45]: *"[MRC] is a stable stopped condition at a position with an acceptable risk given the situation when the decision to enter MRC is taken. If an acceptable risk is not attainable, the position with the lowest risk should be selected. The ADS is brought to this state by the user or the system itself, by performing the [fallback], when a given trip cannot or should not be completed."*. This definition calls attention to the need for arguing that a chosen MRC is sufficiently safe; the safety is determined by three components:

1. The frequency to enter the MRC (i.e., how often each type of MRC occurs).
2. The risk of the position selected (e.g., stopping in lane would typically pose a higher risk than stopping in a parking lot).
3. The rate of resolving the MRC (i.e., for how long, on average, will the vehicle remain in the selected stopped position before it can be recovered).

The ADS is configured with a perception block and a decision-making block, which limit tactical decisions to manoeuvres viable with current perception and actuation

[44] SAE. "J3016:2021 - Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles." Retrieved from https://www.sae.org/standards/content/j3016_202104/.

[45] Gyllenhammar, M., Brännström, M., Johansson, R., Sandblom, F., Ursing, S., & Warg, F. (2021). Minimal risk condition for safety assurance of automated driving systems. In *CARS: 6th International Workshop on Critical Automotive Applications: Robustness & Safety*. (Paper from the SALIENCE4CAV project).

capabilities. If the ADS cannot fulfil its DDT—that is the act of operating the vehicle on a tactical and operational level, e.g., steering, braking, and monitoring the environment—given these capabilities, it should abandon the user-defined goal and go to an MRC. Thus, the ADS needs to understand when a sub-optimal capability (due to a temporary or permanent performance degradation) is sufficient for continued operations and be able to cope with any kind of capability degradation. Better self-diagnostics would increase the ability of tactical decisions to keep the strategic goal and adapt, e.g., by changing the route to avoid conditions outside the degraded capabilities, instead of going to MRC.

The concepts of MRM and MRC have also been investigated in the context of cooperative or collaborative CAVs[46], i.e., where multiple vehicles work together as described in Section 6.2.2. If one vehicle malfunctions, the possibilities to halt it while allowing others to continue operating, possibly with reduced performance, are investigated depending on the type of interaction that has been implemented between the CAVs. The aim is to maintain safety while minimizing productivity loss. Two types of MRCs are introduced: *Global MRCs*, which shut down the entire system-of-systems when safety is severely compromised or productivity is no longer possible due to dependencies between collaborating vehicles, and *Local MRCs*, which only affect one or a group of vehicles, allowing the overall operations to maintain some level of productivity. We also introduce the concept of *concerted MRMs*, which are MRMs performed jointly by several AVs to reduce risk during transitional manoeuvres, e.g., one vehicle moving out of the path of a vehicle with degraded capabilities in order to enable it to achieve the best possible MRC with regards to both safety and continued productivity.

Some simulations of interactions between collaborating CAVs have been performed, e.g., to explore the use of concerted MRMs. Figure 13 shows a snapshot from a simulation where one mining vehicle leaves a tunnel to clear the way for another vehicle which may have suffered some failure forcing it to abandon its task. If the path can be cleared to allow the vehicle to leave the working area rather than stopping the machine inside the tunnel, a stop in production may be avoided.

[46] Vu, V., Warg, F., Thorsén, A., Ursing, S., Sunnerstam, F., Holler, J., Bergenhem, C. & Cosmin, I. (2023, June). Minimal Risk Manoeuvre Strategies for Cooperative and Collaborative Automated Vehicles. In *2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)* (pp. 116-123). IEEE. (Paper from the SALIENCE4CAV project).

Figure 13 Screenshot from simulation of concerted MRM (Simulation by Epiroc).

6.3 Safe machine learning

6.3.1 Out-of-distribution detection

Out-of-distribution (OoD) detection refers to the identification of data samples that deviate from the distribution that has been seen during the training of a ML model. In the context of automated driving, or safety critical systems, OoD detection can be a crucial part by supporting the system with identifying situations or objects that the model is not familiar with. To exemplify, an AV may encounter road signs that it has not seen before, and perhaps misunderstanding the meaning of the sign, thereby pose a risk. OoD detection and similar outlier detection methods can then alert the system to novel scenarios, suggesting that the system proceed with caution instead of proceeding as usual.

In our paper "*Evaluation of Out-of-Distribution Detection Performance on Autonomous Driving Datasets*"[47], we focus on examining the performance of deep neural networks (DNNs) in handling of OoD samples in the context of automated driving. The study tests Mahalanobis distance (MD) as a metric for OoD detection in semantic segmentation DNNs. The key result is the identification of a trade-off between reducing misclassification risks and maintaining pixel coverage. This trade-off is crucial because overly conservative OoD detection (high sensitivity to OoD samples) can lead to the rejection of many pixels, reducing the model's usefulness in real-world applications. Conversely, less strict OoD detection might not adequately safeguard against misclassifications of critical objects, such as pedestrians.

[47] Henriksson, J., Berger, C., Ursing, S., & Borg, M. (2023, July). Evaluation of Out-of-Distribution Detection Performance on Autonomous Driving Datasets. In *2023 IEEE International Conference On Artificial Intelligence Testing (AITest)* (pp. 74-81). IEEE. (Paper from the SALIENCE4CAV project).

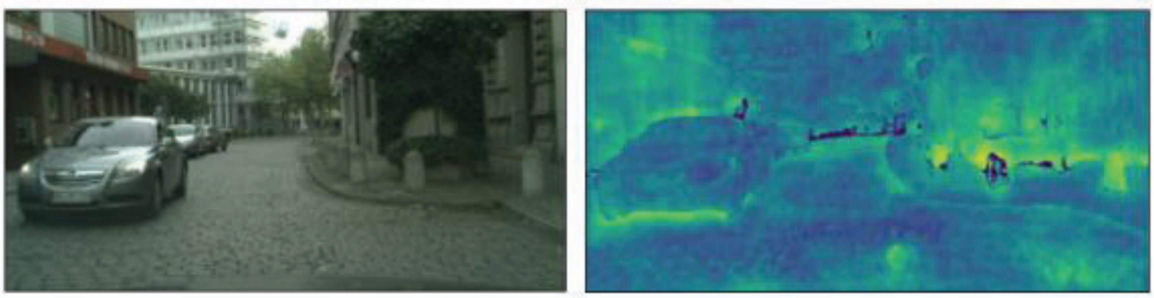


Figure 14 A Visualization of the Mahalanobis Distance, applied to a front looking camera. Brighter colours refer to higher uncertainty of the predicted pixel.

The paper presents risk-coverage as a trade-off, where a safety case can be constructed based on an accepted level of risk of pixel errors and shows a reduction of risk by reducing the amount of allowed predictions done to pixels. By extending the system with an exclusion criterion, the automotive systems can be designed with a target error rate to be more reliable and safer. In practical terms, this means an AVs perception system could effectively detect and react to potentially hazardous situations it wasn't explicitly trained to handle, thereby enhancing overall safety. The research underscores the importance of incorporating robust OoD detection mechanisms as part of the safety measures in AV technology, ensuring they can operate safely in diverse and unforeseen environments.

In follow-up work[48], we analyse and discuss how OoD detection can be used in three different phases of ADS development for safety-related purposes: (1) In the development phase used to identify limitations in the training dataset through highlighting scenarios where the detection rate is low (alternatively suggest ODD reduction); (2) For use in shadow-mode, i.e., when a function is active in production vehicles to test its capabilities but not allowed to interfere in decision-making or affect the actuators, as a way to test the expansion of ODD boundaries and highlight where more training data is needed; (3) in the operational phase to help identify uncertainties in the ML model and trigger safe fallback if the uncertainty goes above a defined threshold.

6.3.2 Impact of edge cases

ML-enabled approaches are considered to substantially support the detection of obstacles and traffic participants around an AV. State-of-the-art network like YOLO (you-only-look-once) provide bounding boxes around detected objects including a classification and a confidence level that can be used for, e.g., trajectory planning. Training is typically based on high quality annotations to provide the ground truth data. However, if traffic participants like pedestrians or bicyclists are partially occluded, e.g., because they are carrying objects, ML-enabled systems may be

[48] Henriksson, J., Ursing, S., Erdogan, M., Warg, F., Thorsén, A., Jaxing, J., Örsmark, O. & Toftås, M. Ö. (2023, April). Out-of-Distribution Detection as Support for Autonomous Driving Safety Lifecycle. In International Working Conference on Requirements Engineering: Foundation for Software Quality (pp. 233-242). Cham: Springer Nature Switzerland.

challenged, or relevant objects missed entirely. In this work[49], the impact of systematically challenging a neural network (NN) was investigated by feeding edge cases caused by partial occlusions of pedestrians in video frames from the KITTI dataset. Conclusions was firstly that the training of such ML-enabled systems should be adjusted to be more robust to disturbance effects, secondly, an ML-enabled system may contain a cascade of multiple NN that are all trained differently to consult in cases of low confidence of the primary NN, and thirdly, a separately trained NN also contribute to the explainability of NN to get indications why there is a drop in the detection performance.

6.3.3 Simulation-aided approach to safety analysis of learning-enabled components

Artificial Intelligence (AI) techniques through Learning-Enabled Components (LEC) are widely employed in Automated Driving Systems (ADS) to support operation perception and other intelligent driving tasks relating to planning and control. Therefore, risk management becomes a critical aspect in the system development as well as in the safety engineering. The challenge arises however due to the inherent stochasticity of LEC algorithms, which in combination with the complexity of operational conditions makes it difficult to assess the system failure logic or to estimate the hazardous events. To address this issue, this work[50] is focused on the development of a simulation-aided approach to the identification of fault behaviours of LEC through a framework as shown in Figure 15. This framework consists of the following services:

- A simulation-aided operational data generation service with the operational parameters extracted from the corresponding system models and specifications.
- A Fault Injection (FI) service aimed at high-dimensional sensor data to evaluate the robustness and residual risks of LEC.
- A Variational Bayesian (VB) method for encoding the collected operational data and supporting an effective estimation of the likelihood of operational conditions.

[49] Henriksson, J., Berger, C., & Ursing, S. (2021, September). Understanding the impact of edge cases from occluded pedestrians for ML systems. In 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA) (pp. 316-325). IEEE.

[50] Su, P., Warg, F., & Chen, D. (2023). A Simulation-Aided Approach to Safety Analysis of Learning-Enabled Components in Automated Driving Systems. In *26th IEEE International Conference on Intelligent Transportation Systems (ITSC 2023)*.

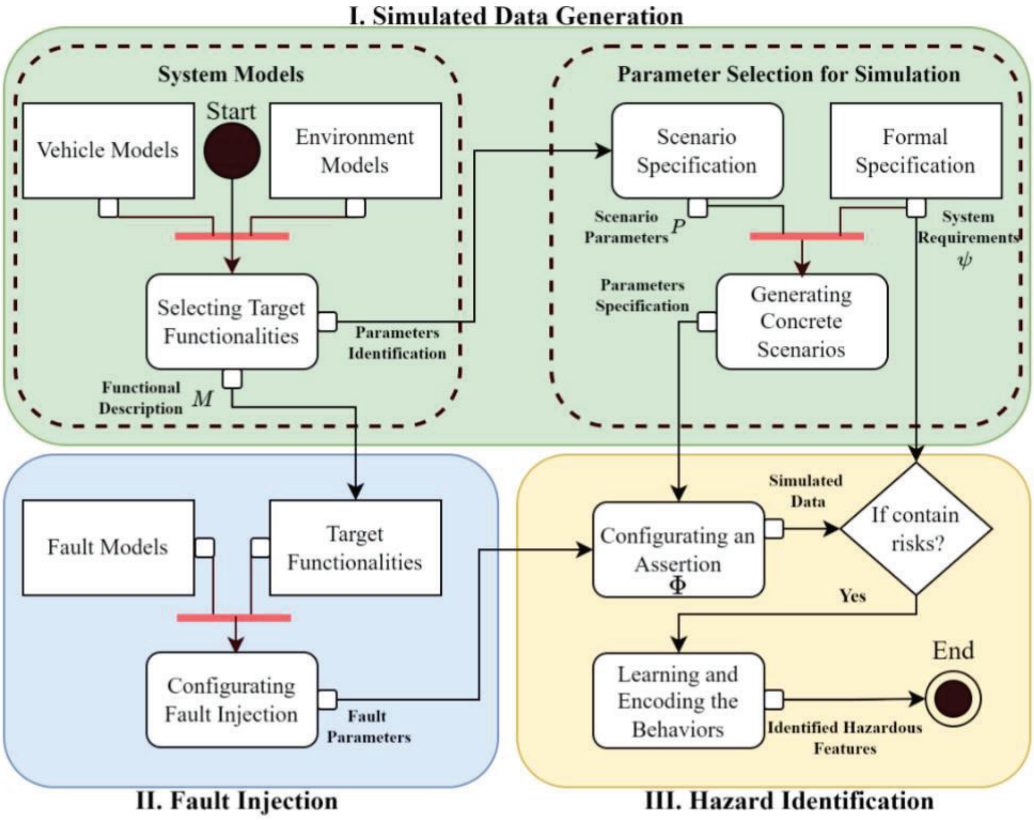


Figure 15. A simulation-aided approach to the identification of fault models for LEC based on FI.

Within this framework, the system models are used to specify the target ADS and its operational environments, supported by the domain-specific language EAST-ADL. An additional scenario description method is also used to capture the related operational conditions for the configuration of simulations. The configurations for fault injection-based simulation cases are given by combinations of such system and environmental parameters. The faults are injected according to a systematic reasoning about the system models, failure modes, and the location and intensity of their occurrences. The system emissions by fault injection operation are encoded by Variational Auto-Encoder (VAE), which is a generative deep learning model combining auto-encoder (AE) and probabilistic models for unsupervised classification of the fault conditions.

The implementation is support by a platform, shown in Figure 16, which consists of three parts, implemented by two computers and one Jetson Nano node respectively. The first part (Computer1) provides a virtual traffic and road environment, virtual sensors for vehicles (radar, camera, etc.) based on the open-source simulator CARLA[51]. A virtual control system is responsible for controlling the ego vehicle operating in this virtual environment. The second part is an embedded node responsible for implementing the LEC for object detection, object tracking, and distance estimation based on a Python and CUDA environment in Nvidia Jetson Nano. This node communicates with Computer1 by RTMP for video streaming. The third part (Computer2) is responsible for fault injection based on a hardware

[51] http://carla.org/

perspective of LEC, regarding in particular the weight parameters of target deep neural networks.

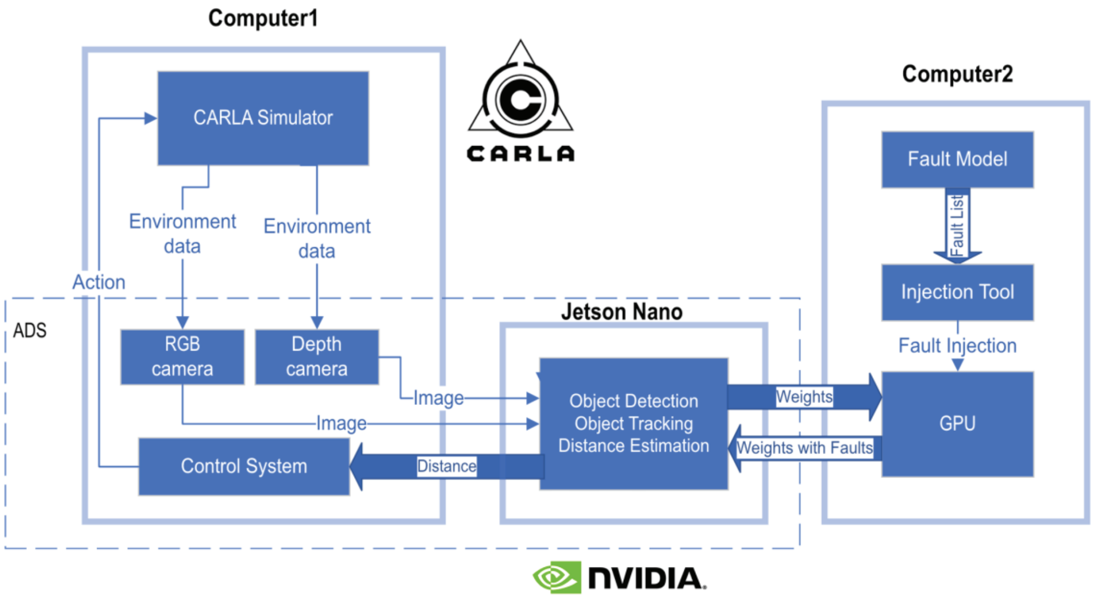
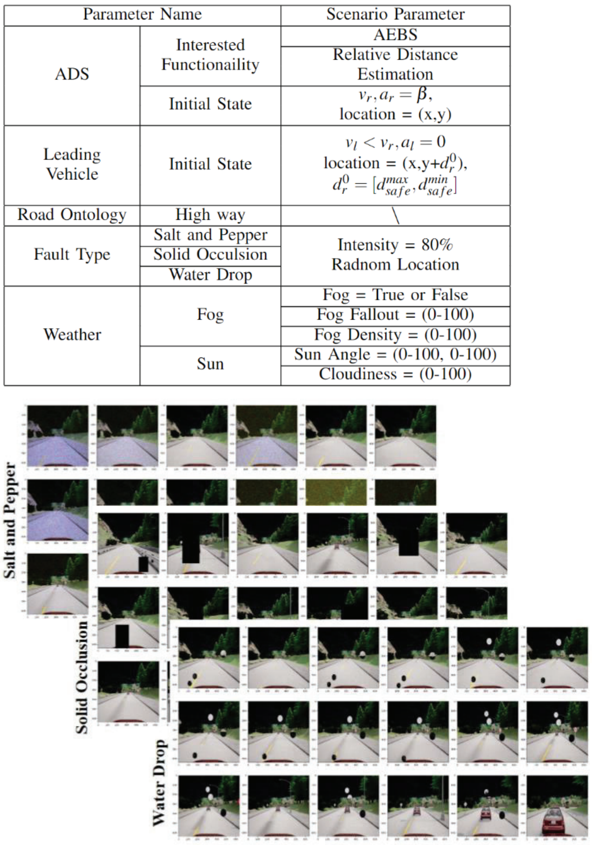


Figure 16. The architecture of the fault injection simulation system.

As a case study, the behaviour of an Autonomous Emergency Braking (AEB) system with camera-based operation perception is conducted. A set of fault types of cameras, including solid occlusion, water drop, salt and pepper, are modelled and injected into the perception module of the AEB system in different weather conditions. See Figure 17. The results indicate that this framework enables to identify the critical faults under various operational conditions.



Parameter Name		Scenario Parameter
ADS	Interested Functionaility	AEBS
		Relative Distance Estimation
	Initial State	$v_r, a_r = \beta$, location = (x,y)
Leading Vehicle	Initial State	$v_l < v_r, a_l = 0$ location = (x,y+d_r^0), $d_r^0 = [d_{safe}^{max}, d_{safe}^{min}]$
Road Ontology	High way	\
Fault Type	Salt and Pepper	Intensity = 80% Radnom Location
	Solid Occulsion	
	Water Drop	
Weather	Fog	Fog = True or False
		Fog Fallout = (0-100)
		Fog Density = (0-100)
	Sun	Sun Angle = (0-100, 0-100)
		Cloudiness = (0-100)

Figure 17. An example of simulation case and related faults of camera being injected.

6.4 Variational safety across product-line and product lifecycle

Many of the ADSs features are mission- and safety critical, while also being expected to have a long lifetime, often up to 20 years. As ADSs are complex products, they involve parts from multiple suppliers and exist in multiple variants. Therefore, the management of multiple variants (and product lines) and integration of components from different suppliers are critical issues.

6.4.1 Product line engineering

For successful management of product variations, it is necessary to manage the traceability and (non)conformities of changes across product lifecycle and product-line regarding some local dependability constraints or the overall operational risks. Nevertheless, the complexity of interdependences among the environmental assumptions and the configurations of functional and technical safety concepts often makes any approach without a systematic variability management ineffective and error prone. Fundamentally, the success relies on the consolidations of heterogeneous technologies, multidisciplinary knowledge, and engineering efforts. One key issue is related to the management and reconciliation of the impacts of product-line variability, actual changes as well as their unexpected side effects. The combination of CI and CD creates connections from system development directly with actual system operation feedback for product maintenance and improvements.

The project work in this regard investigates the challenges and novel concepts for establishing a framework for effective management of the variability of operational safety. The goal is to enable an effective management of the operational safety concerns both product-line and product lifecycle. Figure 18 illustrates how safety release cycles are related to software release cycles. Target for CI/CD is to have a complete safety case at the end of the integration phase assuring the safety of the improvements made to the vehicle functionality during that specific development cycle.

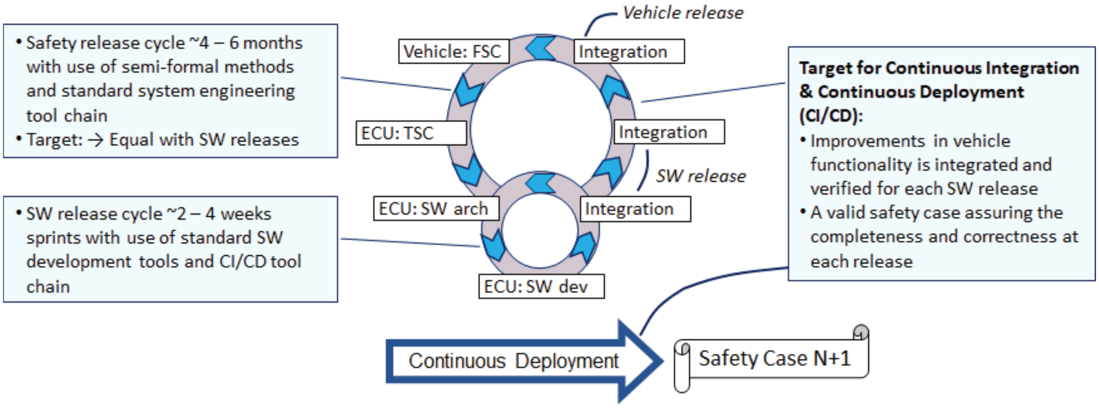

• Safety release cycle ~4 – 6 months with use of semi-formal methods and standard system engineering tool chain
• Target: → Equal with SW releases
• SW release cycle ~2 – 4 weeks sprints with use of standard SW development tools and CI/CD tool chain
Vehicle release
Vehicle: FSC
Integration
ECU: TSC
Integration
SW release
ECU: SW arch
Integration
ECU: SW dev
Target for Continuous Integration & Continuous Deployment (CI/CD):
• Improvements in vehicle functionality is integrated and verified for each SW release
• A valid safety case assuring the completeness and correctness at each release
Continuous Deployment
Safety Case N+1

Figure 18 Illustration of Safety Case continuous deployment development cycles.

6.4.2 Key system parameters, variation points and their interdependencies

The variation points characterizing variational safety concepts across product-line and product lifecycle are mainly product or product-line specific. The definitions are dominated by the system architectural parameters, constraints, and V&V measures that define the variability of safety concepts, their interdependencies, and binding criteria. These parameters encompass the specific attributes and characteristics of the safety goals (SG) and safety concepts (i.e., FSC and TSC), relating to operational environments, functional decomposition, hardware allocation, safety increments, and the design of hardware platforms (e.g., ECU) that vary across the product-line and product lifecycle. Safety goals are further related to the operational scenarios and design decisions that impact the design and selections of safety concepts. A systematic specification of these variation points is essential for understanding the sensitivity of updates or changes, such as due to emergent operational environments, reconfigurations of system safety functions, and integrations of alternative components. The complexity is elaborated in Figure 19. Important aspects to handle include:

- A change/addition in functionality could be at any level of abstraction or architectural layer of the system or sub-system.
- Changes affects properties at various hierarchical levels in a very complex relationship.
- Applying component-based design, analyse safety properties, assert safety contracts, and model relevant dependencies by tool support would master the complexity.
- Safety case can be generated continuously in a CI/CD build chain based on the aggregation of safety-contract fragments.

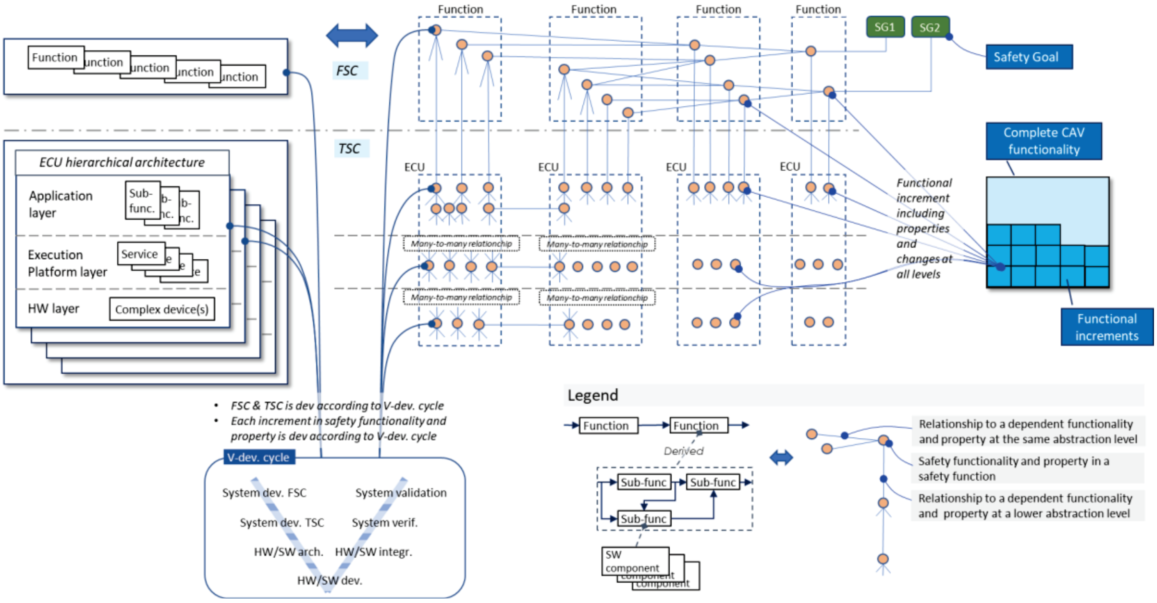

Function
Function
Function
Function
SG1
SG2
Safety Goal
FSC
TSC
ECU
ECU
ECU
ECU
ECU hierarchical architecture
Application layer
Sub-func.
Execution Platform layer
Service
HW layer
Complex device(s)
Many-to-many relationchip
Many-to-many relationchip
Many-to-many relationchip
Many-to-many relationchip
Complete CAV functionality
Functional increment including properties and changes at all levels
Functional increments
FSC & TSC is dev according to V-dev. cycle
Each increment in safety functionality and property is dev according to V-dev. cycle
V-dev. cycle
System dev. FSC
System validation
System dev. TSC
System verif.
HW/SW arch.
HW/SW integr.
HW/SW dev.
Legend
Function
Function
Derived
Sub-func
Sub-func
Sub-func
SW component
Relationship to a dependent functionality and property at the same abstraction level
Safety functionality and property in a safety function
Relationship to a dependent functionality and property at a lower abstraction level

Figure 19 Continuous deployment in agile development – Variability & dependency complexity of safety functionality.

6.4.3 Variability modelling and decision binding

Several variability modelling techniques are available for capturing the variation points and binding decisions. For example, EAST-ADL provides the support for feature modelling with well-managed design-space information and configuration decisions. There are also more generic variability modelling techniques available, including Cardinality Based Feature Modelling (CBFM), Common Variability Language (CVL), and a UML-based modelling methodology SimPL. These techniques could be useful for supporting generic variability configuration.

For ADS, a variability model should especially allow a formal reasoning about the sensitivity of variations regarding safety concepts. The key is related to the modelling and assessment of interdependencies across safety goals and safety concepts, i.e., how variations in the requirements, constraints, mechanisms, resources, and V&V measures affect each other. This also involves the usage of logic-based languages and formal methods to analyse and reason about the implications of variabilities. Statistical methods and other data-driven techniques become important for quantifying the effects of changes in the operational environment or system configurations. Moreover, each binding of variational safety concepts generates some specific contracts regarding functional and technical safety concepts. Each binding decision involves a systematic reasoning about the assumptions of system operation and related safety goals in terms of both original and expanded ODDs, the compliances of variational functional and technical safety concepts.

For managing complex interdependencies across continuous integration and development cycles, a novel concept, referred to as High-Dimensional Safety Contracts, has been proposed to enable a modularized specification of compliances beyond basic functional operation across the lifecycle stages. In particular, for safety assurance over various maintenance and evolution cycles, the contract mechanism contains an extension for additional dimensions of compliance, including failure modes, safety mechanisms, and expected V&V measures. This offers several advantages over conventional component specifications. Existing tools for contract-based design such as OCRA[52], CHASE[53], AGREE[54] can possibly be extended to provide support for automatic generation and management of such contracts, ensuring consistency and accuracy of variations.

6.5 Goal fulfilment

6.5.1 Deliverables and project objectives

The defined deliverables in the project application were the following four reports reflecting the topic areas and expected method development defined in Section 5:

- *Report on safety assurance for CAVs*
- *Report on safety design for CAVs*
- *Report on safety of ML in CAVs*
- *Report on operational safety in CAVs*

The reports have been written, to a large extent also referring to the published research papers containing most of the project results.

Additional deliverables were A *dissemination plan* with the dissemination strategy and listing of publications, talks, and events. A *public website for the project*, which has been created and been publicly available throughout the duration of the project (http://salience4cav.se/), a *project public report* (this document), and a *final seminar and demonstration of results*. Due to time constraints, we were unable to hold a final seminar/demonstration before the project end. However, the project results will be presented at the SAFER stage 5 final event[55], and results have previously been presented through several talks/events during the project.

Additional project objectives were: (1) to *contribute to the research for an industrial Ph.D. on the topic of ML*. This has been achieved since the Ph.D candidate successfully defended the thesis[56] at the end of the project. In addition, the project has contributed to the research of a second industrial Ph.D. candidate who held a halfway seminar[57] during the project and will continue to pursue the Ph.D. in other projects, and it provided one study for a university Ph.D. student; (2) *dissemination of project results through scientific publications and talks,* with an aim of at least 7 published or submitted papers. At the end of the project, 12 talks/events have been held and 15 publications (11 peer-reviewed and published, 2 submitted papers, and 2 additional reports) produced; (3) it was expected that some project results are *relevant for*

[52] https://ocra.fbk.eu/
[53] https://chase-cps.github.io/chase/
[54] http://loonwerks.com/tools/agree.html
[55] https://www.saferresearch.com/ (stage 5 final event to be held on 2024-03-08).
[56] Henriksson, J. (2023). Outlier Detection as a Safety Measure for Safety Critical Deep Learning. https://research.chalmers.se/en/publication/537689
[57] Gyllenhammar, M. (2022). *Efficient Strategies for Safety Assurance of Automated Driving – half-time PhD seminar and discussion.*

standardization—during the course of the project, project members have been active in particular in the development of the upcoming *"ISO/CD TS 5083 Road vehicles– Safety for automated driving systems"* and *"ISO/CD PAS 8800 Road Vehicles—Safety and artificial intelligence"*, providing both information on the ongoing standardization and benefitting from the knowledge and results gained in the project.

Positive deviations have more publications and talks that the original aim, contribution to the research of three Ph.D. students instead of one, and early benefit of project results in standardization. A negative deviation is a lack of published results in the area of continuous/quantitative assurance cases, where the work has taken longer than anticipated. We aim at one or two publications after the project end (work has started but is not completed at the time of writing this report) and continuation of this work within other projects.

6.5.2 Contribution to the FFI program

This project started in January 2021. The main goals of the FFI program at the time were to reduce the environmental impact of road traffic, reduce the number of injured and killed in traffic, and increased competitiveness of the Swedish vehicle industry[58]. The aim of the sub-program *road safety and automated vehicles* was to contribute to increased automation in the transport sector, including aspects such as efficiency and environmental friendliness, as well as contribute to vision zero (no deaths or serious injuries in traffic).

The main goals and results of the SALIENCE4CAV project are related to safety of AVs, which is mainly in line with the goal of reducing injuries in traffic. We have contributed with methods for safety analysis and safety assurance, e.g., the work on precautionary safety, human interaction safety analysis, and safety of ML. The work on cooperative and collaborative vehicles also partly contribute to the goal of increased efficiency, as one target has been to enable increased use of automated vehicles, e.g., in confined areas with mixed traffic, and to reduce the productivity losses due to faults in one constituent AV in a collaborative task. We also believe we have contributed to increased competitiveness through the work on enabling a more efficient way-of-working with continuous deployment and variability management, but also by: conducting research in an advanced area expected to contribute to growth in the vehicle industry; promoted cross-domain (road and mine vehicles) knowledge transfer, collaboration between OEMs, Tier 1 suppliers, service providers, research institute and academia; contributed to the work of three Ph.D. students; as well as increased the competence in the area of automation for individuals from the participating partners.

In addition, the FFI programme shall promote equality within vehicle research and development, as well as support the sustainable development goals set out in Agenda 2030[59]. We believe the work contributes to the agenda 2030 goals *3.6, reduce fatalities and injuries in road traffic*, goal *9.5 build resilient infrastructure, promote*

[58] https://www.vinnova.se/globalassets/mikrosajter/ffi/dokument/ffi-fardplan-2019.pdf
[59] https://www.globalamalen.se/

inclusive and sustainable industrialization and foster innovation, and *11.2, provide access to safe, affordable, accessible and sustainable transport systems for all,* as AVs hold the potential for safer, more accessible—e.g., to be used by persons not able to drive—and more efficient—e.g., by increased cooperation and vehicle sharing—transportation. The field is male dominated, but in this project 6 women have participated in the research, two of them contributing to research papers for the first time.

Since the project started, a new FFI roadmap has been published (2023[60]). In this roadmap, the new main goals are that FFI has: *demonstrated solutions that make society's road transport fossil-free, safe, equality and efficient; developed sustainable solutions that have been implemented and accepted by users and society*; and *contributed, through innovation, partnership and collaboration, to the development of skills, infrastructure, policy, regulatory frameworks and business models in the road transport system*. While not directly considered in this project as it started before the new roadmap was published, we note that the results also fit well with the new goals regarding safety, equality, and efficiency in the transport sector, and to contribution to new skills through partnership and collaboration.

[60] https://www.vinnova.se/globalassets/mikrosajter/ffi/dokument/fardplan/ffi-roadmap.pdf?cb=20230615135940

7 Dissemination and publications

7.1 Dissemination

How are the project results planned to be used and disseminated?	Mark with X	Comment
Increase knowledge in the field	X	Contributions published in open access papers listed in Section 7.2 and disseminated in 12 talks, seminars, and workshops.
Be passed on to other advanced technological development projects	X	Work on several topics continue in other research projects that have already started, e.g., SUNRISE—Validation of AVs (RISE, AV safety assurance), TADDO—prestudy for reliable DevOps (Qamcom, safety contracts for safe CD, DevOps), FAMER—safe perception systems with ML (RISE, Zenseact, safety of ML).
Be passed on to product development projects	X	Project members also part of development teams bring knowledge back directly to development.
Introduced on the market		
Used in investigations / regulatory / licensing / political decisions	X	Participation and input to standardization based on project results.

7.2 Publications

The following publications are based on project results.

Title	Authors	Venue for publication	PR[61]
Understanding the Impact of Edge Cases from Occluded Pedestrians for ML Systems	Jens Henriksson, Christian Berger, Stig Ursing	Euromicro Conference on Software Engineering and Advanced Applications (SEAA 2021)	X
Precautionary Safety for Autonomous Driving Systems: Adapting Driving Policies to Satisfy Quantitative Risk Norms	Gabriel Rodrigues de Campos, Roozbeh Kianfar, Mattias Brännström	24th IEEE International Conference on Intelligent Transportation (ITSC 2021)	X
Minimal Risk Condition for Safety Assurance of Automated Driving Systems	Magnus Gyllenhammar, Mattias Brännström, Rolf Johansson, Fredrik Sandblom, Stig Ursing, Fredrik Warg	6th International Workshop on Critical Automotive Applications: Robustness & Safety (CARS) at EDCC 2021	X
ADS Safety Assurance Methods - Future Directions	Magnus Gyllenhammar, Carl Bergenhem, Fredrik Warg	6th International Workshop on Critical Automotive Applications: Robustness & Safety (CARS) at EDCC 2021	X
Human Interaction Safety Analysis Method for Agreements with Connected Automated Vehicles	Fredrik Warg, Martin Skoglund, Matthew Sassman	IEEE 94th Vehicular Technology Conference (VTC 2021-Fall)	X

[61] An ’X’ in the column ’PR’ indicates a peer-reviewed publication.

Developing SEooC – Original Concepts and Implications when Extending to ADS	Rolf Johansson and Håkan Sivencrona	6th International Workshop on Critical Automotive Applications: Robustness & Safety (CARS) at EDCC 2021	X
Uncertainty Aware Data Driven Precautionary Safety for Automated Driving Systems Considering Perception Failures and Event Exposure	Magnus Gyllenhammar, Gabriel Rodrigues de Campos, Fredrik Sandblom, Martin Törngren and Håkan Sivencrona	33rd IEEE Intelligent Vehicles Symposium (IV 2022).	X
Holistic Perspectives on Safety of Automated Driving Systems - Methods for Provision of Evidence	Magnus Gyllenhammar, Gabriel Rodrigues de Campos, and Martin Törngren	Preprint DOI: 10.36227/techrxiv.20331243.v1	
Out-of-Distribution Detection as Support for Autonomous Driving Safety Lifecycle	Jens Henriksson, Stig Ursing, Murat Erdogan, Fredrik Warg, Anders Thorsen, Johan Jaxing, Ola Örsmark, and Mathias Örtenberg Toftås	29th International Working Conference on Requirement Engineering: Foundation for Software Quality (REFSQ 2023)	X
Minimal Risk Manoeuvre Strategies for Cooperative and Collaborative Automated Vehicles	Victoria Vu, Fredrik Warg, Anders Thorsén, Stig Ursing, Fredrik Sunnerstam, Jimmy Holler, Carl Bergenhem, Irina Cosmin	9th International Workshop on Safety and Security of Intelligent Vehicles (SSIV), held in conjunction with DSN 2023	X
Evaluation of Out-of-Distribution Detection Performance on Autonomous Driving Datasets	Jens Henriksson, Christian Berger, Stig Ursing and Markus Borg	The 5th IEEE International Conference on Artificial Intelligence Testing (AITest 2023)	X
A Simulation-Aided Approach to Safety Analysis of Learning-Enabled Components in Automated Driving Systems	Peng Su, Fredrik Warg, DeJiu Chen	26th IEEE International Conference on Intelligent Transportation Systems (ITSC 2023) - Workshop on Beyond Traditional Sensing for Intelligent Transportation	X
Managing continuous assurance of complex dependable systems	Fredrik Warg, Omar Jaradat, Anders Cassel, Dejiu Chen, Negin Nejad, Anders Thorsén, Stig Ursing	Report based on workshop held at 10th Scandinavian Conference on System and Software Safety (SCSSS 2022)	
A Unified Taxonomy for Automated Vehicles: Individual, Cooperative, Collaborative, On-Road, and Off-Road	Fredrik Warg, Anders Thorsén, Victoria Vu, Helen Ebrahimi	Preprint DOI: 10.48550/arXiv.2304.02705	
Choosing Risk Acceptance Criteria for an Automated Driving System	Fredrik Sandblom, Gabriel Rodrigues de Campos, Peter Hardå, Fredrik Warg, Fredrik Beckman	*Submitted for review*	

Work on two additional papers based on the work described in Section 6.1.3 and 6.4 respectively, with the tentative titles *"Safety Contracts for Decoupling of Complex Interdependencies across Continuous Integration and Development Cycles of Automotive Systems,"* and *"Leveraging System Ontology for Effective Synthesis and Analysis of Safety Contracts with Probabilistic Models for Automated Driving Vehicles"* are ongoing with the aim to finalize and publish after the end of the project.

8 Conclusions and future research

The development of automated vehicles is continuing as OEMs are pursuing different business models, focusing on diverse features such as collaborative vehicles in confined areas, convenience functions for passenger cars, shuttle buses, or robotaxi services, with different levels of automation, and in some cases complemented with remote assistance. However, a common challenge is safety assurance and a need to continuously improve the systems. This report summarises the results from the FFI project SALIENCE4CAV, which aimed to contribute to some of the challenging issues related to safety assurance for automated vehicles, and in particular to methods suitable for enabling frequent updates of automated driving systems.

While this, and many other projects, have contributed to increased knowledge regarding safety assurance for AVs, there are still open questions. Some areas of future work identified within the topics addressed by SALIENCE4CAV are:

- Closing the DevOps loop for safety contracts, allowing the same mechanism to be used all the way from defining the ODD/traffic environment to the use of field data as feedback to the next development cycle.
- The practical obstacles which still make use of safety contracts difficult, e.g., toolchains, introduction in the design flow, efficient formulation of contracts and refinement between abstraction levels, etc.
- How to best combine the benefits from existing safety assurance methods to tackle all ADS safety assurance challenges identified in the context of continuous assurance.
- Develop useful safety patterns for cooperative and collaborative vehicles accounting for varying level of availability of infrastructure support and supervisory systems.
- Management of collaboration of heterogeneous AVs (e.g., different capabilities, ODDs, manufacturers), including cooperative strategies for MRC and ODD monitoring.
- Scheduling of AVs in confined areas for management of dynamic autonomous operating zones.
- General criteria to derive safety requirements and corresponding metrics for ML components.
- How out-of-distribution detection, and other potential ML safety measures impacts system-level technical requirements. The effectiveness of ML methods and measures need to be better understood.
- Ability to better use simulation for evaluation of ML safety requirements and effectiveness of measures.
- Continued work on tool framework and case study for variational safety concepts.

9 Participating parties and contact persons

List of contact persons for participating organizations:

Name	Email	Partner Affiliation
Fredrik Warg (project coordinator)	fredrik.warg@ri.se	RISE Research Institutes of Sweden
Jan Pålsson	jan.palsson@agreat.com	Agreat
Ola Örsmark	ola.orsmark@comentor.se	Comentor
Jimmy Holler	jimmy.holler@epiroc.com	Epiroc Rock Drills
DeJiu Chen	chen@md.kth.se	KTH Royal Institute of Technology
Carl Bergenhem	carl.bergenhem@qamcom.se	Qamcom Research and Technology
Stig Ursing	stig.ursing@semcon.com	Semcon Sweden
Fredrik Beckman	fredrik.beckman@magna.com	Magna (formerly Veoneer)
Gabriel Rodrigues de Campos	gabriel.campos@zenseact.com	Zenseact