Complementary Motivations for Blank Cells in Interconnections Analysis to main paper ”The Safety Life Cycle of Automated Driving Systems and Interdependences of Development Methods”

Magnus Gyllenhammar, Gabriel Rodrigues de Campos, and Martin Törngren, *Senior Member, IEEE*

***Abstract*—This is a complementary document to the main paper ”*The Safety Life Cycle of Automated Driving Systems and Interdependences of Development Methods*”. Here the blank cells in the interconnections classifications, presented in the main paper, are motivated. The main paper thus captures the connections between the development methods, whereas this complementary document motivates the cases where the absence of a connection is indicated in TABLE I of the main paper.**

I. The Development Methods

The main paper presents a safety life cycle framework and proceeds to map the 16 development methods identified in [1] onto this framework. The interconnections between the methods are analysed and classified, including the lack of interconnections. The classification includes 240 different cells resulting from analysing the interconnections between the 16 methods with each other (subtracting self-connections between the methods). Out of these, 116 are provided with a classification of the connection and 124 are blank – suggesting a lack of connection between the methods. The 16 methods from [1], are visualised in Fig. 1. Note that *process arguments* category has been removed due to its inherent close connection to the development phase of the life cycle and its lack of informative connections to the other development methods. The following section motivates the 124 blank cells in the classification of TABLE I of the main paper, while the connections are instead motivated in the main paper.

II. Motivating the Absence of Interconnections Between Methods

Contract-based Design (CBD), is used for the design phase and the system design in particular and does consequently not influence the Operational Design Domain (ODD) nor the Hazard Analysis and Risk Assessment (HARA) methods. While the system can be composed according to a contract, the definitions of the ODD and applicable hazards will not be facilitated by the use of CBD. Also the runtime certification (RT Cert- .) method lack indirect connection to the ODD. However, it might be valuable as a supportive tool for the formalisation of restricted operational domains (RODs), which is instead captured by the connection between RT Cert. and degradations.

Similarly, the Supervisor architectures (Sup. Arch.) is determined in the light of the ODD and while some capabilities of the ADS might impact the choice of ODD the Sup. Arch. itself does not directly support the ODD, its formulation or selection.

The ODD and the HARA are both without direct connections to the Threat assessment (TA), Dynamic Risk Assessment (DRA), Dynamic Safety Management (DSM) and Precautionary Safety (PCS) methods. Instead, any connection from these runtime methods to the ODD and the HARA is indirectly provided via the collection of operational data (OpsData). There is one exception, though, namely the ability of DSM to provide updates to the risk assessment of the HARA during runtime by explicit quantification of the HARA factors. It might be argued that also PCS holds this connection, however, this would instead be via the QRN. And such updates would rather impact the decision-making of the ADS rather than the QRN (i.e. the HARA) itself. Also Out-of-Distribution (OoD) detection is only potentially connected indirectly to the HARA and CBD via OpsData.

For CBD, the data from Field Operational Tests (FOTs) and OpsData only provide general feedback to the development and design phase, the data does not directly influence the use of CBD. Thus, the runtime methods, as argued for the ODD and HARA above, also do not have any connection or provide any support to the use of CBD.

When devising Sup. Arch. the use of EVT is of no direct benefit as the architectural decisions and supervisory setups are not determined through statistical analysis of edge cases. However, the capabilities of the architecture might be analysed through EVT modelling of OpsData or similar, but not warranting a direct connection.

Similar to how ODDs might be indirectly influenced by CBD and Sup. Arch., the FOTs and OpsData collection are also not directly impacted by the HARA, CBD or Sup. Arch. That said, these methods might potentially guide the areas and conditions for which the FOTs should be conducted to capture relevant performance data for these methods. However, this would at most constitute an indirect connection for the benefit

M. Gyllenhammar and G. Rodrigues de Campos are with Zenseact

M. Gyllenhammar and M. Törngren are with the Mechatronics division at KTH, Royal Institute of Technology

The research has been partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation, and partially supported by the Swedish Innovation Agency (Vinnova, through the TADDO2 FFI project and the TECoSA centre for Trustworthy edge computing systems and applications).

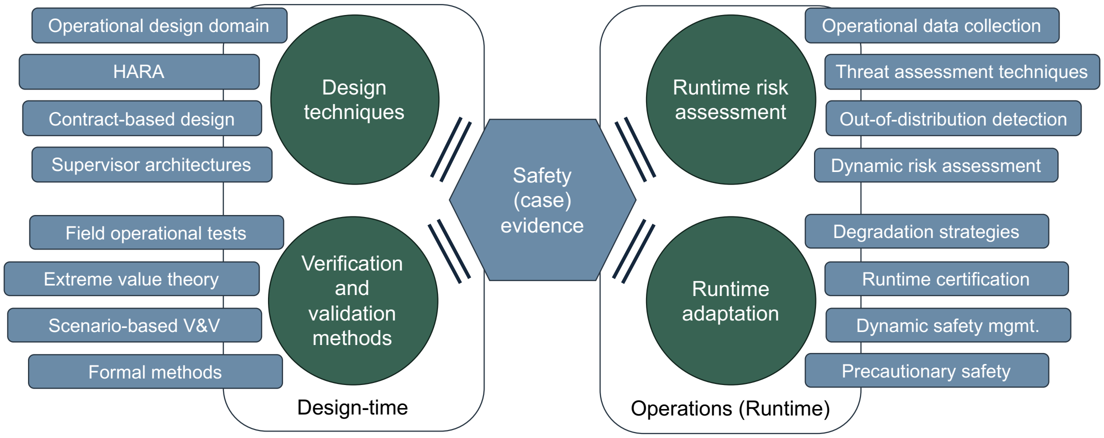

Fig. 1. Presents an overview of the 16 methods from [1] which are considered in the analysis and mapping presented here and in the main paper.

of the present analysis. When it comes to OpsData, also the ODD does not directly impact the collection. Of course, it influences where the ADS operates in the first place, but it might also be cases where OpsData is captured from outside the ODD, if the ADS fails to stop operations before exiting. The connection between the ODD to the OpsData would at any rate rather be a consequence of the entire system design and development rather than as an inherent consequence of the ODD itself. Thus motivating the blank cell. Also formal methods (FMs) lack direct influence on both FOTs as well sal OpsData.

The use of EVT as a statistical model will not benefit and will not be influenced by particular outcomes from the ODD, the HARA, CBD or Sup. Arch. Furthermore, the use of FMs will also not make any impact.

The way scenarios are captured and modelled for Scenario-based Verification and Validation (V&V) (Scenarios), are not impacted by Sup. Arch., neither is this influenced by the different risk assessment and runtime adaptation methods, except through an indirect connection via FOTs or OpsData. The use of FMs are also not influenced by any of the runtime methods, here also including OpsData. Moreover, the way FMs are deviced and used throughout the safety life cycle are also not impacted by Sup. Arch., FOTs, EVT or the use of Scenarios.

In general, the four runtime adaptation methods do not provide any direct links back to any of the V&V methods nor to the runtime risk assessment methods. The adaptation methods draw upon these for their definitions and to be able to operate. However, there is not feedback signal the other way. Any possible such signal would then travel via OpsData.

Considering the use of EVT it will not impact the collection of OpsData, even though it might be used to analyse the statistics resulting from the collection. Furthermore, this is true alsoe for the other risk assessment methods as well as degradation strategie, RT Cert. and DSM. EVT can work on statistics but will not impact the design nor the runtime aspects of these methods. EVT could, however, provide monitoring capabilities on the fleet level statistics, as suggested in [2].

Considering the data-driven nature of OoD detection methods, there is no connection between such methods and the HARA, CBD, Sup. Arch., Scenarios, FMs, TA or DRA.

In some ways, one could consider OoD as a potential threat metric, which would then go within the TA category. However, the outputs of the OoD detection is difficult to correlate to a physical threat, as the TA methods generally are. Consequently, there is no motivation for a direct connection between OoD and TA.

Considering CBD and Sup. Arch., these methods do not impact any of the runtime methods. Neither do they affect the ADS's ability to perform risk assessment or runtime adaptation during operations nor is the design of these methods related to the use of CBD or Sup. Arch. Degradation strategies are not dependent on Sup. Arch. in this context, the connection is in the opposite direction, where supervision and fall back channels are entirely dependent on the ADS's ability to operate in some degraded mode.

DRA is not connected to FMs, since the risk assessments made are more data-driven than formally specified. Some of the scenarios underpinning the DRA approach might of course be formally defined, but that connection is in such case captured via scenarios – not directly to FMs. Also PCS is data-driven rather than relying on formal approaches directly.

When considering the definitions of RT Cert. these do not directly depend on different scenarios, thus motivating the blank cell from scenarios to RT Cert.

RT Cert. is furthermore not influenced by OpsData. Neither in the design of the certificates nor for the execution of them. There might be an indirect connection where fleet level statistics from OpsData could impact some of the runtime evidence that could go into the evaluation of the certificates. However, such a connection is arguably rather via the other runtime assessment methods rather than OpsData per se. This argument is true also for degradations.

Lastly, RT Cert. and DSM constitute different methods to solve similar problems as PCS does. The problems are solved in different ways and there is no direct connection between these methods. At least not as they are developed at the time of writing of this analysis.

References

[1] M. Gyllenhammar, G. Rodrigues de Campos, and M. Törngren, "The road to safe automated driving systems: A review of methods providing safety evidence," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 4, pp. 4315–4345, 2025.

[2] D. Åsljung, C. Zandén, and J. Fredriksson, "A Risk Reducing Fleet Monitor for Automated Vehicles Based on Extreme Value Theory," *techrxiv*, 5 2022.

Magnus Gyllenhammar received his Ph.D. on safety strategies for ADSs in 2025 from KTH Royal Institute of Technology, Stockholm, Sweden. His research focuses on finding efficient strategies for safety argumentation of ADSs, especially focusing on precautionary safety and situation awareness in relation to the fulfilment of quantitative risk acceptance criteria. He received his MSc. in Engineering Physics, from Chalmers University of Technology, in 2016. In 2018, he joined Zenseact (then Zenuity) and has since worked on creating and realising data-driven strategies for verification and safety argumentation of ADSs.

Gabriel Rodrigues de Campos received his Ph.D. in Automatic Control in 2012 from Grenoble University/Grenoble INP, France. He is currently a researcher with Zenseact in Gothenburg, Sweden. Prior to joining Zenseact, he was a postdoctoral fellow with the Department of Signals and Systems, Chalmers University of Technology, Sweden and the DEIB, Politecnico di Milano, Italy. His research interests include cooperative and distributed control, safety assurance, and threat-assessment and decision-making techniques.

Martin Törngren has an engineering background in Mechatronics. After starting a company in the mid 90s, specializing in advanced tools for developers of embedded control systems, he embarked on an academic career, becoming a Professor in Embedded Control Systems at KTH in 2002. His core research interests are in cyber-physical systems design methodology including architecting, safety, and model-based engineering. Networking, multidisciplinary research and industrial collaboration have been characteristic throughout his career. He is the initiator of the Innovative Centre for Embedded Systems (www.ices.kth.se), launched in 2008, and the initiator and director of the TECoSA research center on Trustworthy Edge Computing Systems and Applications at KTH (www.tecosa.center.kth.se).