Systematic Identification and Analysis of Hazards for Automated Systems

Lina Putze, lina.putze@dlr.de; and **Eckard Böde,** eckard.boede@dlr.de

■ **ABSTRACT**
The introduction of automation into technical systems promises many benefits, including performance increase, improved resource economy, and fewer harmful accidents. In particular, in the automotive sector, automated driving is seen as one key element in Vision Zero by eliminating common accident causes such as driving under the influence, reckless behavior, or distracted drivers. However, this is contrasted by new failure modes and hazards from the latest technologies. In this article, we address the problems of finding common sources of criticality for specific application classes and identifying and quantitatively assessing new sources of harm within particular automated driving systems.

■ **KEYWORDS:** automated driving; hazard analysis; risk assessment; criticality; SOTIF; scenario identification; open context

INTRODUCTION – THE PROBLEMS OF IDENTIFYING RISKS FOR AUTOMATED DRIVING

Accidents due to speeding, distraction, or driving under the influence of alcohol – human misbehavior, intended or unintended, is an important factor in accident statistics. Self-driving vehicles are supposed to increase road safety by reducing the "human" risk factor. Although hazards associated with humans, like a collision due to a distracted driver, might be mitigated, the new technologies come with unknown risks and failure modes. The research topic, *Automation Risks*, focuses on identifying and assessing hazards and scenarios likely to trigger critical situations in the interaction of automated driving systems with their environment. In this article, we will focus on investigating automated driving systems since the methods presented have been developed in close collaboration with partners from the automotive industry. Nonetheless, we are actively adapting to other domains, like the maritime industry.

The safety of road vehicles is a well-known issue in the automotive industry. Due to the rising complexity of interacting safety-critical components, even conventional driving systems need to undergo a systematic safety process corresponding to ISO26262:2018 (ISO2018]). To keep development costs and efforts to a minimum, it is essential to include safety considerations from the beginning of the concept phase and throughout the entire development process because integrating changes in the system during early design phases is significantly easier. Knowledge about the common sources of criticality, for example, from accident databases, is an essential prerequisite for these first safety considerations. Moreover, a comprehensive safety concept requires a systematic identification and analysis of system-specific sources of harm. In the automotive domain, several methods exist for a so-called hazard and risk analysis (HARA), which is well-established in developing road vehicles.

Common hazard and risk analysis methods emphasize functional safety, which focuses on identifying and mitigating possible hazards caused by malfunctioning behavior of safety-related electrical and electronic systems. Assistance systems currently on the market, like adaptive cruise control, lane-keeping assistance, and combinations thereof, still require a human driver to monitor the vehicle and the environment and intervene when necessary. Nonetheless, many of those systems already take over parts of the driving tasks by providing braking, acceleration, and steering support while relying on sensor data that captures the internal and external environment. This comes with new potential sources of harm that take root in the system's specification. Let us consider an automatic emergency braking function (AEB). Despite the absence of faults and malfunctions, such hazards might occur due to incorrect interpretation of sensor input. For example, a poster on the roadside with a picture of a pedestrian crossing the road could be perceived as a natural person resulting in a breaking maneuver that could trigger a collision. This demonstrates that additional examination beyond the functional safety of the system is needed. We need to ensure that the system is robust concerning incorrect or unexpected sensor input, can comprehend situations correctly, and plans and acts responsibly based on these perceptions. These issues concerning the safety of the intended functionality (SOTIF) are addressed by ISO 21448:2022 (ISO 2022).

As assistance systems still have the driver as a redundant and immediately available fallback, such systems only

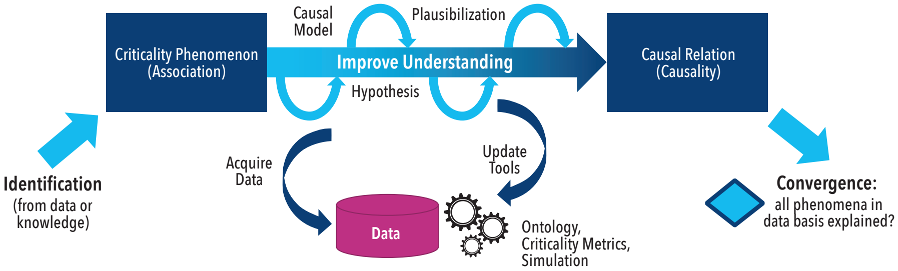

Figure 1. Basic concept of the criticality analysis

require evidence that the safety concept is fail-safe because the system does not provoke any additional risks, for example, by unintended interventions. In contrast to well-established systems, conditionally or highly automated driving functions like a traffic jam chauffeur temporarily release the driver from monitoring the environment for a certain time. This important step in the Levels of Driving Automation comes with additional safety difficulties as the abandonment of the driver as supervising instance involves the loss of a comprehensive and immediately available fallback. Therefore, it is necessary to prove that the system takes all the actions required to mitigate critical situations and that these actions are always carried out correctly and with the right timing: an operational-safe concept is required.

This is particularly problematic since automated driving systems driving on public roads face the challenge of operating safely in an open context. This arbitrarily complex, infinitely dimensional environment includes myriad factors that might lead to harm. Thus, it is infeasible to describe all relevant scenarios explicitly and specify the intended behavior. Moreover, hazards cannot be sufficiently reconstructed from existing real-world data. While there is extensive data for conventional driving systems, the challenges for automated driving systems differ from those for the human driver. For example, falling leaves in autumn are not generally a problem for the human eye, but if they hit the lens of a camera, object detection is not feasible anymore. Therefore, we cannot solely rely on data considering conventional systems and need extensive data that reflects the impact of automated systems on criticality.

To address these outlined issues, our research into *Automation Risks* is based on two main pillars: First, there is the criticality analysis which aims at finding common factors associated with criticality. Its focus is not on a specific system but on abstract application classes, such as the function of a highway chauffeur. Hence, the scope is in a pre-development phase where working groups comprising representatives from regulation authorities, standardization bodies, and industry define standard guidelines that every manufacturer of such a system must meet. In this setting, the criticality analysis will be a systematic approach to identify potential sources of criticality and specify a complete, well-defined set of criticality phenomena to be used as the basis for a homologation concept. Second, we work on a methodology that can be employed to perform a comprehensive hazard and risk analysis for specific highly automated systems that accompany the development process. This automation risks method aims to identify specific scenarios for further verification and validation and define safety goals as a basis for a fail-operational safety concept. The method intends to integrate functional safety (ISO 26262:2018 (ISO 2018)) and SOTIF (ISO 21448:2022 (ISO2022)) concerns.

STRUCTURING THE OPEN CONTEXT – CRITICALITY ANALYSIS

The first method we present is the criticality analysis. Its purpose is to investigate and structure the open context that constitutes the environment of automated vehicles. This includes not only the problem of identifying factors, parameters, and scenarios that have an essential impact on criticality but also abstracting these artifacts and mapping them on a finite set of criticality phenomena. This abstraction structures the criticality-inducing factors into comprehensive but manageable lists that can serve as a foundation for systematic verification and validation processes that enable a homologation for classes of automated systems. Furthermore, it helps to understand the underlying causalities to derive generic safety principles and mechanisms that avoid or mitigate the effects of critical situations.

Therefore, criticality analysis relies on a combined approach of expert-based and data-driven methods that precedes the design phase of specific systems. For example, it can be applied to urban traffic to set up a foundation for developing automated systems in this domain. In addition, it can support the operation and subsequent updates of corresponding systems in a DevOps process by continuously assessing changes in their domain. That might involve specific effects of amendments or enactments of laws and guidelines – a recent example would be the approval of e-scooters for German streets in 2019 – or even effects of climatic or societal changes. One of the fundamental principles of criticality analysis is that it does not only focus on the view of a single vehicle but also looks at the criticality of traffic. In this way, criticality analysis makes it possible to create generally accepted catalogs of criticality phenomena managed by regulation bodies and used by all manufacturers.

The basic approach of the criticality analysis is shown in Figure 1 and consists of three steps which we will present individually in the following.

1. *Identification and selection of criticality-triggering elements*: In the first step of the criticality analysis, candidates for criticality phenomena are selected for which a high correlation with a criticality increase is assumed. Expert knowledge, which is stored, for example, in the form of domain ontologies, test catalogs for vehicle approval, or accident databases, serves as a basis for the selection. Another source is data-driven approaches that systematically evaluate data from driving tests on test fields or in real traffic and data from specific computer simulations.
2. *Plausibilization and elaboration of interactions between criticality phenomena*: In the next step, the individual selected candidates for criticality phenomena are further analyzed. To make their influence on criticality, measurable criticality metrics are employed that quantify specific aspects of criticality. A typical example of such a metric is the time to collision (TTC), indicating the minimal time until a collision occurs, provided no action is taken. To achieve a comprehensive causal understanding of how the different phenomena affect certain aspects of criticality, we model the underlying causal assumptions based on causal theory by Judea Pearl (Pearl 2009). This theory allows the qualitative and quantitative investigation of causal queries based on constructing a so-called causal graph that represents the causal relationships of the different factors on a certain abstraction level. *Figure 2* illustrates such a causal graph for the criticality phenomenon stationary occlusion of traffic participants.
3. *Consolidation and abstraction of criticality phenomena/convergence*: The last step of the criticality analysis maps the identified and relevant criticality phenomena to a manageable and finite set of classes of criticality phenomena.

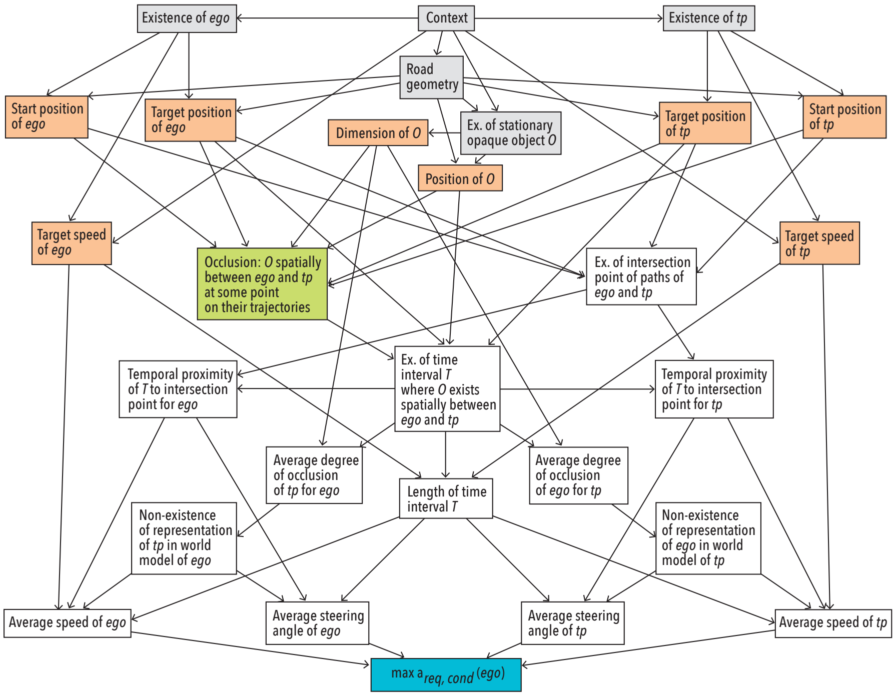
Existence of *ego*
Context
Existence of *tp*
Road geometry
Start position of *ego*
Target position of *ego*
Dimension of *O*
Ex. of stationary opaque object *O*
Target position of *tp*
Start position of *tp*
Position of *O*
Target speed of *ego*
Occlusion: *O* spatially between *ego* and *tp* at some point on their trajectories
Ex. of intersection point of paths of *ego* and *tp*
Target speed of *tp*
Temporal proximity of *T* to intersection point for *ego*
Ex. of time interval *T* where *O* exists spatially between *ego* and *tp*
Temporal proximity of *T* to intersection point for *tp*
Average degree of occlusion of *tp* for *ego*
Length of time interval *T*
Average degree of occlusion of *ego* for *tp*
Non-existence of representation of *tp* in world model of *ego*
Non-existence of representation of *ego* in world model of *tp*
Average speed of *ego*
Average steering angle of *ego*
Average steering angle of *tp*
Average speed of *tp*
max $a_{req, cond}(ego)$

Figure 2. Causal graph for the criticality phenomenon "occlusion of a participant (tp)

This assumes that such a manageable set must exist and that the number of criticality phenomena cannot be unlimited. If this were the case, the amount of data relevant to safe driving would surely exceed the processing capacity of human drivers. However, since we know that humans can drive a vehicle safely even in very complex situations, we can assume that there is a compact representation of the criticality phenomena. The procedure for generating the criticality classes takes each new criticality phenomenon from step 2 and compares it to the already identified classes of criticality phenomena. If these are similar, they are merged into a standard class. Otherwise, a new class is created. The process is continued until it is determined with sufficient statistical certainty that all new phenomena found in step 1 are only ever mapped to already known classes.

During the execution of the method, individual parts, particularly in Step 2, are iterated repeatedly. This is done until the underlying mechanisms are sufficiently understood. A manageable finite set of abstracted criticality phenomena remains, covering all criticality-triggering causes for the investigated system class in a given environment. However, let us note that the method can be presented here only in a highly simplified form, and the figure notably omits details on where and how the feedback loops tie in with the process. For a comprehensive description of the methodology, please refer to Neurohr et al. 2021.

HAZARD AND RISK ANALYSIS FOR AUTOMATED SYSTEMS

The second method we elaborate on is the automation risks method (Kramer et al. 2020) which defines a comprehensive approach to the hazard and risk analysis of automated driving functions. It addresses both functional safety and SOTIF (safety of the intended functionality) by sustaining existing safety processes of the standards ISO 26262:2018 and ISO 21448:2022 and complementing them where necessary (ISO2018, ISO2022). The focus is on hazards that are inherent in the system but are triggered by external influences of the automated function, such as situations where the automated driving function does not react appropriately to its current environment. This includes non-detection or misclassification of objects, such as a bicyclist not detected or misclassified as a pedestrian, erroneous recognition of non-existing objects, and wrong predictions of future events, for example, due to wrong dynamic models. Therefore, the method builds on established analytical techniques for hazard analysis and risk assessment while it adds significant enhancements to enable the applicability to automated systems.

The proposed method is designed to accompany the entire development process. It is beneficial to initiate its application early during the concept phase so that safety considerations can be integrated into the system as early as possible. As shown in Figure 3, the method contains several feedback loops between the concept phase and development that enable the consideration of adjustments in the system, especially the integration and analysis of risk mitigation measures based on the previously gained knowledge, such as the implementation of redundancies or the definition of a higher safety distance.

The approach involves two main parts: the identification of hazardous scenarios (Steps (1) – (5) in Figure 3. Overview of the automation risks method) and the quantification of corresponding risks (steps (6) and (7)).

The first part aims to identify hazards, understand the underlying causal relationships, and deduce scenarios that might

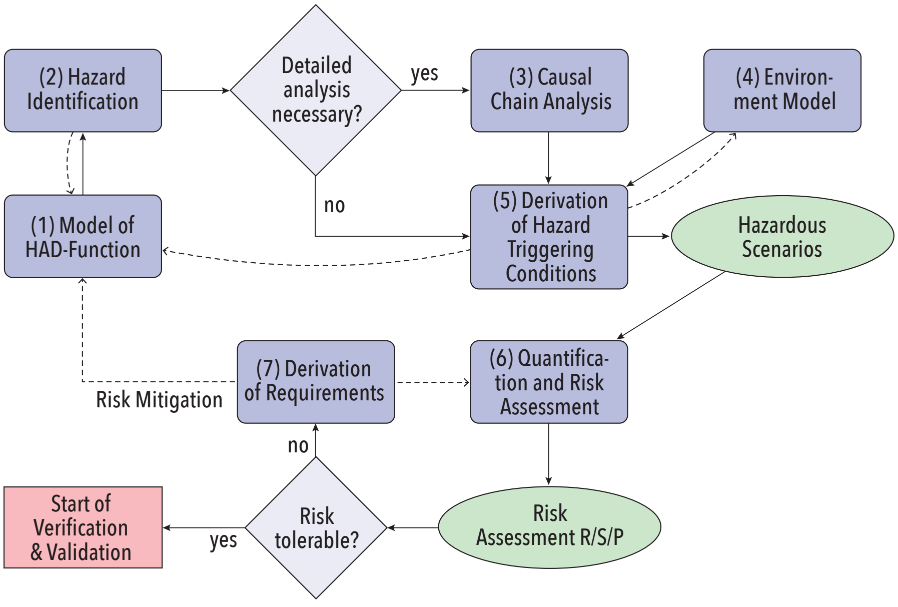
(2) Hazard Identification
Detailed analysis necessary?
yes
(3) Causal Chain Analysis
(4) Environment Model
no
(1) Model of HAD-Function
(5) Derivation of Hazard Triggering Conditions
Hazardous Scenarios
Risk Mitigation
(7) Derivation of Requirements
(6) Quantification and Risk Assessment
no
Start of Verification & Validation
yes
Risk tolerable?
Risk Assessment R/S/P

Figure 3. Overview of the automation risks method (Kramer et al. 2019)

ID	Basic Scenario	Basic Maneuver	Correct if (context)	Keyword	Incorrect Vehicle Behavior	Observable Effect(s) in Scenario	Additional Scenario Conditions (necessary for Top Level Event)	Potential Top Level Event
1	slower turn into path challenger	decelerate/ braking	front distance < safety distance	no	necessary breaking maneuver not initialed	ego continues with constant speed	challenger with significantly lower speed or critical Time-To-Collision	front/side collision with challenger
2				less	breaking maneuver not strong enough	Ego does not decelerate to prevent collision	challenger with significantly lower speed or critical Time-To-Collision	front/side collision with challenger

Figure 4. *Table for identification of hazards on vehicle level (Kramer et al. 2019)*

Functional Unit	Function			Key-word	Local Failure/ Functional Insufficiency	Basic Scenario	System Effect(s) in Scenario	Incorrect Vehicle Behavior	ID(s) of IVB	Possible System Cause(s)	Environ-mental Condition	Relevant for human driver?
	Input	Compu-tation	Output									
Sensors > Front camera > object recognition	camera image	segmen-tation	seg-mented camera image	no	segmented camera image not generated	slower turn into path challenger	challenger not detected by front camera > maneuver planning without information about the challenger	necessary braking maneuver not initiated	1	HW-failure, degradation or design fault	none	no statement
					no segments in camera image recognized	s/a	s/a	s/a	s/a	no night vision lacking sensibility at dark	darkness	likely (human vision also impaired by darkness)

Figure 5. *Table for identification of hazards on component level (Kramer et al. 2019)*

trigger hazardous events. These hazardous scenarios serve as inputs to the following quantification part. They can also serve as a basis for comprehensive scenario-based testing within the verification and validation process and define a starting point for improvements in the system.

The investigation is based on an initial system description that involves at least an item definition and a functional architecture that describes an architectural model representing system functions, like sensor fusion or trajectory planning and their interactions. To identify hazards caused by incorrect behavior of the automated function, we employ a keyword-based brainstorming approach inspired by the hazard and operability study (HAZOP) (Ericson 2005, 365-381), a technique originated from the chemical industry. The main idea is to combine a set of basic scenarios with a set of basic maneuvers that the automated function could perform with a list of keywords to derive possible incorrect behavior of the automated system that might lead to harm. An example of such a table applied to a highway-chauffeur function is provided in Figure 4.

In the next step, we employ a second HAZOP-inspired approach to examine local failures and functional insufficiencies and their effects on the system and vehicle level by applying keywords to the individual functional units.

Based on the identified hazards, we aim to derive scenario properties that might provoke them. Therefore, we use a modified fault tree analysis (Ericson 2005, 183-222) which analyzes the causal chains starting from the top-level event of a hazard during a basic scenario.

A unique feature is that we denote environmental conditions in the tree wherever necessary for the propagation of a fault. We can derive the triggering scenario properties by reducing the fault tree to these environmental conditions and identifying so-called minimal cut sets. An exemplary dependency graph is shown in Figure 6.

The quantification aims to derive a risk assessment that can be used to determine safety goals based on the afore-identified scenario properties. Therefore, it mainly

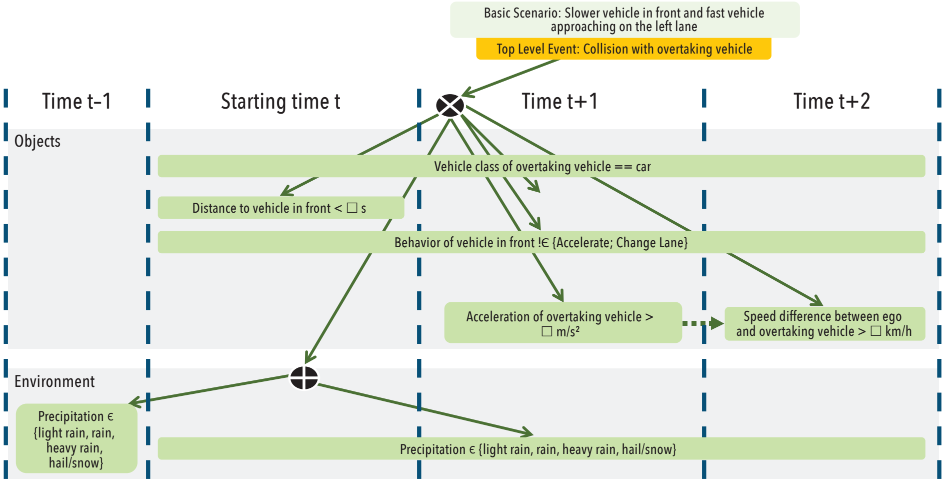
Basic Scenario: Slower vehicle in front and fast vehicle approaching on the left lane
Top Level Event: Collision with overtaking vehicle
Time t-1
Starting time t
Time t+1
Time t+2
Objects
Vehicle class of overtaking vehicle == car
Distance to vehicle in front < □ s
Behavior of vehicle in front !∈ {Accelerate; Change Lane}
Acceleration of overtaking vehicle > □ m/s²
Speed difference between ego and overtaking vehicle > □ km/h
Environment
Precipitation ∈ {light rain, rain, heavy rain, hail/snow}
Precipitation ∈ {light rain, rain, heavy rain, hail/snow}

Figure 6. *Exemplary part of an environmental fault tree reduced to the environmental conditions and chronologically ordered into discrete time steps*

builds on probability estimation. Relying on the probabilities of occurrence of the single environmental conditions and the conditional probabilities that an error propagates in the fault tree, we estimate the probability of a hazard occurring with the help of the single minimal cut sets representing the triggering scenario properties. This serves as a basis for the risk assessment according to the automotive safety integrity level (ASIL) of the ISO 26262:2018 (ISO 2018).

SUMMARY AND OUTLOOK

In this paper, we presented two methods that enable systematic investigation of criticality causes and their effects in the context of automated systems.

Criticality analysis aims at identifying a comprehensive list of all potential sources of criticality in a given application field which serves as input for certification authorities and test organizations to develop detailed homologation guidelines. The method is being developed in the VVMethoden project in close cooperation with representatives from the automotive industry.

The second approach describes an extension of a hazard and risk analysis in which functional safety is combined with SOTIF (safety of the intended functionality). This approach was developed in the PEGASUS project, where it was extensively tested using the example of a highway chauffeur function. A comprehensive description of the approach and the evaluation can be found in (Böde et al. 2019). Furthermore, we have investigated to what extent the approach can be adopted in other application domains. Vander Maelen describes the application of this method to a collision warning system in the maritime domain (Vander Maelen et al. 2019).

Currently, we are working on elaborating the methods, simplifying their application, and investigating other use cases. In two internal projects, we are investigating the suitability of these approaches for hazard detection in automated road traffic (https://verkehrsforschung.dlr.de/de/projekte/kokovi) and for automated ship navigation in port areas (https://verkehrsforschung.dlr.de/de/projekte/das-projekt-futureports-fuer-hochautomatisierte-digitalisierte-und-intermodal-vernetzte). ■

ACKNOWLEDGMENTS

This paper reports on collaborative work by many colleagues from DLR and partners in the projects PEGASUS (https://www.pegasusprojekt.de/en/home) and VVMethoden (https://www.vvm-projekt.de/en/). These projects received funding from the German Federal Ministry for Economic Affairs and Climate Action.

In particular, we want to thank our DLR (formerly OFFIS) colleagues Matthias Büker, Birte Neurohr, Christian Neurohr, Sebastian Vander Maelen, and Lukas Westhofen, the scientific leaders Werner Damm and Martin Fränzle, as well as our industrial cooperation partners Martin Butz (Bosch), Martin Bollmann (ZF), Ulrich Eberle (Stellantis) and Roland Galbas (Bosch) for their substantial contributions.

REFERENCES

- Böde, E., M. Büker, W. Damm, M. Fränzle, B. Kramer, C. Neurohr, and S. Vander Maelen. 2019. "Identifikation und Quantifizierung von Automationsrisiken für hochautomatisierte Fahrfunktionen." Technical report, OFFIS e.V.
- Ericson, C. A. 2005. *Hazard Analysis Techniques for System Safety*. John Wiley & Sons, Inc.
- Kramer, B., C. Neurohr, M. Büker, E. Böde, M. Fränzle, and W. Damm. 2020. "Identification and quantification of hazardous scenarios for automated driving." *International Symposium on Model-Based Safety and Assessment*: 163–178.
- Neurohr, C., L. Westhofen, M. Butz, M. H. Bollmann, U. Eberle, and R. Galbas. 2021. "Criticality analysis for the verification and validation of automated vehicles." *IEEE Access 9*: 18016-18041. doi:10.1109/ACCESS. 2021.3053159.
- Pearl, J. 2009. *Causality* (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511803161
- Vander Maelen, S., M. Büker, B. Kramer, E. Böde, S. Gerwinn, G. Hake, and A. Hahn. 2019. "An Approach for Safety Assessment of Highly Automated Systems Applied to a Maritime Traffic Alert and Collision Avoidance System." *2019 4th International Conference on System Reliability and Safety (ICSRS)*: 494-503, doi:10.1109/ICSRS48664.2019.8987712.
- ISO (International Organization for Standardization). 2022. ISO 21448:2022. Road vehicles — Safety of the intended functionality. Geneva, CH: ISO.
- ———. 2018. ISO 26262:2018. Road vehicles – Functional safety. Geneva, CH: ISO.
- SAE (Society of Automotive Engineers). 2021. SAE J3016:2021. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. Geneva, CH: SAE/ISO.

ABBREVIATIONS OF STANDARDS, CORRELATED WITH REFERENCE-LIST CITATIONS

ISO 21448:2022	(ISO 2022)
ISO 26262:2018	(ISO 2018)
SAE J3016:2021	(SAE 2021)

ABOUT THE AUTHORS

Lina Putze is a researcher at the DLR Institute on Systems Engineering for Future Mobility. Her background is in mathematics.

Eckard Böde is a manager of the research group System Concepts and Design Methods at the DLR Institute on Systems Engineering for Future Mobility. His background is in computer science and model-based safety assessment of cyber-physical systems.