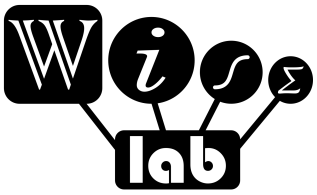

WISE Requirements Analysis Framework for Automated Driving Systems

On-Road Safety of Automated Driving System (ADS)

Taxonomy and Safety Analysis Methods

Krzysztof Czarnecki
Waterloo Intelligent Systems Engineering (WISE) Lab
University of Waterloo
Canada

July 21, 2018

For updates and related documents, see
https://uwaterloo.ca/waterloo-intelligent-systems-engineering-lab/projects/wise-requirements-analysis-framework-automated-driving

Document history

Date	*Updates*	*Contributor(s)*
August 14, 2017	Version 1.0 created	Krzysztof Czarnecki[1]
November 30, 2017	Comments by Rick Salay[1] addressed	Krzysztof Czarnecki[1]
July 21, 2018	Version 1.1 Section on methods revised to include driving behavior safety	Krzysztof Czarnecki[1]

[1]Waterloo Intelligent Systems Engineering Lab
University of Waterloo

Abstract

This document provides a taxonomy of on-road safety of an Automated Driving System (ADS) and describes safety analysis methods applicable at requirements and early design stage of ADS development. In particular, it covers Driving Behavior Safety Assurance, Safety of The Intended Functionality (SOTIF) Assurance, and the Hazard Analysis and Risk Assessment (HARA) part of Functional Safety, and relates the three methods to System-Theoretic Process Analysis (STPA).

Table of Contents

1. Scope
2. Related Documents
2.1 Industry Guidance
2.2 STPA Literature

3. Basic Terms
3.1 Safety, Harm, Risk, and Hazard
3.2 Crash, Unstabilized Situation, and Harmful Event
3.3 Transport Crash Classification

4. On-Road Safety of ADS

5. Safety Analysis Methods
5.1 System-Theoretic Process Analysis (STPA)
5.2 Causes of Hazardous Control Actions in Automated Driving
5.3 Driving Behavior Safety and Its Assurance
5.3.1 Identification of Crash Types and Safety Requirements on Driving Behavior
5.3.2 Verification and Validation of Specified Driving Behavior and Residual Risk Assessment
5.4 Safety of The Intended Functionality (SOTIF)
5.5 Functional Safety (FuSa)

References

1. Scope

This document provides a taxonomy of on-road safety of an Automated Driving System (ADS) and describes safety analysis methods applicable at requirements and early design stage of ADS development, including automated driving task specification and functional concept design.

The document was developed with the intent to be consistent with prior art to the extent possible, including relevant industry standards and scientific literature. The majority of the terms are borrowed from applicable industry standards; new terms are defined as necessary (see Appendix A for a summary of term sources). The following format is used when referring to terms defined in the industry guidance documents: [XXX](N.M….), where [XXX] references a document listed in Section 2, and (N.M….) is the definition number specified in the referenced document. For example, [FuS](1.103) refers to definition 1.103 in ISO 26262:2011. When quoting a definition, the references inside quotation marks only give the definition reference since the quotation already provides the document reference.

The taxonomy and safety analysis methods target the on-road safety of a passenger car, equipped with level 4 or 5 *driving automation* according to SAE J3016. In particular in its current scope, the document does not deal with ADS requesting a fallback-ready user to take over control and perform the dynamic driving task fallback. A passenger car is a motor vehicle "designed and constructed primarily for the carriage of persons and their luggage, their goods, or both, having not more than a seating capacity of eight, in addition to the driver, and without space for standing passengers." [FUS](1.86). "On-road" refers to publicly accessible roadways, including parking areas and private campuses that permit public access, and private access roads and garages, "that collectively serve users of vehicles of all classes and driving automation levels (including no driving automation), as well as motorcyclists, pedalcyclists, and pedestrians." [LA] The scope is further restricted the subject vehicle being *in-transport* [TAC](2.2.34). On-road safety is scoped to mean the absence of unreasonable risk of *transport crashes* [TAC] (2.4.9), with the ADS of the subject vehicle being a causal or contributing crash factor.

An *Automated Driving System* (ADS) is an E/E system that performs level 3, 4, or 5 driving automation [LA](3.2) In particular, ADS performs the complete Dynamic Driving Task (DDT) [LA](3.13), which consists of operational and tactical aspects of driving, and potentially also performs parts of the strategic driving tasks (the latter is beyond the scope of SAE 3016 [LA]). A passenger car equipped with ADS that is subject to the safety analyses is referred to as *subject vehicle*.

The taxonomy covers key terms related to different types of on-road safety of ADS, in particular Driving Behavior Safety, Safety of The Intended Functionality (SOTIF), and Functional Safety (FuSa).

The document also describes the System-Theoretic Process Analysis (STPA), as a general safety assurance method, and three methods targeting specific hazard sources in the context of

automated driving: Driving Behavior Safety Assurance, SOTIF Assurance, and the Hazard Analysis and Risk Assessment (HARA) of FuSa. These methods are applicable at requirements and early design stage of ADS development, including automated driving task specification and functional concept design.

2. Related Documents

2.1 Industry Guidance

This document builds on the following industry guidance:

[LA] Surface Vehicle Recommended Practice — Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE J3016:Jul2018
This SAE recommended practice defines levels of driving automation and related concepts, including Automated Driving System (ADS).

[SOT] Road vehicles — Safety of the intended functionality. Working Document, ISO/WD-PAS 21448.1:2017-11-15
This working document represents work in progress to develop an ISO Publicly Available Specification (PAS) providing guidance on design, verification and validation measures to achieve safety of the intended functionality, that is, avoid unsafe behavior that stems from technological and system definition shortcomings. The scope of the upcoming first edition of the PAS targets driving automation at levels 0, 1, and 2. While this first edition can be taken into account when developing driving automation at levels 3, 4, and 5, additional measures may be necessary. These will be addressed in future editions.

[FuS] Road vehicles — Functional safety. ISO 26262:2011
This ISO standard defines terms and activities to be used in ensuring the functional safety of electrical and/or electronic (E/E) systems within motor vehicles. Functional safety is the absence of unreasonable risk from malfunctioning behavior caused by failures or unintended behavior with respect to design intent.

[TAC] Manual on Classification of Motor Vehicle Traffic Crashes. ANSI D16.1-2017
This ANSI standard provides a taxonomy and guidance on classifying motor vehicle traffic crashes. It defines different types of transport crashes, including collision and noncollision crashes, to be used in collecting and analyzing data for crash databases.

[CC] Model Minimum Uniform Crash Criteria (MMUCC) Guideline. Fourth edition, DOT HS 811 631, Governors Highway Safety Association (GHSA), July 2012

This guideline provides a uniform schema for collecting, storing, and analyzing motor vehicle crash data in the United States. The classification schema is based on ANSI D16.1-2007.

2.2 STPA Literature

The following scientific literature is used as a source of information on STPA:

[Lev03] N. Leveson. A New Approach to Hazard Analysis for Complex Systems. In Proceedings of International Conference of the System Safety Society, Ottawa, August 2003
This paper provides a concise summary of the basic ideas behind STAMP and STPA.

[Lev11] N. Leveson. Engineering a Safer World: Systems Thinking Applied to Safety. MIT Press, 2011
This book provides a comprehensive description of STAMP and STPA.

[ALW17] A. Abdulkhaleq, D. Lammering, S. Wagner, J. Röder, N. Balbierer, L. Ramsauer, T. Raste, H. Boehmert. A Systematic Approach Based on STPA for Developing a Dependable Architecture for Fully Automated Driving Vehicles. In 4th European STAMP Workshop 2016, Procedia Engineering 179 (2017) 41 – 51
This paper applies STPA to fully automated driving (levels 3, 4, 5). In particular, it provides a sample control structure for use in STPA.

2.3 Literature on Scene, Situation, and Scenario

The following scientific literature is used as a basis to define the terms scene, situation, and scenario:

[GBF14] S. Geyer, M. Baltzer, B. Franz, S. Hakuli, M. Kauer, M. Kienle, S. Meier, T. Weissgerber, K. Bengler, R. Bruder, F. Flemisch, H. Winner. Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance. IET Intelligent Transport Systems, vol. 8, no. 3, 2014, pp. 183-189

[UMR15] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, M. Maurer. Defining and substantiating the terms scene, situation, and scenario for automated driving. In Proc. 2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC), 2015, pp. 982-988

3. Basic Terms

3.1 Safety, Harm, Risk, and Hazard

This document adopts the terms *safety*, *harm*, and *risk* from ISO 26262.

ISO 26262 defines *safety* [FuS](1.103) as "absence of unreasonable risk" of harm. This definition is in contrast to defining safety as freedom from *harmful events* [Lev11]; the absence of a harmful event may simply mean that the event has not yet occurred. Thus, safety is more adequately defined as the absence of unreasonable risk of harm.

Harm is defined as "physical injury or damage to the health of persons." [FuS](1.56) Thus, ISO 26262 limits the definition of harm to personal injury; however, ANSI D.16.1 subdivides harm into (personal) *injury* and (property) *damage*.

ISO 26262 defines *risk* [FuS](1.99) as "combination of the probability of occurrence of harm (1.56) and the severity (1.120) of that harm." *Unreasonable risk* [FuS](1.136) is defined as "risk (1.99) judged to be unacceptable in a certain context according to valid societal moral concepts."

A *hazard* is potential source of harm caused by (1) deficiencies in the specified behavior, (2) performance limitations of the intended functionality (3) malfunctioning behavior, (4) foreseeable misuse, or (5) security vulnerability. This definition extends the one from ISO 26262 [FuS](1.57), which is focused on malfunctioning behavior [FuS](1.73) due to failures [FuS](1.39) or unintended behavior (case 1), with cases (2,4,5).

3.2 Crash, Unstabilized Situation, and Harmful Event

The document is scoped to safety related to *transport crashes*.

A *crash* is "an unstabilized situation which includes at least one harmful event." [TAC](2.4.6) An *unstabilized situation* is "a set of events not under human control." [TAC](2.4.4) A *harmful event* is "an occurrence of injury or damage." [TAC](2.4.1) Traffic safety literature suggest using the term "crash" rather than "transport accident" because the term "accident" emphasizes "randomness," obscuring the controllable causes such as human factors [Eva04]. In fact, editions of ANSI D16.1 prior to the eighth one have used the term "accident" in place of "crash".

A *transport crash* is "a crash that involves a transport vehicle in-transport" [TAC](2.4.9). A *transport vehicle* consists of "one or more devices or animals and their load […]." [TAC](2.1.4) Examples of transport vehicles are car, airplane, train, snowmobile, and horse and rider. When applied to motor vehicles, *in-transport* means "on a roadway or in motion within or outside [the public road.]" [TAC](2.2.34) This definition includes vehicles in traffic on public and private roadways, and (possibly soft) shoulders; it also includes disabled vehicle on a roadway. It does not include vehicles parked legally in designated parking spaces; however, "in roadway lanes

used for travel during some periods and for parking during other periods, a parked motor vehicle should be considered to be in-transport during periods when parking is forbidden." [TAC](2.2.34)

The definition of a transport crash excludes situations resulting directly from *cataclysms* [TAC](2.4.5), such as a tornado or an earthquake, but includes situations occurring after the cataclysm has ended, such as road obstruction by fallen trees, and situations resulting from natural events that are not cataclysms, such a tree branches falling on a motor vehicle in traffic. However, a branch falling on a motor vehicle parked legally is not a transport crash, because the vehicle is not in-transport. On the other hand, an illegally parked vehicle being struck by a falling branch or another vehicle is considered a transport crash. Thus, in the case an ADS parks a subject vehicle illegally and the parked vehicle is subsequently struck by another vehicle or a falling tree branch, the ADS would have contributed to the cause of the resulting transport crash.

3.3 Transport Crash Classification

Transport crashes include *motor vehicle crashes*, *railway crashes*, and *airplane crashes*. A *motor vehicle crash* [TAC](2.4.12) is a transport crash "that involves a motor-vehicle in-transport," but does not involve aircraft or railway train in-transport. In particular, a collision between a motor vehicle and a railway train in-transport is classified as a *railway accident* [TAC](2.4.13). A *road vehicle crash* [TAC](2.4.17) is "a transport crash that is either a motor vehicle crash or an *other-road-vehicle crash*." An example of an other-road-vehicle crash is a collision of a pedalcycle in-transport with a pedestrian on a public road. A *traffic crash* is a road vehicle crash in which "(1) the unstabilized situation originates on a [public road] or (2) a harmful event occurs on a [public road]" [TAC](2.4.18). Thus, a road vehicle crash that occurs on a private road, rather than a public road, is not a traffic crash.

A road vehicle crash is further classified by the type of its first *harmful event*: a *collision crash* [TAC](2.6.2) and a *noncollision crash* [TAC](2.6.3). A collision crash is a road vehicle crash in which the first harmful event is a collision of a road vehicle in-transport with another vehicle, pedestrians, or other objects. Noncollision crashes are road vehicle crashes other than collision crashes, including overturning, jackknife, fire or explosion of any parts of road vehicle in-transport, immersion (such as driving into water), occupant falling from a road vehicle in-transport or being thrown against some part of the vehicle inside, and thrown or falling objects striking occupant or the road vehicle in-transport.

The Model Minimum Uniform Crash Criteria (MMUCC) [CC] defines a crash description schema to be used by reporters, classifiers, analysts and users of traffic crash data in the United States. Table 1 gives the MMUCC classification of crashes according to the first harmful event. Other MMUCC crash criteria include location of first harmful event (on roadway, shoulder, median, roadside, parking zone, etc.), impact type (front-to-rear, front-to-front, angle (includes front-to-side), sideswipe in same or opposite direction, rear-to-side, rear-to-rear), weather conditions,

light conditions, roadway surface conditions, contributing circumstances, road configuration, and temporary road structure.

Table 1 MMUCC classification of traffic crashes according to the first harmful event [CC]

Noncollision	**Collision**	
• Overturn/Rollover • Fire/Explosion • Immersion, Full or Partial • Jackknife • Cargo/Equipment Loss or Shift • Fell/Jumped From Motor Vehicle • Thrown or Falling Object • Other Noncollision	**With Person, Motor Vehicle, or Non-Fixed Object**	**Collision With Fixed Object**
	• Pedestrian • Pedalcycle • Other Non-motorist • Railway Vehicle (train, engine) • Animal (live) • Motor Vehicle in-Transport • Parked Motor Vehicle • Struck by Falling, Shifting Cargo or Anything Set in Motion by Motor Vehicle • Work Zone / Maintenance Equipment • Other Non-Fixed Object	• Impact Attenuator / Crash Cushion • Bridge Overhead Structure • Bridge Pier or Support • Bridge Rail • Cable Barrier • Culvert • Curb • Ditch • Embankment • Guardrail Face • Guardrail End • Concrete Traffic Barrier • Other Traffic Barrier • Tree (standing) • Utility Pole/Light Support • Traffic Sign Support • Traffic Signal Support • Fence • Mailbox • Other Post, Pole or Support • Other Fixed Object (wall, building, tunnel, etc.)

4. On-Road Safety of ADS

On-road safety of ADS is defined as absence of unreasonable risk from transport crashes due to hazards caused by the ADS of the subject vehicle in-transport.

On-road safety of ADS is concerned with crashes in which the subject vehicle is involved directly (cf. *contact vehicle* [TAC](2.4.7)) or indirectly (cf. *noncontact vehicle* [TAC](2.4.8)) while the subject vehicle is in-transport. These crashes include not only collisions of the subject vehicle with other vehicles or pedestrians, but also collisions with railway trains at level crossings or aircraft landing on a roadway in emergency. They also include crashes of other road vehicles such as a car or a pedalcycle involving the subject vehicle indirectly, i.e., without contact between the subject vehicle and any vehicle directly involved in the crash. An example of such a crash would be an unsafe cut-in maneuver by the subject vehicle in intense traffic, which would cause subsequent vehicles to crash, without any contact with the subject vehicle. Another example would be a car colliding with roadside infrastructure because of its driver being blinded by high-beam headlights of the subject vehicle at night. In both examples the subject vehicle would be contributing to the cause of the crash as a noncontact vehicle.

5. Safety Analysis Methods

This section describes safety analysis methods that are applicable at requirements and early design stage of ADS development, including automated driving task specification and functional concept design.

The first described method, the System-Theoretic Process Analysis (STPA), is a general safety assurance method that provides a sound overall foundation for the three subsequently presented, more specific, and complementary methods: Driving Behavior Safety Analysis, SOTIF analysis, and the Hazard Analysis and Risk Assessment (HARA) of FuSa. SOTIF [SOT] and FuSa [FuS] are industry standards providing guidance on safety assurance related to hazards caused by, respectively, (i) limited performance of sensor technology and algorithms and foreseeable misuse and (ii) malfunctioning behavior. Driving Behavior Safety Analysis focuses on hazards caused by deficiencies in the specified driving behavior, and is defined in this document.

5.1 System-Theoretic Process Analysis (STPA)

System-Theoretic Process Analysis (STPA) is a hazard analysis technique based on the System Theoretic Accident Modeling Process (STAMP) [Lev11]. The key idea in STAMP is that crashes occur because of inadequate control that takes a system outside its safe envelope.

Consequently, STPA focuses first on identifying safety constraints, that is, safety requirements, rather than hazardous events, and determining control actions that may violate the safety constraints. STPA then examines the control structure of a system to determine the potential causes for the unsafe actions, and suggests improvements to eliminate, reduce, control, or mitigate these actions in design or operation.

The steps of STPA are as follows [Lev11,Lev03]:

1. Identify mishaps, hazards, and safety requirements (see Chapter 7 in [Lev11]).
2. Identify hazardous control actions, that is, actions that could lead to a hazardous state (violation of safety requirements):
    a. A control action required for safety is not provided or not followed.
    b. An unsafe control action is provided.
    c. A potentially safe control action is provided too early or too late, that is, at the wrong time or in the wrong sequence.
    d. A control action required for safety is stopped too soon or applied too long or at too high or too low intensity.
3. Determine how each potentially hazardous control action identified in step 1 could occur and eliminate, control, mitigate hazardous control actions in design or operation.
    a. Consider the following potential causes for hazardous control actions when analyzing the control structure: inadequate enforcement of constraints in the controller part, including process model, algorithms, and coordination among controllers; inadequate execution of control actions (e.g., communication flaws, actuator problems, and time lags); and inadequate or missing feedback (see Figure 4 in [Lev03]).
    b. Design controls and mitigation measures if they do not already exist or evaluate existing measures if the analysis is being performed on an existing design.
    c. Consider degradation of the control structure over time and build protection, including change management, audits, and accident an incident analysis.

Leveson defines a hazard as “a system state or set of conditions that, together with a particular set of worst-case environmental conditions, will lead to an accident (loss).” [Lev11] Thus, a violation of a safety constraint, that is, a safety requirement, is a hazard. This definition is consistent with ISO26262, which defines a hazard as “a potential source of harm.” The STPA process identifies crash types, then high-level hazards, and then safety requirements [Lev11]. Leveson states that “the definition of hazards associated with a system is an arbitrary but important step in assuring system safety.” She further suggests to start with a small number of high-level hazards, arguing that even large systems would have seldom more than a dozen high-level hazards. These high-level hazards are then used to derive more detailed safety requirements.

5.2 Causes of Hazardous Control Actions in Automated Driving

There are different types of potential causes for hazardous control actions (Step 3a in STPA), including deficient specifications, failures of hardware or software components, performance limitations of sensor and actuator technology and algorithms, security vulnerabilities, and impacts from external infrastructure, driving environment, and other technologies (e.g., mechanical or hydraulic). Table 2 lists existing industry guidance to address safety assurance related to these different types of causes.

Table 2 Causes of hazardous control actions and existing industry guidance

Cause for hazardous control actions	**Existing Guidance**
Deficiency in specified driving behavior (this cause is addressed by *driving behavior safety*, see Section 5.3)	ISO 26262 considers deficiencies in specified behavior as a type of "malfunctioning behavior"; however, the standard does not provide guidance on the required driving performance; ISO PWI 22737 is addressing the latter for Low Speed Automated Driving Systems (LSAD).
E/E system failures	ISO 26262 provides detailed guidance to address hazardous control actions due to failures of electric and electronic hardware and computer software
Performance limitations of the specified behavior due to limitations of sensor technology, algorithms (e.g., machine learning), and actuator technology	ISO/PAS 21448 primarily addresses performance limitations due to sensor technology and algorithms; ISO 26262 also addresses actuator technology
Foreseeable misuse (e.g., user confusion, user overload)	ISO/PAS 21448; European Statement of Principles on human-machine interface
Security vulnerabilities	SAE J3061, ISO 21434 (draft), PAS 11281 (draft)
Impacts from active infrastructure and/or vehicle to vehicle communication, external devices and cloud services	Extended Vehicle (ExVe) methodology defined in ISO 20077
Impact from car surroundings (other users, "passive" infrastructure, environmental conditions: weather, EMC...)	ISO/PAS 21448; ISO 26262
Impact from other technologies, e.g., mechanical or hydraulic technology	ISO 26262 defines the concept of "other technologies" and allows allocation of safety requirements to them; other existing standards cover these technologies (e.g., Federal Motor Vehicle Safety Standards (FMVSS) in the U.S.)

5.3 Driving Behavior Safety and Its Assurance

Driving behavior safety of an ADS is the absence of unreasonable risk from transport crashes due to hazards caused by deficiencies in the specified driving behavior of the ADS-operated subject vehicle in-transport.

The on-road driving vehicle behavior is the main factor under the ADS control that influences the risk of a crash. Driving behavior safety requires that this behavior must be such as not to expose the subject vehicle, occupants, and other road users and property to unreasonable risk of crashes.

Assuring driving behavior safety consists of the following steps:

1. Identification of crash types and safety requirements on driving behavior;
2. Verification that the specified driving behavior satisfies to the safety requirements on driving behavior;
3. Validation of the safety of the specified driving behavior in real-life use cases;
4. Assessment of residual risk due to specified driving behavior.

Step 1 of driving behavior safety assurance corresponds to step 1 of STPA (see Section 5.1), that is, identification of mishaps (crash types) and safety requirements (recall that STPA defines hazards as violations of safety requirements). Steps 2 and 3 of behavior safety assurance aim at assuring that the specified behavior is safe. The resulting behavioral specification is input to SOTIF (Section 5.4) and FuSa (Section 5.5) assurance, which correspond to parts of steps 2 and 3 of STPA. SOTIF and the early stage of FuSa determine the unsafe control actions (step 2 of STPA) and then use an overall control structure of the system (such as in Figure 1) to determine causes of unsafe actions (step 3.a of STPA). SOTIF focuses on causes related to technology performance limitations and foreseeable misuse, and FuSa focuses on causes related to E/E system failures. Both SOTIF and FuSa provide measures to address the respective causes (step 3.b of STPA).

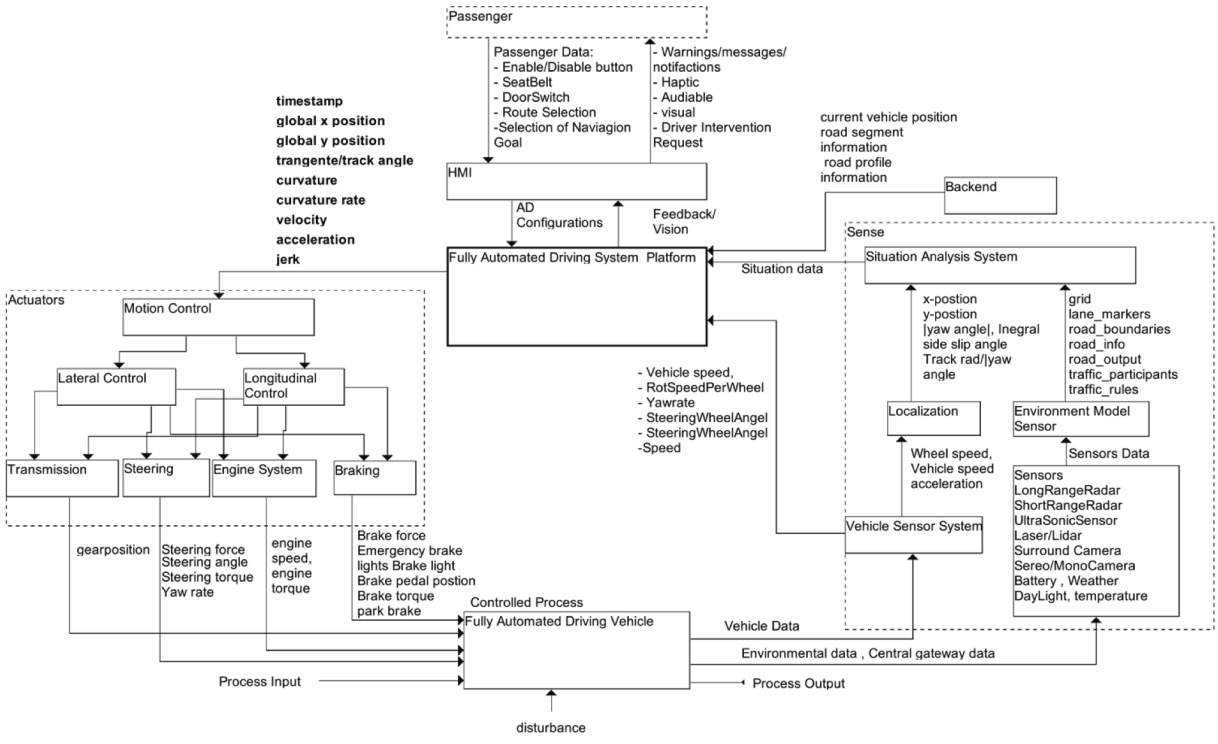

Figure 1 Sample control structure for an ADS-equipped vehicle (Figure 5 from [ALW17])

The following two subsections describe each of the four steps of driving behavior safety assurance.

5.3.1 Identification of Crash Types and Safety Requirements on Driving Behavior

The crash types involving ADS-operated vehicles are the same as those involving only human-operated vehicles (see Table 1). While the causes of the unsafe actions are different for ADS and human drivers, the outcomes in terms of collision types and collision configurations and types of noncollisions (see Section 3 in [S18] for a detailed analysis) are the same.

Further, the safety requirements on driving behavior of an ADS-operated vehicle, which is to operate in traffic consisting of both human-operated and other ADS-operated vehicles, can be derived from the safety requirements on driving behavior of a human-operated vehicle. This derivation and the resulting safety requirements are documented in the companion document "ADS High-Level Quality Requirements Analysis – Driving Behavior Safety" [S18].

The resulting safety requirements on driving behavior of an ADS-operated vehicle are classified into five categories, ordered by priority from highest to lowest [S18]:

1. *Vehicle stability*: Skid and roll stability is required for a vehicle to be controllable. Loosing control may result in colliding with other road users or objects, skidding off the roadway, and rollover.
2. *Assured clear distance ahead (ACDA)*: ACDA is the path distance ahead of the subject vehicle that the ADS can assure to be clear for driving and within which the ADS can bring the vehicle to a halt. ACDA is the minimum standard of care in driving in common law [LOT98]. Measures of ACDA include different types of sight distances, which depend on road geometry and the executed maneuver, and the stopping sight distance. Violating ACDA may lead to a crash, such as a collision, but also a noncollision (e.g., an immersion).
3. *Minimum separation*: The ADS has to assure minimum separation between the subject vehicle and other dynamic and static objects. Multiple measures characterize separation, including distance gaps, time gaps, time to collision, and lateral clearance. Minimum separation has to include sufficient safety margin to accommodate perception, prediction, and control uncertainties. The target values for many of the separation measures are maneuver- and situation-specific. Violating minimum separation may lead to a crash, typically a collision.
4. *Traffic regulations*: Traffic regulations are formal traffic rules required by law in a given geographic area. The majority of traffic regulations are safety-related. They control traffic conflict resolution, such as yielding rules at intersections and passing rules, and prescribe how different road users use the roadway, such as specifying traffic direction, lane restrictions, and parking restrictions. Violating traffic regulations, such as running a STOP sign or a read light, may lead to a crash.

5. *Driving best practices*: Driving best practices are informal traffic rules that refine and complement the formal rules. Examples include rules about how early to signal turns and how to respond to tailgating. Among others, an ADS-operated vehicle should use best practices to anticipate, recognize, and properly respond to likely mistakes of human road users such as those identified in the NHTSA pre-crash scenarios (see Section 4 in [S18]). Disregarding best practices may increase crash risk.

Both ADS- and human-operated vehicles are subject to vehicle stability, ACDA, and minimum separation requirements, which are mainly imposed by laws of physics. The minimum separation between two ADS-operated vehicles may be smaller than when at least one of the vehicles would be human-operated (except for platooning), because the perception-reaction time of an ADS may be faster than that of a human driver. However, the same calculation of ACDA, but with the varying driver or ADS-specific perception-reaction time, braking delay, and braking performance parameters, is used to determine the minimum separation for ADS- or human-operated vehicles, as discussed in the companion document [S18]. Further, in the considered target scenario of substituting the human driver with an ADS for some of the vehicles in traffic, the ADS-operated vehicles are expected to follow the same traffic rules as human drivers. This approach, which has been codified by law in some countries, such as in Germany [GA17], allows human drivers to predict the behavior of ADS-operated vehicles to the extent that they are familiar with the traffic rules. Special traffic rules for ADS-operated vehicles are likely to be introduced in future. For example, the SAE is currently preparing a standard, J3134 “Automated Driving System Lamps”, for external lighting to be used for indicating that a vehicle is operated by an ADS and to signal giving and taking way to pedestrians. This new type of external lighting, specific to ADS-operated vehicles, may be required by law in future, similar to how directional signal lights are required today. The fifth category of safety requirements, best practices, initially consists of existing expert-recommended behaviors for human drivers that are also applicable to ADS-operated vehicles, such as dealing with animals in the roadway or tailgating by human-operated vehicles. This category is open-ended, and is likely to grow with best practices that may be specific to ADS-operated vehicles, such how to deal with “robot bullying” by human road users, or how to react to new situations (where a human driver would simply apply “common sense”). The traffic data collected by ADS-operated vehicles will drive the development of these best practices in future, which will likely be geographically and culturally specific.

The detailed safety requirements in the five categories depend on the driving situation or scenario, and in particular the risk category: *normal driving*, *near-crash*, *crash*, and *fallback*. [ODD18].

In *normal driving* situations and scenarios, the safety requirements from all five categories apply. Unreasonable risk is avoided by ensuring vehicle stability with a significant safety margin, respecting the ACDA rule and minimum separation requirements, adhering to traffic rules, such as not violating red lights, and applying defensive driving practices such as increased following distance at night. Beyond the basic requirement of eliminating unreasonable risk, additional risk reduction, that is, *extra safety*, may be traded off for other objectives in normal driving, such as

for individual progress. For example, on a free roadway and in perfect driving conditions, it may be slightly less risky to go somewhat slower than the speed limit, but the ADS could choose to match the speed limit to improve individual progress. That choice would not be unreasonably riskier than going slightly slower. Similarly, it would be acceptable to take a faster but busier route even if it would be slightly riskier due to the increased traffic.

In contrast, *emergency* situations and scenarios, that is, near-crash, crash, and fallback ones, require performance of emergency maneuvers, which aim at *minimizing the crash risk*, and thus may disregard other driving qualities, such as progress, comfort, and fuel efficiency.

In *near-crash* situations, maintaining vehicle stability is required to ensure controllability, but the ADS may perform crash avoidance maneuvers at the edge of the stability envelope, that is, with little or no safety margin, if required to avoid a crash. In near-crash situations, minimum separation is likely already violated (possibly due to actions of other road users). Further, the crash avoidance maneuver may need to violate traffic rules. For example, the subject vehicle may need to charge forward at red light to avoid being rear-ended, this emergency maneuver may be the safest choice provided the ADS of the subject vehicle can ensure that no vehicles are approaching the intersection from the sides. If the ADS could not assure that, for example, due to limited visibility, then being rear-ended at low speed could be safer than risking a front-to-side collision at high speed.

In *crash* situations, maintaining (or regaining) vehicle stability remains priority while executing an emergency maneuver in order to mitigate the crash, such as emergency braking to reduce impact velocity and emergency steering to avoid hitting a high-risk target (best practice). Again, current formal traffic rules do not apply to these situations; however, it is conceivable that formal crash mitigation rules may be developed for ADS-operated vehicles in future.

Finally, in *DDT-fallback* situations and scenarios, the ADS is about to or has left its Operational Design Domain (ODD) or has experienced a failure that necessitates fallback performance that will transition the vehicle into a *minimal risk condition* [LA]. The fallback maneuver should respect the requirements in all five categories, if feasible, but additional best practices specific to fallback performance, such as where it is safe to stop under given circumstances, should be identified and applied. In particular, emergency stopping in a lane of traffic may pose unreasonable crash risk. The human driver task analysis [MA70] provides some best practices for emergency stops.

The basic motion control task catalog [BM18] and maneuver catalog [M18] refine the high-level safety requirements on driving behavior for specific driving tasks and situations.

5.3.2 Verification and Validation of Specified Driving Behavior and Residual Risk Assessment

The specified driving behavior must be verified with respect to the safety requirements on driving behavior, and subsequently validated and assessed for residual risk, within the entire ODD.

The objective of the verification is to assure that the specified behavior satisfies the safety requirements on driving behavior that were identified in Step 1. The verification can be achieved using a range of methods, including inspections and walkthroughs, prototyping and testing of prototypes, and using formal methods (e.g., [RKH17]).

The objective of the validation is to assure that the behavior specified by the safety requirements does not cause unreasonable risk of crashes in real-life use cases. This can be achieved using a combination of simulation, closed course, and field tests. Long-term field tests are of particular importance to validate the assumptions on the driving environment, including road user behavior. Given the open-ended nature of the driving environment and uncertainty of road user behavior, the driving behavior safety analysis could establish validation targets as part of Step 1, similar to step 2 in the SOTIF process (Section 5.4).

The objective of the residual risk assessment is to review the previous behavior safety assurance steps and evaluate the acceptability of the residual risk considering the findings of these steps. The step could also include a statistical argument about the residual risk, such as using the validation targets and rare event theory.

The verification, validation, and residual risk assessment steps are analogous to the corresponding steps in SOTIF (Section 5.4), but focused on assuring the adequacy of the specified behavior rather than performance limitations due to sensor technologies and algorithms.

5.4 Safety of The Intended Functionality (SOTIF)

Safety of The Intended Functionality (SOTIF) is defined as “absence of unreasonable risk due to hazards caused by performance limitations of the intended behavior or by reasonably foreseeable misuse by the user.” [SOT](3.5)

The SOTIF standard is an extension of ISO 26262 that targets unsafe actions due to performance limitation of the “intended behavior,” which is the “specified behavior including interaction with other systems and functions” [SOT](3.1). Performance limitations are insufficiencies of the implemented functions due to technology limitations, such as sensor performance limitations and noise, limitations of algorithms (e.g., machine learning), and limitations of actuator technology. Hardware and software failures are addressed by ISO 26262 and are out of scope of SOTIF. SOTIF also addresses unsafe actions due to foreseeable misuse by the user, such as user confusion, user overload, and user overconfidence. The current SOTIF standard targets automation levels 0, 1 and 2; it also states that the method can be applied to levels 3, 4 and 5, but additional measures may be necessary.

The SOTIF assurance process starts with a functional description of the system as input, which includes [SOT]:

1. The goals of the intended function;
2. The description and behavior of the functions and functionalities;
3. The dependencies on, and interaction with
    - other vehicle functions and systems;
    - the car driver and passengers;
    - relevant environmental conditions;
    - the interfaces with the road infrastructure;
4. The use cases in which the system is activated;
5. The concepts and technologies for the system and sub systems;
6. The level of automation / authority over the vehicle dynamics;
7. The limitations and their countermeasures;
8. The system architecture supporting the countermeasures;
9. The degradation concept; and
10. The warning strategies.

The SOTIF process extends HARA, safety concept design, and verification and validation testing of the ISO 26262 process. The process consists of the following main steps [SOT]:

1. *Identification of hazardous events* (Clause 6, Objectives 1,3): The first step is to identify hazardous events, which are combinations of hazards (or more precisely *hazardous control actions*) and driving situations. An example would be unintended emergency braking by the subject vehicle while being followed closely by another vehicle, which could lead to a rear-end collision. The hazardous events with severity other than S0 *and* controllability other than C0 are considered for further analysis (using the ISO 26262 classification of severity and controllability).
2. *Establishment of validation targets* (Clause 6, Objective 2): The purpose of this step is to establish performance targets for the analyzed function that would be used as acceptance criteria, such as the maximum acceptable probability of unintended braking per kilometer driven. The validation targets may be derived based on the performance of similar systems that are already in the field, human driver performance from traffic safety statistics, and expert judgment.
3. *Identification of triggering events* (Clause 7, Objective 1): This step identifies scenarios that may trigger the unsafe action, such as emergency braking triggered by a radar reflection from a soda can on the roadway or missed emergency braking because of sun glare or missed detection by a perception algorithm. The standard provides a checklist of possible triggering events related to sensor technologies, algorithms, actuator technologies, and external conditions (such as bad weather or unusual traffic situation).
4. *Evaluation of triggering events* (Clause 7, Objective 2): The purpose of this step is to assess the likelihood of the triggering events and determine their acceptability by comparing their likelihood with the validation targets.
5. *Functional modification* (Clause 8): For those triggering events whose likelihood does not meet the validation target, the functional specification needs to be modified, either by

improving the function, e.g., increasing sensor or algorithm performance, or by restricting the ODD of the system.

6. *Verification* (Clause 10): The purpose of this step is to verify that the system and the components perform as specified in the known potentially unsafe scenarios, and that they are covered sufficiently by the tests within the entire ODD. Verification uses a combination of unit and integration tests, requirements-based tests, robustness tests, and simulation and vehicle-level tests.
7. *Validation* (Clause 11): The purpose of this step is to show that the system and the components do not cause an unreasonable level of risk in real-life use cases. Validation uses a combination of requirements-based tests, simulation, closed-course and long-term field tests.
8. *Evaluation of residual risk* (Clause 12): The purpose of this step is to review the SOTIF steps and evaluate the acceptability of the residual risk considering the findings of the SOTIF steps.

The SOTIF process is iterative. Step 5 (functional modification) may be executed after the SOTIF evaluation (steps 1-4) and also after any of the steps 6, 7, and 8, if needed. Each functional modification triggers renewed SOTIF evaluation, followed by verification, validation, and evaluation of residual risk.

Step 1 of SOTIF corresponds to step 2 in STPA. Steps 3-4 of SOTIF correspond to step 3.a in STPA. Finally, step 5 of SOTIF corresponds to step 3.b in STPA.

5.5 Functional Safety (FuSa)

Functional safety (FuSa) of ADS is defined as absence of unreasonable risk of crashes due to hazards caused by malfunctioning behavior of the ADS.

ISO 26262 defines *malfunctioning behavior* [FuS](1.73) as “failure (1.39) or unintended behavior of an item (1.69) with respect to its design intent.” In particular, malfunctioning behaviors of the ADS are failures or unintended behaviors of ADS with respect to the intended functionality of the ADS. ISO 26262 defines *intended functionality* (1.68) as “behaviour specified for an item (1.69), system (1.129), or element (1.32) excluding safety mechanisms (1.111)” [FuS]. In essence, FuSa focuses on addressing hazards due to failures of the E/E system, including hardware and software failures.

ISO 26262 provides a complete functional safety assurance process, which covers system level activities, such as hazard analysis and risk assessment, the development of a safety concept, and system-level verification and validation, and also fulfilling functional safety requirements in hardware and software development and verification and validation activities.

The remainder of this section describes the *hazard analysis and risk assessment* (HARA), which is performed early in the functional safety assurance process. HARA shares the discovery

of hazardous events with SOTIF, but focuses on different causes of hazardous events than SOTIF. SOTIF targets mainly performance limitations of sensor technologies and algorithms and foreseeable misuse, whereas FuSa HARA targets hardware and software failures. FuSa HARA also includes hazards due to behavior specification deficiencies in its scope, but does not provide concrete guidance for discovering and addressing these deficiencies for automated driving. Driving behavior safety covers the latter aspect (Section 5.3).

More precisely, ISO 26262 defines *HARA* [FuS](1.58) as “a method to identify and categorize hazardous (1.57) events of items (1.69) and to specify safety goals (1.108) and ASILs (1.6) related to the prevention or mitigation of the associated hazards in order to avoid unreasonable risk (1.136).” In FuSa, a hazard [FuS](1.57) is “a potential source of harm (1.56) caused by malfunctioning behavior (1.73) of the item (1.69).” A *hazardous event* (1.59) is “a combination of a hazard (1.57) and an operational situation (1.83).” ISO 26262 defines an *operational situation* [FuS](1.83) as “a scenario that can occur during a vehicle's life,” such as driving, parking, or maintenance. In contrast to ISO 26262, SOTIF makes a distinction between situation and scenario, as also discussed elsewhere [ODD]. An Automotive Safety Integrity Level (ASIL) [FuS](1.6) is “one of four levels to specify the item's (1.69) or element's (1.32) necessary requirements of ISO 26262 and safety measures (1.110) to apply for avoiding an unreasonable residual risk (1.97), with D representing the most stringent and A the least stringent level.” ASILs are used to define development requirements on functions whose failures may lead to hazards at the given ASIL.

ISO 26262 defines the following steps for HARA [FuST]:

1. Identification of analysis participants (experts, stakeholders, engineers, etc.)
2. Identification of possible functional faults of the item (e.g., using FMEA or HAZOP)
3. Selection of situations or scenarios under consideration
4. Alignment of the risk assessment matrix
    4.1. Combination of hazards and situations into hazardous events
    4.2. Description of the expected impacts to involved persons (note that ISO 26262 does not consider impacts on property)
    4.3. Determination of risk parameters (severity, controllability, and exposure)
    4.4. Determination of the potential risk (ASIL) from the standardized risk matrix
5. Evaluation of the results
    5.1. Determine the worst-case scenario per malfunction (top event)
    5.2. Define safety goals to prevent the top events with ASIL >= A
6. Verification of the analysis results

For HARA applied to ADS intended functionality, step 2 needs to consider malfunctioning behaviors such as absence of the specified behavior in operation, presence of unspecified behavior in operation, and presence of required behavior but with degraded performance. In essence, this step identifies potentially unsafe control actions in the sense of STPA; the functional safety concept phase of FuSa that follows HARA determines the potential causes for the unsafe control actions due to faults, failures, and their interactions in the elements of the control structure. Further, step 3 selects all situations and stories from the scenario catalog for

the ODD in order to create hazardous events. Step 4 assesses risks of each hazardous event by determining their severity, controllability, and exposure, and Step 5 derives *functional safety requirements* that will need to be addressed by the subsequently developed functional safety concept. A *functional safety requirement* (1.53) is defined as “a specification of implementation-independent safety (1.103) behavior, or implementation-independent safety measure (1.110), including its safety-related attributes.” Safety-related attributes include information about ASIL, that is, severity, exposure, and controllability. Note that ISO 26262 defines controllability as “ability to avoid a specified harm (1.56) or damage through the timely reactions of the persons involved, possibly with support from external measures (1.38).” It also states that “persons involved can include the driver, passengers or persons in the vicinity of the vehicle's exterior.” An ADS-operated vehicle does not have a driver, but involved persons may include a (possibly remote) fallback-read user, passengers, or persons in the vicinity of the vehicle's exterior. For example, a sudden hard braking by an ADS-operated vehicle creates a hazard of a rear-end collision. If the following vehicle is human-operated, the driver of that vehicle is considered an involved person, who may be able to control the hazard.

References

Note: Key references are listed in Section 2.

[ALW17] A. Abdulkhaleq, D. Lammering, S. Wagner, J. Röder, N. Balbierer, L. Ramsauer, T. Raste, H. Boehmert. A Systematic Approach Based on STPA for Developing a Dependable Architecture for Fully Automated Driving Vehicles. In 4th European STAMP Workshop 2016, Procedia Engineering 179 (2017) 41 – 51

[BM18] K. Czarnecki. Automated Driving System (ADS) Task Analysis – Part 1: Basic Motion Control Tasks. Waterloo Intelligent Systems Engineering Lab (WISE) Report, University of Waterloo, 2018

[CC] Model Minimum Uniform Crash Criteria (MMUCC) Guideline. Fourth edition, DOT HS 811 631, Governors Highway Safety Association (GHSA), July 2012

[Eva04] L. Evans. Traffic safety. Science Serving Society of Bloomfield Hills, Michigan, 2004

[FuS] Road vehicles — Functional safety. ISO 26262:2011

[FuST] Functional safety according to ISO 26262. Module K3 – Concept Development. SGS TUV Saar, 2013

[GA17] K. Czarnecki. English Translation of the German Road Traffic Act Amendment Regulating the Use of “Motor Vehicles with Highly or Fully Automated Driving Function” from July 17, 2017, https://www.researchgate.net/publication/320813344

[GBF14] S. Geyer, M. Baltzer, B. Franz, S. Hakuli, M. Kauer, M. Kienle, S. Meier, T. Weissgerber, K. Bengler, R. Bruder, F. Flemisch, H. Winner. Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance. IET Intelligent Transport Systems, vol. 8, no. 3, 2014, pp. 183-189

[LA] Surface Vehicle Recommended Practice — Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE J3016:SEP2016

[Lev03] N. Leveson. A New Approach to Hazard Analysis for Complex Systems. In Proceedings of International Conference of the System Safety Society, Ottawa, August 2003

[Lev11] N. Leveson. Engineering a Safer World: Systems Thinking Applied to Safety. MIT Press, 2011

[M18] K. Czarnecki. Automated Driving System (ADS) Task Analysis – Part 2: Structured Road Maneuvers. Waterloo Intelligent Systems Engineering Lab (WISE) Report, University of Waterloo, 2018

[ODD18] K. Czarnecki. Operational Design Domain for Automated Driving Systems – Taxonomy of Basic Terms. Waterloo Intelligent Systems Engineering Lab (WISE) Report, University of Waterloo, 2018

[RKH17] A. Rizaldi, J. Keinholz, M. Huber, J. Feldle, F. Immler, K. Althoff, E. Hilgendorf, and T. Nipkow. Formalising and Monitoring Traffic Rules for Autonomous Vehicles in Isabelle/HOL. In International Conference on Integrated Formal Methods. Springer, 2017, pp. 50-66

[S18] K. Czarnecki. Automated Driving System (ADS) High-Level Quality Requirements Analysis – Driving Behavior Safety. Waterloo Intelligent Systems Engineering Lab (WISE) Report, University of Waterloo, 2018

[SOT] Road vehicles — Safety of the intended functionality. Working Document, ISO/WD-PAS 2148.1:2017-03

[TAC] Manual on Classification of Motor Vehicle Traffic Accidents. ANSI D16.1-2017 (8th edition)

[UMR15] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, M. Maurer. Defining and substantiating the terms scene, situation, and scenario for automated driving. In Proc. 2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC), 2015, pp. 982-988