Cite as:

M. Loba, N. F. Salem, A. Dotzler, and M. Maurer, “Toward a Harmonized Approach – Requirement-based Structuring of a Safety Assurance Argumentation for Automated Vehicles,” submitted for publication.

BibTeX:

```
@inproceedings{loba_2025,
      author={{Loba}, Marvin and {Salem}, Nayel Fabian and {Nolte}, Marcus and {Dotzler},
          Andreas and {Maurer}, Markus},
      title={{Toward} a {Harmonized} {Approach} -- {Requirement}-based {Structuring} of a {
          Safety} {Assurance} {Argumentation} for {Automated} {Vehicles}},
      year={2025},
      publisher={submitted for publication}
}
```

Toward a Harmonized Approach – Requirement-based Structuring of a Safety Assurance Argumentation for Automated Vehicles

Marvin Loba*, Nayel Fabian Salem*, Marcus Nolte*, Andreas Dotzler†, Markus Maurer*

*TU Braunschweig
Institute of Control Engineering, Braunschweig, Germany

†MAN Truck & Bus SE, Munich, Germany

***Abstract*—Despite increasing testing operation on public roads, media reports on incidents show that safety issues remain to this day. One major cause factoring into this circumstance is high development uncertainty that manufacturers face when deploying these systems in an open context. In particular, one challenge is establishing a valid argument at design time that the vehicle will exhibit reasonable residual risk when operating in its intended operational design domain. Regulations, such as the European Implementing Regulation 2022/1426, require manufacturers to provide a safety assurance argumentation for SAE-Level-4 automated vehicles. While there is extensive literature on assurance cases for safety-critical systems, the domain of automated driving lacks explicit requirements regarding the creation of safety assurance argumentations. In this paper, we aim to narrow this gap by elaborating a requirement-based approach. We derive structural requirements for an argumentation from literature and supplement these with requirements derived from stakeholder concerns. We implement the requirements, yielding a proposal for an overall argumentation structure. The resulting "safety arguments" argue over four topic complexes: The developed product, the underlying process including its conformance/compliance to standards/laws, as well as the argumentations' context and soundness. Finally, we instantiate this structure with respect to domain-specific needs and principles.**

***Index Terms*—safety argumentation, automated vehicles**

I. Introduction

In recent years, the testing operations of vehicles equipped with SAE-Level-4 automated driving systems has been advanced steadily on public roads, with growing fleets and expanding operational design domains. Consequently, the question arises as to why automated road vehicles have not yet been commercialized on large scale.

One reason lies in uncertainty when it comes to deploying such systems in an open context[1] like the road traffic system. Automated vehicles are exposed to various kinds of uncertainty, e.g., regarding measurements or the prediction of the behavior of other road users. Knowledge gaps are inevitable, resulting in incomplete specification of requirements which, in turn, condition incomplete testing. These functional and systemic causes lead to an inherent risk to automated vehicles' operation that can be reduced but never eliminated [2], [3].

Due to these effects of uncertainty, established practices of mainly consolidating the evidence stemming from activities in the safety lifecycle is not sufficient any longer to target the preparation of a valid basis for releasing SAE-Level-4 systems. Instead, to account for uncertainty, there is a need for a coherent argument that makes assumptions explicit regarding how the absence of *unreasonable risk*[2] is achieved and argues how valid these assumptions remain during field operation.

Frequently also referred to as "safety case" (see section II-A for a terminological delimitation), one common approach to respond to this task is a "safety assurance argumentation". Crafting such an artifact is expected by regulation [5, Annex I, Appendix 1, Part 2, 1.1] and standards [4], [6]–[8]. Although it is possible to realize argumentations at different levels of formalization, a semi-formal representation (e.g., Goal Structuring Notation (GSN), see [9]; originally proposed by Kelly [10]) appears to be a suitable compromise, as a textual degree of freedom is largely sustained while concepts like hierarchization and modularization are utilized to manage complexity of the argumentation.

Safety assurance argumentations for complex systems have been comprehensively researched and addressed by literature for decades [10]–[14]. Regarding the safety assurance of automated vehicles, emerging standards, research publications, and best practices impact the evolving state of the art in the domain – carrying implicit knowledge on the underlying line of argumentation which yields the processes, methods, and requirements presented in respective documents. Nonetheless, although extensive literature deals with different aspects on creating safety assurance argumentations, the state-of-the-art lacks an explicit provision of requirements for structuring a GSN-based safety assurance argumentation – especially when aiming to construct an argumentation that accounts for needs and principles that are particular relevant to the domain of

This work was supported by the German Federal Ministry for Economic Affairs and Climate Action within the project "Automatisierter Transport zwischen Logistikzentren auf Schnellstraßen im Level 4 (ATLAS-L4)".

[1] Refers to an environment that cannot be fully specified at design time, either due to its complexity, unpredictability, or temporal development [1].

[2] "Risk judged to be unacceptable in a certain context according to valid societal moral concepts" [4, Part 1, 3.176].

automated driving.

Hence, in this paper, first, we analyze terminological inconsistencies and address these by providing an ontology (Section II-A). This is followed by an overview of related work (Section II-B). Second, we derive requirements for the creation of a safety assurance argumentation based on relevant literature and identified stakeholder concerns (Section III). Third, we provide an overview of our proposed argumentation structure as a result of the domain-specific implementation of the specified requirements (Section IV). Finally, we discuss open issues with respect to the presented approach before concluding our paper (Section V).

II. Background

A. Terminology

The terms "safety case" and "safety assurance argumentation" are often used interchangeably. A conceptual distinction is illustrated by Fig. 1 to clarify on their relationship, facilitating communication between stakeholders having concerns with respect to the documentation of system safety.

The overarching concept is an "assurance case", defined as an "auditable artifact that provides a convincing and sound argument for a claim on the basis of tangible evidence under a given context" [8, 3.1.1]. While the principles of an assurance case apply equally for different properties of a complex system whose proof is pursued [15], the specific concern for a safety case is the emergent property *safety*. Hence, the latter can be understood as a dedicated instantiation of an assurance case.

Multiple standardized definitions (e.g., [6, 4.2.37], [16, 3.15], and [4, 3.136]) exist that share certain characteristics attributed to a safety case. Correspondingly, there is an objective to prove *safety* by a structured *argument* that is supported by *evidence* and considered in a specific environment (*context*). In accordance, Fig. 1 visualizes that evidence supports the claim of sufficient safety as the argumentation objective. Yet, as safety is defined as absence of unreasonable risk in the context of road vehicles [4, Part 1, 3.132], the basis of the argumentation relies on residual risk. Implications for the starting point of the argumentation are discussed in Section IV-A.

Distinguishing a "safety assurance argumentation" from a "safety case" emphasizes the particular task of building a coherent argumentation that goes beyond consolidating evidence generated during safety assurance processes.[3] Instead, a dedicated argumentation artifact is required that demonstrates the contributions of documented work products to achieve the absence of unreasonable risk. This is pursued by systematically decomposing claims using strategies and references to evidence and context. The modeled claims as well the evidences are valid in a specific context – see [17] for a discussion of context dependency for automotive safety arguments. In line with [9], context elements especially comprise justifications of claims and assumptions that need to be made explicit at different points in the argumentation.

Thus, the safety case comprises the safety assurance argumentation, which in this paper is understood as a GSN-based model, but also the documentation associated with evidence and context elements referenced within the argumentation. This interpretation is shared by [18, 1] concerning an "assurance argument" and its instantiation as a "safety argument" when the considered property is safety. Accordingly, a safety argument forms a "safety case" once it is considered together with the "materials it references." However, we recommend the usage of "argumentation" instead of "argument", as "safety arguments" are a common label for distinct branches within an argumentation (see [13], [12]).

B. Related Work

Structure and content of assurance cases are covered by standards [8], best practices [15], and publications that provide the required "tools" like GSN [9]. With respect to developing safety cases for complex systems, comprehensive guidance is available [10], [13], [14] that deals with challenges/pitfalls and responds with methodological approaches.

In [12], [13], [19]–[21] the task of structuring a safety case is addressed, especially supported by differentiating safety argument types such as risk, confidence, and operational arguments. In [19], a layered approach for safety argumentations is proposed as an adaption of the risk/confidence argument approach – emphasizing the necessity for conceptualizing a structured approach to create safety argumentations.

The standard UL 4600 addresses the creation of safety cases for autonomous systems. Conformance to this standard promotes sufficiency of a claim-based safety case, as the standard "puts forth assessment criteria to determine the acceptability of a safety case" [6, 1.2.3]. Still, neither does it present a process nor does it direct the construction of a concrete argumentation. An example to allow for operationalizing implicit principles of UL 4600 is presented in [22], with argumentation patterns being proposed in the context of safety performance indicators.

Domain-specific literature on safety argumentations includes work in the context of functional safety [4], [17], [23] or safety of artificial intelligence [7] for road vehicles. While the aforementioned references are equally applicable to conventional vehicles, specific needs for structuring an argumentation for vehicles equipped with automated driving systems must be accounted for in particular.

Manufacturers partially disclose safety assurance approaches of automated vehicles to the public, e.g., manifested as text-based safety reports. Yet, a coherent line of argumentation is not apparent with these representations and merely implicitly captured. Publications detailing safety case approaches (e.g., [24]) indicate how certain aspects could be incorporated into an organization's internal GSN-based argumentation. Aurora [25] presents a hybrid representation that is oriented towards GSN, revealing a superordinate structure of an argumentation – yet, is rather tailored for external stakeholder communication because it can be navigated interactively but refrains from providing evidence to the claims made.

[3] This perspective is supported by requirements defined in the recently published ISO PAS 8800:2024 [7, 7.3.4 e)].

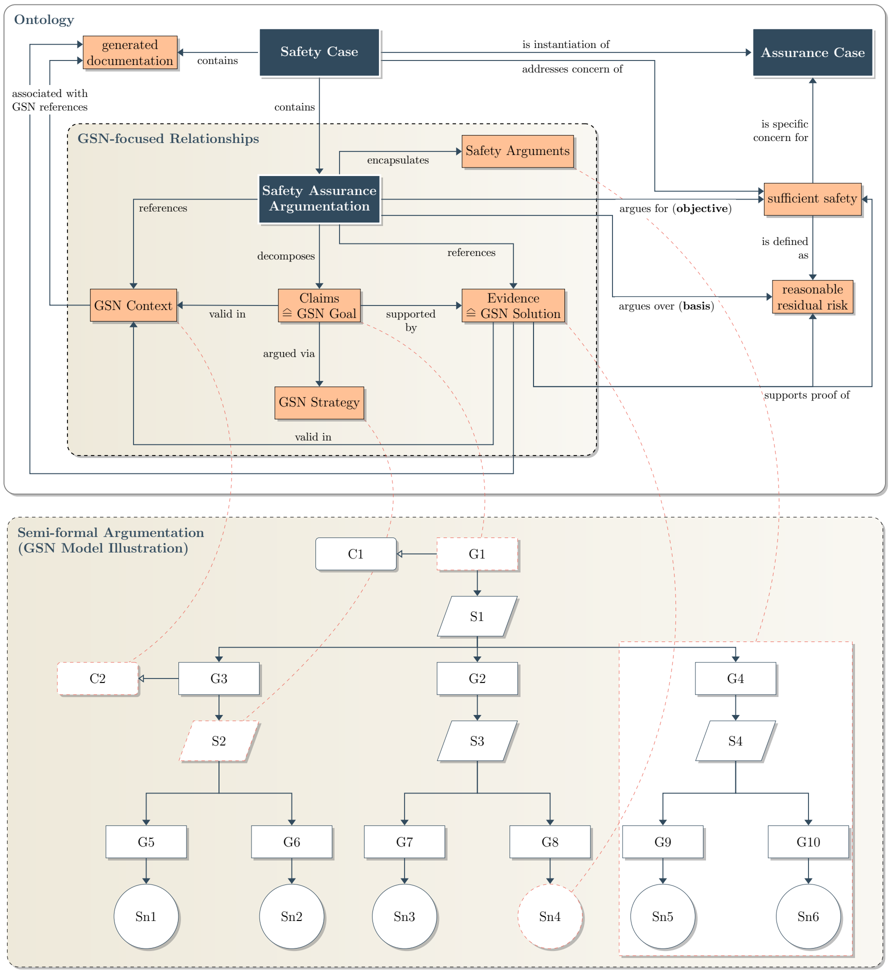

Fig. 1. Proposed ontology in the context of safety cases. ■ and 🟧 indicate artifacts and other ontology elements. GSN goals, strategies, contextual artifacts, and solutions are represented by ▭, ▱, ▢, and ○, respectively.

Different frameworks target the creation of GSN-based safety argumentations for automated vehicles [26]–[28]. Still, these lack traceability to an underlying basis, i.e., miss an evident reasoning for the resulting argumentation structure.

III. Requirements

In the following, relevant literature is examined to derive macro- and microstructural requirements in Section III-A. The former refer to requirements towards the superordinate structure, i.e., the distinction of individual safety arguments. The latter refer to requirements towards the subordinate structure, i.e., the specific contents that shall be covered in the downstream argumentation contained in the safety arguments. Based on our experience in creating and assessing safety argumentations, stakeholder concerns are identified and translated into supplementary structural requirements in Section III-B.

A. Literature-based Requirement Derivation

Following Hawkins et al. [12, p. 6], a clear distinction between "risk arguments" and "confidence arguments" is a key factor to providing compelling safety argumentations. Risk arguments shall capture the direct causal chain of risk mitigation (▶ **R1**) whereas confidence arguments shall support the confidence in the risk argument, i.e., its adequacy [11], [13].

Assurance Claim Points have been introduced to explicitly capture this relationship and indicate the assertions in a risk argument whose adequacy is argued for in separate confidence arguments [9], [12]. Hence, there are fragments to the "overall confidence argument" (▶ **R2**) distributed within the safety argumentation [12]. Assurance Claim Points are also used in automotive safety arguments [23].

Kelly [11] introduces the "conformance/compliance argument" as an additional safety argument type that argues for adherence to relevant standards, regulations, and legislation. In [15, 2:5.2.1] the categorization via aforementioned argumentation types is adopted but the authors use the label "conformance argument" only. While literature often refers to compliance with standards (e.g., [27], [28]), the Assurance Case Working Group defines conformance as "voluntary adherence to a standard, specification, guide, process or practice" and compliance as “forced adherence to a law, regulation, rule or process” [15]. However, against the assumption in [15] that compliance subsumes under conformance, we deem distinguishing the two dimensions helpful to nurture clarity in stakeholder communication and preserve the corresponding argumentation focus. This is especially relevant, as regulatory requirements are mandatory whereas arguing for conformance includes an upstream identification (and potentially disqualification) of relevant normative requirements. This is also in line with the distinction made in [29]. Thus, we propose a conformance argument (▶ **R3**) and a compliance argument (▶ **R4**) that encapsulate arguments that the development adheres to normative and regulatory requirements, respectively.

Arguing “safety through direct appeal to features of the implemented item” is often termed as product argument and arguing through “appeal to features of the development and assessment process” is often termed as process argument [4, Part 10, 5.3.1]. This classification is supported by other ISO documents [7, 8.5.1] as well as research [20], [26], [30], leading to ▶ **R5**.

Preceeding explanations yield the following requirements:

Macrostructural Requirements

The superordinate safety argumentation shall include a...

- ☑ **risk argument** that argues over risk reduction. ▶ **R1**
- ☑ distributed overall **confidence argument** that argues why elements or their assertion in the risk argument should be trusted. ▶ **R2**
- ☑ **compliance argument** that argues for adherence to regulatory requirements. ▶ **R3**
- ☑ **conformance argument** that argues for adherence to normative requirements. ▶ **R4**
- ☑ risk argument comprising a **product argument** and a **process argument**. ▶ **R5**

While the overarching goal of the risk argument is to argue over risk management, Kelly emphasizes that this is directly related to arguing over the appropriate management of hazards [10]. This includes the elimination or mitigation of all identified hazards posed by the system as well as linking it to the resulting risk. Similarly, Hawkins et al. highlight that “everything that is included as part of a risk argument must have a direct role as part of the causal chain to the hazard” [13], consequently yielding ▶ **R6**.

Palin and Habli [17, Fig. 3] consider a “Through Life Safety Argument” as part of the “High Level Vehicle Safety Argument Pattern” they present – marking another requirement that emerges from the demand to account for the operational phase, i.e., to argue over the whole system lifecycle (▶ **R7**). This concern becomes also evident in [14], as the authors extend the top-level claim of sufficient safety by the notion of “throughout its entire operational life.”[4]

Wagner and Carlan incorporate the claim that the developing organization is trustworthy in the superordinate structure of their argumentation framework [27], allocating it next to the risk argument. This consideration is related to arguing over an implemented safety culture and also addressed by UL 4600 [6] as well as Aurora [25]. This aspect is captured via ▶ **R8**.

Preceding explanations yield the following requirements:

Microstructural Requirements

The subordinate safety argumentation shall argue over...

- ☑ hazards posed by a system and discuss how these **hazards are managed** by adequate measures. ▶ **R6**
- ☑ **system lifecycle** considerations, including operational aspects related to post-deployment activities. ▶ **R7**
- ☑ how the process accounts for both procedural but also underlying organizational aspects, such as establishment of a **safety culture**. ▶ **R8**

B. Additional Requirements Based on Stakeholder Concerns

While the elicitation of macro- and microstructural requirements stems from identifying common principles according to literature, the need for additional requirements arises when stakeholder concerns are considered. Internal stakeholders (e.g., function developers, managers, or safety engineers) involved in the argumentation’s creation often possess implicit knowledge that enables comprehension of all aspects of the argumentation. In order to allow for conscious assessments by external stakeholders, e.g., in the course of audits by certification agencies or type approval authorities, we encourage to explicate this knowledge. This intention is especially tied to the objective of achieving a safety argumentation structure that is self-explanatory to the highest degree possible.

▶ R9 Contextualization Argument On the one hand, this necessiates a sufficient contextualization of the argumentation objective, i.e., providing sufficient context that, in turn, establishes an adequate argumentation basis for the downstream argumentation complexes. This contextualization can be understood as an “onboarding” of external stakeholders. From our experience, implicit knowledge is also associated with individual concepts, terminology, and abbreviations leveraged by an organization when creating the argumentation. Complementary, we deem a basic contextualization of the system of interest and its operation as important context dimensions, ideally encapsulated in a devoted contextualization argument.

▶ R10 Soundness Argument Additionally, we propose a soundness argument that argues over different measures to account for uncertainty. We consider an argument to “sound” if domain experts can judge that the remaining uncertainty from an argumentation has been sufficiently mitigated. In the

[4] In this regard, Fenn et al. [21] extend the concept of Assurance Claim Points by introducing “operational claim points” to allow for establishing operational arguments that can be understood as a runtime-focused perspective associated with the risk argument.

context of a safety assurance argumentation, different sources of uncertainty exist, including uncertainty regarding the validity of claims' inference, the scope and relevance of context, as well as the relevance and the validity of evidence [8, 4.1].

To enable comprehension of external stakeholders, the soundness argument shall argue over all applied methods that were used for the creation and maintenance of the argumentation to ensure its soundness. As already introduced, Assurance Claim Points can be utilized to reduce uncertainty in the appropriateness of GSN elements and their assertions within a graphical argument. Hence, the soundness argument may include the reasoning of the "overall confidence argument" (see [13]) where this reasoning provides insights on the selection of elements in the risk argument that are associated with Assurance Claim Points but also on explanations how the aggregation of Assurance Claim Points purpoesfully contributes to an overall satisfactory level of confidence. Other measures include for example independent reviews or methods to identify and manage weaknesses (e.g., by using *challenges* and *defeaters*) as complementary steps to developing a risk argument [13]. As an example for quantitative assessments, Herd and Burton propose the use of Subjective Logic to propagate uncertainty in GSN-based argumentations [31].

Preceeding explanations yield the following requirements:

Supplementary Requirements

The superordinate safety argumentation shall include a...

☑ **contextualization argument** addressing relevant context dimensions to allow for comprehension of the downstream argumentation. ▶ **R9**

☑ **soundness argument** that argues over applied methods to account for uncertainty in the argumentation's overall validity. ▶ **R10**

IV. Argumentation Approach

Fig. 2 illustrates our proposed structure of a safety assurance argumentation that satisfies the defined requirements. In the remainder of this section, we will provide a summary of the proposed structure and describe its instantiation due to domain-specific principles in the field of automated driving. Measures argued over in the soundness argument are agnostic to the technology since the methods considered for dealing with uncertainties are argumentation-theoretical in nature. Hence, we refrain from discussing its contents in-depth.

The underlying line of argumentation follows a GSN model that is oriented towards the VVMethods project's argumentation framework in [26]. However, several aspects are adapted/extended to achieve an argumentation that fulfills all specified requirements. This includes introducing a contextualization and soundness argument, explicitly addressing conformity, or distinguishing risk acceptance criteria regarding their abstraction level, as we discuss in the subsequent subsections.

A. Top-level Claim

The claim of a system being safe has to be accompanied by a definition of what constitutes safe operation, as also suggested in [1], [12], [13], [17], [27]. The need for justifying the top-level claim is also formulated as a normative requirement, linked to comments on this justification's critical character since it "drives the assurance case's formulation" and "serves as a means for communicating" [32].

Following Fleischer [33], we argue that, from a linguistic point of view, safety is an "open signifier", leaving the term with both enabling and impeding effects on interdisciplinary communication. This is due to the term's openness, conditioning both an alleged societal consensus on the objective of deploying "safe" automated vehicles while implicit and deviating stakeholder understandings aggravate the explicit determination of a level of safety that is accepted by society. The range of stakeholder perspectives on safety and risk in the field is also discussed by Salem et al. [34].

From an engineering perspective, there is far-reaching consensus in the domain of automated driving that safety is defined as absence of unreasonable risk (see [4], [24], [35]). This definition acknowledges that inherent risk prevents achieving freedom from risk during operation. In line with [36], we deem it especially important to avoid unfulfillable stakeholder expectations of "zero risk" (associated with a "Vision Zero") by explicitly representing and communicating residual risk.

Correspondingly, we consider the absence of *unreasonable* risk as a favorable top-level claim. This approach is also taken in literature, both within the domain (e.g., [26], [28]) but also for assurance cases in general – yet, with a slightly different wording ("no intolerable risk" according to [15]). To still account for stakeholder expectations and facilitate communication by using established labels, e.g., with respect to an assessor's aim to assess whether the system is "safe" when reviewing a safety case, we propose the contextualization argument as possibility to allocate further explanation how the concepts of residual risk and safety relate.

B. Contextualization Argument

Despite the aforementioned potential to argue over the justification of the top-level claim by defining safety via risk or, conversely, relate risk to safety, content dimensions have to be contextualized so that external stakeholders can comprehend argumentation. For instance, a system's definition and a description of its operating role and environment can pose top-level contextual elements [8]. We regard the following documentation as highly relevant for automated vehicles:

- Operational Concept according to [37], including
  - An Operational Design Domain (cf. [38]), and a
  - Behavior specification and associated competencies (cf. [39], [40])
- Concept of Operations according to [37]
- System Description (cf. [4], [35])
- Concept explanations (e.g., the introduced *inherent risk*).

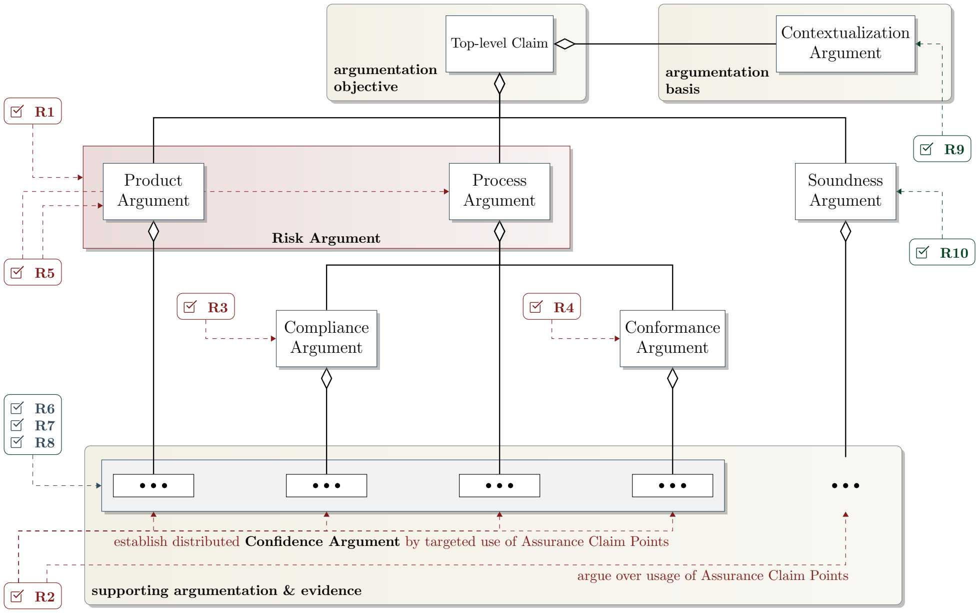

Fig. 2. Conceptual illustration for structuring a safety assurance argumentation based on the implementation of the requirements specified in this paper. The satisfy relationship (– – ▶) represents the allocation of elicitated requirements, with macrostructural and supplementary requirements posing the primary basis for distinguishing the safety arguments. The aggregation relationship (—◇) visualizes hierarchical decomposition. The microstructural requirements are associated with the supporting argumentation contained in both the product and process argument.

C. *Process Argument*

The process argument deals with aspects that contribute to answer whether the developing organization is capable of developing an automated vehicle that is free from unreasonable risk. On the one hand, this comprises covering cultural aspects. In this regard, the argumentation addresses the establishment of a safety culture [4, Part 1, 3.137] within an organization, e.g., by arguing over safety policies or safety-related trainings and onboarding procedures of employees (▶ **R8**).

Furthermore, this refers to arguing over the development process and post-deployment activities (cf. [37]) that guide the safety assurance activities. The argumentation needs to provide an adequate information basis regarding the definition and assessment of relevant sub-processes in subsequent phases as well as the proof for the deployment of these processes. This proof may be provided by evidence emerging from conducted reviews, attesting that defined processes are practiced.

The distinction of sub-processes can be derived from technical processes that subsume under the lifecycle processes consistent with system engineering standards [37]. The lifecycle perspective is not only accounted for by the associated operation and maintenance sub-processes that define post-deployment activities (▶ **R7**) – but also by arguing that the processes are scrutinized and, in case of identified deficiencies, adapted in order to achieve continuous improvement.

One major aspect factoring into the assessment of the processes' suitability is the adherence to normative and regulatory requirements. Therefore, the process definition is supported by the adjacent conformity and compliance arguments.

1) Conformity Argument: Even if the codification of the state of the art is one of the objectives of standardization, there is no agreed-upon state of the art that prescribes which normative documents are to be taken into account when developing automated vehicles. This situation is made particularly difficult by the fact that the normative landscape is dynamic. Normative documents, which exhibit varying degrees of maturity and present both complementary but also competing approaches, currently appear at high frequency [36]. Therefore, a comprehensive and critical (see also [41, 2.1.3]) analysis is required that provides a rationale for selecting normative documents, i.e., for gathering the relevant normative requirements that determine the definition of the development process.

As the analysis of standards involves multiple assumptions, it is particularly important to guarantee traceability within the argumentation. This traceability shall be established between normative requirements associated with the underlying standards in the conformity argument and the resulting decisions for the process design argued for in the process argument. As Kelly [11] explains, an overlap of a conformance argument with the risk argument should exist.

2) Compliance Argument: The compliance argument follows argumentation principles that are comparable to those of the conformity argument. Yet, arguing for adherence with regulatory requirements demands an ex-ante translation into engineering requirements in the first place. This task is especially aggravated, as legal texts are often characterized by their open nature. There still is a lack of court rulings that provide practical interpretations of legal clauses in the context of automated vehicles. Additionally, as discussed in [42], [43], challenges are present due to differences in the conceptualization of *safety* in different legal frameworks.

D. Product Argument

While the process argument provides evidence for the organization's capabilities, the product argument provides evidence that the vehicle possesses the capability to not pose unreasonable risk when operating inside its operational design domain. The main argumentation principle supporting this claim is the fulfillment of stakeholder-dependent risk acceptance criteria, i.e., the system satisfies specified risk thresholds.

To this end, we propose distinguishing between "global" and "scenario-based" risk acceptance criteria. A similar delimitation of complementary perspectives is presented in [24], [28]. The global perspective refers to a scenario-independent evaluation of the aggregated system performance by statistical means. This requires gathering data during the automated vehicle's operation in its designated operating environment. In contrast, scenario-based acceptance criteria correspond to a scenario-based risk evaluation.

From an argumentation standpoint, both argumentation strands follow the same pattern: Acceptance criteria of the respective abstraction level need to be defined in accordance with stakeholder expectations, evaluated to be met, and be maintained. Arguing for maintenance reflects in conducting field operation, gathering evidence, and ensuring that safety-related incidents do not violate the criteria after deployment.

In terms of scenario-based acceptance criteria, in line with ISO 21448 [35], we argue over residual risk in known and unknown scenarios the vehicle might encounter during its operation. On the one hand, sufficient confidence needs to be established that residual risk in unknown scenarios will not result in the violation of any acceptance criterion. On the other hand, the risk reduction in known hazardous scenarios has to be carried out sufficiently. This comprises estimation of the actual risk posed by the vehicle, specification of the tolerable risk target, and implementation of safety measures to iteratively reduce the risk until it is at least reduced to a tolerable level for the respective scenarios under consideration. Following [44], the former two activities relate to risk assessment and the latter corresponds to risk treatment. The argumentation dealing with the risk treatment is associated with a safety concept that contains the safety requirements and derived measures – consequently, yielding the argumentation that all identified hazards are sufficiently mitigated or eliminated, as suggested by literature (▶ **R6**).

V. Conclusion and Future Work

In this paper, we tackled the issue of creating a safety assurance argumentation for automated vehicles. To this end, first, we proposed an ontology that distinguishes between the artifacts "assurance case", "safety case", and "safety assurance argumentation" and connected them with relevant concepts and GSN model elements. Thereby, we aim to contribute to facilitated stakeholder communication by providing a harmonized terminology that dismantles inconsistencies.

Second, we derived requirements for structuring a safety assurance argumentation based on commonalities and differences in relevant literature. We defined supplementary requirements as a result of considering stakeholder concerns derived from our experience. We implemented all requirements, yielding a requirement-based argumentation structure.

Third, we instantiated the resulting structure based on domain-specific principles, i.e., presented the core argumentation principles of a detailed GSN model underlying this paper.

While the state of the art for safety assurance processes is not explicitly defined, normative documents capture respective requirements. In contrast, the field lacks standardization in terms of informing the creation of GSN-based safety assurance argumentations. We deem a harmonized requirement-based approach valuable to promote consistency in argumentations.

However, by nature, the structure of arguments is always characterized by subjectivity. To account for associated uncertainty, we particularly emphasize the relevance of making assumptions in the argumentation as well as underlying knowledge explicit. Thus, the introduced "soundness argument" and "contextualization argument" can pose important concepts that require further research, e.g., with respect to the questions of how to adequately represent evidence uncertainty or how beneficial contextualization can be achieved.

With respect to different stakeholders affected by automated vehicles' development and deployment, one research complex we want to investigate in the future refers to manifestations of assurance cases. It might be reasonable to have a core assurance case model that addresses basic argumentation principles which apply for different properties – and derive views for different stakeholders, such as a conformity or a compliance case for certification agencies or legal stakeholders, respectively. The idea of having multiple assurance cases for a system whose selection is based on needs and characteristics of different audiences is also supported in [8, 4.1]

As emphasized by Nolte et al. in [43], value conflicts such as the weighing of mobility against physical wellbeing is decisive when aiming to achieve public acceptance of automated vehicles. The discussed argumentation allows for considering different dimensions of harm, e.g., the harm to mobility. With risk being defined as "combination of the probability of occurrence of harm and the severity of that harm" [4, p. 3.128], the concept of stakeholder-dependent risk acceptance criteria we introduced can, hence, apply for various kinds of risk that are prioritized differently by the relevant stakeholders. In the future, we want to further research how the *budgeting* of risk can be realized and accounted for in the argumentation – for instance, the specification of tolerable target risk (see Section IV-D) requires acknowledging that the accepted risk associated with physical harm is influenced by the risk to mobility that society is willing to accept, as parametrization of speed in behavior planning determines the trade-off of physical wellbeing and mobility to all road users.

In future work we want to provide in-depth insights into our GSN model and discuss the explicit lines of argumentation. Additionally, we want to address following research questions:

- How do we conceptualize the evolving character (adaption, extension, instantiation...) of a safety assurance argumentation from a process perspective?
- What are systematical means to establish a sufficient degree of traceability between different representations for arguing safety, e.g., between safety reports, GSN-based models, and formal representations?

ACKNOWLEDGMENT

We thank Olaf Franke, Jonas Kruss, and Klaus Lamm from MAN Truck & Bus SE for their valuable input and contributions to the proposed argumentation approach.

REFERENCES

[1] S. Burton and R. Hawkins, "Assuring the safety of highly automated driving: State-of-the-art and research perspectives," Tech. Rep., 2020.

[2] M. Maurer, "Elektronische Fahrzeugsysteme – Jahresbericht: Akademisches Jahr 2017/2018," Tech. Rep., Ed.: Gerrit Bagschik.

[3] M. Nolte *et al.*, *Toward a comprehensive assurance argument for the release of automated vehicles – challenges, insights, and first results from the research project 'vvmethods'*, Presentation, 28. SAFETrans Industrial Day, 2021.

[4] *Road vehicles — Functional safety*, International Organization for Standardization International Standard 26262, 2018.

[5] *COMMISSION IMPLEMENTING REGULATION (EU) 2022/1426*, 2022.

[6] *Standard for Safety — Evaluation of Autonomous Products*, UL Standards & Engagement Standard 4600, 2023.

[7] *Road vehicles — Safety and artificial intelligence*, International Organization for Standardization Publicly Available Specification 8800, 2024.

[8] *System and software engineering — Systems and software assurance — Part 2: Assurance case*, International Organization for Standardization Standard/International Electrotechnical Commission/Institute of Electrical and Electronics Engineers International Standard 15026-2, 2022.

[9] *Goal Structuring Notation Community Standard Version 3*, The Assurance Case Working Group, 2021.

[10] T. P. Kelly, "Arguing Safety – A Systematic Approach to Managing Safety Cases," Ph.D. dissertation, University of York, 1998.

[11] T. Kelly, "Safety Cases," in *Handbook Saf. Princ.* N. Moller, S. Ove Hansson, J.-E. Holmberg, and C. Rollenhagen, Eds., Hoboken, NJ, USA: John Wiley Sons, Inc., 2018, pp. 361–385. DOI: 10.1002/9781119443070.ch16.

[12] R. Hawkins, T. Kelly, J. Knight, and P. Graydon, "A New Approach to creating Clear Safety Arguments," in *Adv. Syst. Saf.* C. Dale and T. Anderson, Eds., London: Springer London, 2011, pp. 3–23. DOI: 10.1007/978-0-85729-133-2_1.

[13] R. Hawkins, *Developing Compelling Safety Cases*, arXiv: 2502.00911, 2025.

[14] R. Hawkins *et al.*, *Guidance on the Safety Assurance of Autonomous Systems in Complex Environments (SACE)*, arXiv: 2208.00853, 2022.

[15] *Assurance Case Guidance Challenges, Common Issues and Good Practice — Version 1*, The Assurance Case Working Group, 2021.

[16] *Assuring the operational safety of automated vehicles — Specification*, British Standards Institution Publicly Available Specification 1881, 2022.

[17] R. Palin and I. Habli, "Assurance of Automotive Safety – A Safety Case Approach," in *Comput. Saf., Rel., Secur.* D. Hutchison *et al.*, Eds., vol. 6351, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 82–96. DOI: 10.1007/978-3-642-15651-9_7.

[18] M. S. Graydon and S. M. Lehman, "Examining Proposed Uses of LLMs to Produce or Assess Assurance Arguments," Tech. Rep., 2025.

[19] J. Birch *et al.*, "A layered model for structuring automotive safety arguments (short paper)," in *2014 10th Eur. Dependable Comput. Conf.*, pp. 178–181. DOI: 10.1109/EDCC.2014.24.

[20] I. Habli and T. Kelly, "Process and product certification arguments: Getting the balance right," *ACM SIGBED Review*, vol. 3, no. 4, pp. 1–8, 2006. DOI: 10.1145/1183088.1183090.

[21] J. Fenn, R. Hawkins, and M. Nicholson, "A New Approach to Creating Clear Operational Safety Arguments," in *Comput. Saf., Rel., Secur.. SAFECOMP 2024 Workshops*, A. Ceccarelli *et al.*, Eds., vol. 14989, Cham: Springer Nature Switzerland, 2024, pp. 227–238. DOI: 10.1007/978-3-031-68738-9_17.

[22] D. Ratiu, T. Rohlinger, T. Stolte, and S. Wagner, *Towards an Argument Pattern for the Use of Safety Performance Indicators*, arXiv: 2007.13807, 2024.

[23] HORIBA MIRA Ltd and Motor Industry Software Reliability Association, *MISRA: Guidelines for Automotive Safety Arguments*. 2019.

[24] F. Favarò *et al.*, *Building a credible case for safety: Waymo's approach for the determination of absence of unreasonable risk*, https://waymo.com/blog/2023/03/a-blueprint-for-av-safety-waymos, 2023.

[25] Aurora Innovation, *Aurora's safety case framework*, https://safetycaseframework.aurora.tech/gsn, 2023.

[26] J. Reich, *Assurance Argumentation Framework*, Presentation, VVM Final Event, Stuttgart, Germany, 2023.

[27] M. Wagner and C. Carlan, *The open autonomy safety case framework*, arXiv: 2404.05444, 2024.

[28] H. Kodama *et al.*, "A Case Study of Continuous Assurance Argument for Level 4 Automated Driving," in *Comput. Saf., Rel., Secur.* A. Ceccarelli, M. Trapp, A. Bondavalli, and F. Bitsch, Eds., vol. 14988, Cham: Springer Nature Switzerland, 2024, pp. 150–165. DOI: 10.1007/978-3-031-68606-1_10.

[29] S. Swaminathan, J. Wishart, J. Zhao, B. Russo, and S. Rahimi, "Adapting the Technology Readiness Level (TRL) Framework to Automated Vehicle Development," in *WCX SAE World Congr. Experience*, Detroit, Michigan, United States, 2025, pp. 2025-01–8671. DOI: 10.4271/2025-01-8671.

[30] Y. Luo, Z. Li, and M. van den Brand, "A Categorization of GSN-based Safety Cases and Patterns:" in *Proc. 4th Int. Conf. Model-Driven Eng. Softw. Develop.*, Rome, Italy: SCITEPRESS - Sci. and Technol. Publications, 2016, pp. 509–516. DOI: 10.5220/0005734305090516.

[31] B. Herd, J.-V. Zacchi, and S. Burton, "A Deductive Approach to Safety Assurance: Formalising Safety Contracts with Subjective Logic," in *Comput. Saf., Rel., Secur.. SAFECOMP 2024 Workshops*, A. Ceccarelli *et al.*, Eds., vol. 14989, Cham: Springer Nature Switzerland, 2024, pp. 213–226. DOI: 10.1007/978-3-031-68738-9_16.

[32] *System and software engineering — Systems and software assurance — Part 2: Assurance case*, Institute of Electrical and Electronics Engineers Standard – Adoption of ISO/IEC 15026-2:2011, 2011.

[33] T. Fleischer, "Safety and Acceptance – A View of Two Mysteries," Presentation, Oberseminar EFS, virtual, 2023.

[34] N. F. Salem *et al.*, "Safety and Risk – Why their Definitions Matter," in *Handbook Assisted Automated Driving*, ser. ATZ/MTZ-Fachbuch, 4th ed., in press, Wiesbaden, Germany: Springer Vieweg.

[35] *Road vehicles — Safety of the intended functionality*, International Organization for Standardization International Standard 21448, 2022.

[36] M. Nolte, M. Loba, N. F. Salem, and M. Maurer, *Herausforderungen für die Produktcompliance im Feld des automatisierten Fahrens — Überblick und kritische Diskussion der aktuellen Normenlandschaft*, H. Steege and K. Chibanguza, Eds. Nomos, in press.

[37] *System and software engineering — Systems lifecycle processes*, International Organization for Standardization Standard/International Electrotechnical Commission/Institute of Electrical and Electronics Engineers International Standard 15288, 2023.

[38] *Road Vehicles — Test scenarios for automated driving systems – Specification for operational design domain*, International Organization for Standardization International Standard 34503, 2023.

[39] N. F. Salem *et al.*, "An Ontology-based Approach Toward Traceable Behavior Specifications in Automated Driving," *IEEE Access*, vol. 12, pp. 165 203–165 226, 2024. DOI: 10.1109/ACCESS.2024.3494036.

[40] *Behaviour taxonomy for automated driving system (ADS) applications — Specification*, British Standards Institution Standard 1891, 2025.

[41] P. Koopman, A. Kane, and J. Black, "Credible Autonomy Safety Argumentation," in *Saf.-Crit. Syst. Symp. (SSS)*, Bristol, UK, 2019.

[42] M. Nolte *et al.*, *Anmerkungen zu Sicherheit und Risiken autonomer Straßenfahrzeuge — Teil 1*. C.H. BECK, NZV – Neue Zeitschrift für Verkehrsrecht, in press.

[43] M. Nolte *et al.*, *A Review of Conceptualizations of Safety and Risk in Current Automated Driving Regulation*, arXiv: 2502.06594, 2025.

[44] N. F. Salem *et al.*, "Risk management core—toward an explicit representation of risk in automated driving," *IEEE Access*, vol. 12, pp. 33 200–33 217, 2024. DOI: 10.1109/ACCESS.2024.3372860.