PEGASUS Safety Argumentation

Proposal for a framework to support an approval recommendation particularly aimed at highly automated driving functions

Summary

A method is proposed with the aim of formulating a coherent and verifiable safety argumentation in a structured and formalised manner. The PEGASUS Safety Argumentation proposes four layers for this upon which elements are located. Dimensions of integrity are proposed as quality criteria for the four layers with respect to their elements. The central assumption of the PEGASUS Safety Argumentation is: if a chain of arguments, which was created following the method of the PEGASUS Safety Argumentation, stands up to a critical examination, an approval recommendation can be given.

Fundamental Paradigms

Today’s visions of the automotive future are shaped by the idea of Automated Driving Systems (ADS). In order to transform these visions into reality, many questions still have to be answered besides the continually advancing technological development. Which requirements must Automated Driving Systems fulfil? How can safety and reliability be verified? [1]

To make a contribution to answering these questions, a framework has been developed as part of the PEGASUS joint project to support an approval recommendation, particularly for the highly automated driving functions which enable self-driving cars – the PEGASUS Safety Argumentation. Leading the activities here are three paradigms: structuring, formalisation and coherence.

Structuring is meant here as organising the PEGASUS Safety Argumentation into four layers. This structure attempts to bring more clarity into the continuing discussions regarding an approval, particularly surrounding highly automated driving functions. A differentiation must be made between two kinds of layers: layers outside the focus of PEGASUS, which can be placed within the proposed framework in order to put PEGASUS into a wider context, and

---

1 The PEGASUS research project is funded by the Federal Ministry for Economic Affairs and Energy. More information: www.pegasusprojekt.de/en
layers and their elements which directly contribute to answering the question of how to verify safety and reliability.

Formalisation within the PEGASUS Safety Argumentation means making the individual elements of a layer explicit by means of a defined, standardised and ideally established notation. The choice of the notation is dependent on the layer and the contained elements to be formalised.

Coherence is defined as the linking of individual elements within one of the postulated four layers of the PEGASUS Safety Argumentation as well as in between layers. Only once there is success in reasonably linking elements across layers, can one show how safety and reliability can be verified and how an argumentation can be made for socially accepted, highly automated mobility in the larger context.

The PEGASUS Safety Argumentation

Table 1 shows an overview of the four layers, their elements and possible formalisations. In the following, the individual layers and their elements are explained in more detail.

Table 1 - Overview of the four layers, their elements and possible formalisations

<table>
<thead>
<tr>
<th>Layers</th>
<th>Elements</th>
<th>Possible Formalisations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ADS acceptance model</td>
<td>Model for technology acceptance for highly automated driving functions</td>
<td>Structural Equation Modelling (SEM)</td>
</tr>
<tr>
<td>2 Logical structure of the safety argument</td>
<td>Safety goals, strategies, solutions, justifications, assumptions and context for achieving the safety goals and their links</td>
<td>GSN</td>
</tr>
<tr>
<td>3 Methods and tools</td>
<td>Documentation of the solutions from layer 2</td>
<td>ASPICE, SOTIF</td>
</tr>
<tr>
<td>4 Evidence</td>
<td>Results from the application of layer 3</td>
<td>e.g. OpenSCENARIO/OpenDRIVE</td>
</tr>
</tbody>
</table>

Layer 1 – ADS acceptance model

PEGASUS embeds the first layer of the ADS acceptance model in a large context. The specifics are not the focus of PEGASUS. The key element of this layer is a scientific model for describing the dependence of the social acceptance for Automated Driving Systems from
several factors. A key premise here is that individual or social acceptance cannot be explained with a single cause. This premise is in line with established models on technology acceptance such as the Technology Acceptance Model (TAM) [2] [3], the Theory of Planned Behaviour (TPB) [4], and the Unified Theory of Acceptance and Use of Technology (UTAUT) [5]. Depending on the model, different factors are postulated: Attitude, Perceived Usefulness, Perceived Ease of Use, Subjective Norms, Perceived Behavioural Control, Performance Expectancy, Effort Expectancy, Social Influence, Perceived Enjoyment. Rahman et al. examined the utility of the models mentioned for describing the acceptance of Advanced Driver Assistance Systems (ADAS) and give an overview of studies which build upon the models mentioned [6]. Rahman et al. come to the conclusion that in principle all three models can be used in the context of acceptance of ADAS. However, further work is necessary to adapt the theories and models to the domain, to develop them further and to explain a greater proportion of the variance (cf. [7], [8]).

For the formalisation, an approach from multivariate statistics such as Structural Equation Modelling (SEM) is proposed, particularly the flow charts associated with it [9]. To put it briefly, this class of procedure involves modelling interconnections between latent variables, which themselves cannot be directly measured and thus are only accessible via observed items. Since in PEGASUS the focus is on the verification of safety and reliability of highly automated driving functions, it is proposed that these be subsumed under the factor of Performance Expectancy\(^2\). This allows a connection to be made between the first layer and the second layer, the presentation of the logical structure of the safety argument. As part of the PEGASUS Safety Argumentation, layers 2, 3 and especially 4 are to be understood as an operationalisation of the Performance Expectancy factor (in particular here safety and reliability).

Layer 2 – Logical structure of the safety argument

The logical structure of the safety argument as defined by the PEGASUS Safety Argumentation is all about deriving the steps necessary to establish a verification for safety and reliability from higher level goals. The logical structure of the argumentation is made explicit in order to answer the following question: why is a result, which has been achieved in

\(^2\) Since none of the existing models are further developed and no new model is proposed as part of PEGASUS, we are attempting here to define an interface to the existing research in this field.
layer 3 or 4 for the PEGASUS Safety Argumentation, relevant for the verification of safety and reliability? In principle, various formalisations are applicable for this: an example is ISO 15026-2 [10] or the Goal Structuring Notation (GSN) [11] [12], which are also used in the example in Figure 3. Both approaches formulate action-guiding goals, which to a certain extent serve as a monitor of success in the argumentation chain. By means of a standardised graphical notation, safety goals are defined as elements of the second layer, plus strategies and solutions for achieving these safety goals. Further information regarding justifications, assumptions and context can also be consigned to these.

As an extension to the GSN Community Standard Version 1 [12], it is proposed that the individual elements of the logical structure of the safety argument (safety goals, strategies, solutions, justifications, assumptions and context) be allocated with a credibility ranking (see Table 2). This is intended to show the “maturity layer” of each element. For example, a peer reviewed and published method, which is applied for the purposes of generating evidence in layer 3 of the PEGASUS Safety Argumentation, is ranked higher than an unpublished method.

<table>
<thead>
<tr>
<th>Credibility</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>Expert estimation (unpublished)</td>
</tr>
<tr>
<td>Stage 2</td>
<td>Expert estimation (published)</td>
</tr>
<tr>
<td>Stage 3</td>
<td>Expert estimation (peer-reviewed published)</td>
</tr>
<tr>
<td>Stage 4</td>
<td>Law</td>
</tr>
</tbody>
</table>

When formulating safety goals³, we differentiate between two layers of abstraction: safety goals that can be reasonably formulated independently of a specific highly automated driving function (general safety goals), and safety goals that can only be established for a specific highly automated driving function (specific safety goals). Specific safety goals can also be understood as an application of general safety goals to a specific HAD function.

A publication for example by the NHTSA can be referred to when defining general safety goals for achieving safety and reliability [13]. The authors formulate 12 priority design elements. General safety goals can be derived from these priority design elements. Not all of the NHTSA’s proposed priority design elements are applied in the PEGASUS project⁴. The

---

³ When formulating safety goals, paying attention to the resolvability of the vocabulary is recommended.

⁴ This should not be understood as a valuation of the relevance. It is instead a result of how the PEGASUS joint project is focussed.
main goals in PEGASUS [1] can particularly be subsumed under the priority design element “validation methods” and the general safety goals derived from this element. Specific safety goals can now be derived from this general safety goal for the highly automated Autobahn Chauffeur (AC) function. The measures to derive those concrete safety goals are found in layer 3.

Layer 3 – Methods and Tools

The leading question for layer 3 of the PEGASUS Safety Argumentation is: How can one prove that the layer 2 specific safety goals have been achieved? As part of the overall PEGASUS method (see Figure 1), a continuous and flexible tool chain is being described. In PEGASUS, both, data-driven (from available sources such as accident databases, NDS studies or simulator studies) (2) and theory-led (ontological approach)(4) logical scenarios and the associated parameters are defined and stored in a database (7)-(12). Their constituent parameters are then varied (13) and played out from the database, so that they are available for tests (simulations or test drives) (14) and can be evaluated using suitable metrics (17). The description of the individual steps is not the subject of this work. Rather, it is about pointing out, how methods and tools can be linked to specific safety goals as part of the PEGASUS Safety Argumentation. This will later be depicted as an example (see Figure 3).

Figure 1- PEGASUS Method for Assessment of Highly Automated Driving Function (HAD-F)
As a formalisation, it is recommended that each evidence-producing method be documented. Standards such as ASPICE [14] or SOTIF [15] can be used for this. Documentation of the methods should include at least one of the following two elements: a specific ID and a specific name which should match the ID and name of the corresponding solution in layer 2. This ensures coherence and provides the information necessary to enable reproducibility and verifiability of the evidence-producing method and its result.

**Layer 4 – Evidence**

A result is generated from applying the evidence-producing methods located in layer 3. The challenge is now to embed the result and evaluate its contribution to achieving a specific safety goal. Only through this step can a result become evidence as intended by the PEGASUS Safety Argumentation. The leading question for layer 4 is therefore: Can a result be considered as evidence for achieving a specific safety goal? The key difference between the terms “Result” and “Evidence” in the sense of the PEGASUS Safety Argumentation is therefore how they can be assigned to specific safety goals.

The methods and tools described in layer 3 can be as widely differing as the specific safety goals that are to be achieved. This can then lead to heterogeneous results and to formats in which evidence can lie. A uniform formalisation cannot therefore be reasonably proposed for all results from layer 4. The formats OpenDRIVE [16] and OpenSCENARIO [17] are examples of formats for a possible formalisation.

The paradigm of scenario-based tests is intrinsic to the overall PEGASUS method. A complete test space described by means of (logical) scenarios could also be seen as evident in the sense of the PEGASUS Safety Argumentation, if a corresponding specific safety goal (for example the complete description of possible scenarios with which the AC could be confronted in the field) is defined. With this example of a specific safety goal, a particular challenge is to verify the completeness. If this verification cannot be made in this example, the test space can be considered as a result, but not as evidence as required by the PEGASUS Safety Argumentation. But what is the consequence of this evaluation?

At this point we arrive at a key point of the PEGASUS Safety Argumentation – the critical testing of the argumentation chains. If the argumentation chains, which connect the
individual elements of the layers, can withstand critical evaluation, an approval recommendation can be made. Two aspects are meant by this critical evaluation: one is that the chain can be checked top-down from the safety goal to the solution, i.e. it can be shown that a continuous and coherent argumentation chain for goal achievement can be compiled. The other is that the argumentation chain can in principle also be tested bottom-up from starting with the solution. The focus here is the question of whether the results generated are evident for a safety goal. It might happen that a result cannot be brought into connection with a safety goal. This could occur when an unanticipated result occurs during a test process or a method is applied which is not (yet) foreseen in layer 2 of the safety argumentation. Here it appears reasonable to account for these circumstances and as such foresee the bottom-up testing in principle in the PEGASUS Safety Argumentation. This step is explicitly not meant to adapt achieved results to safety goals in retrospect to make the results look better. Under the premise that the development of methods to approve Automated Driving Systems are still in their infancy, it should however be assumed that the experience made over time with a process like the PEGASUS Safety Argumentation will likely lead to an iterative development of all elements on all layers. The possibility for further development can, for example, be accounted for by the "Credibility" introduced in layer 2. As an example, a bottom-up argumentation for reformulation of a safety goal with low credibility can be more easily argued for than for one with high credibility. The addition of a safety goal found in this way, which was not previously anticipated, seems unproblematic. In the case of the safety goal in this example (the complete description of all possible scenarios with which the AC could be confronted with in the field), the question would be whether this goal would have a higher credibility than a possible alternative method which might show that completeness cannot be achieved. In this example, one would have to clarify what is meant exactly by completeness.

A prerequisite, which has not been mentioned yet, but is necessary for the testability of an argumentation chain – top-down or bottom-up – is (besides the consideration of the paradigms structuring, formalisation and coherence) the integrity of the elements of the layer in question. This will be examined in more detail in the following.

Integrity

Integrity in the sense of the PEGASUS Safety Argumentation is intended as a quality criterion for the individual elements of the four layers. They are oriented around the quality criteria of
empirical research and thus form a cross-section layer. The integrity does not necessarily have to correlate with the credibility, however it can be expected that an element with high credibility also has a high integrity. Different levels of integrity are imaginable in principle for the different layers and their elements. They are listed as an example in Table 3.

Table 3 - Overview of the Layers

<table>
<thead>
<tr>
<th>Layer</th>
<th>Possible levels of integrity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ADS acceptance model</td>
<td>Validity</td>
</tr>
<tr>
<td>2 Logical structure of the safety argument</td>
<td>Objectivity, Closedness, Robustness</td>
</tr>
<tr>
<td>3 Methods and tools</td>
<td>Reliability, Reproducibility</td>
</tr>
<tr>
<td>4 Evidence</td>
<td>Correctness, Representativeness</td>
</tr>
</tbody>
</table>

The validity of the Automated Driving Systems acceptance model generally refers to the level to which the contents of the features, which are to be operationalised, match (particularly in this case with regard to the factors safety and reliability) with the measurement in the test processes. Remember at this point that the layers 2, 3 and especially 4 can be interpreted as operationalisation of the factors of the model in layer 1.

![Figure 2- Schematic depiction of the 4 layers and the integrity](image-url)
Objectivity refers here to the definition of safety goals. General safety goals in particular should therefore be independent of the defining subject.

Closedness is meant here as the application of the general paradigm “Coherence” on layer 2 of the PEGASUS Safety Argumentation. This highlights the aim of achieving closed argumentation chains. If the individual elements of the argumentation chain are interpreted as nodes and the connections as edges in the sense of a graph, the aim is to achieve as few nodes not connected by edges as possible.

Robustness refers to the strength of the connection between elements of the argumentation chain, and can be thought of as a weighted connection between two nodes in a graph. This concept is closely related with “Credibility”, which to a certain extent can be interpreted as a node (or an element in the sense of the PEGASUS Safety Argumentation) in the argumentation chain.

Reliability refers to how reliable evidence-producing methods can be applied, and is closely connected to the concept of reproducibility; tests under the same conditions lead to the same results and are thus free of random errors.

On layer 4, correctness of the achieved results and their representativeness play a special role. A correct result is free of (random) errors. Representative results allow the conclusion of a spot check, as they are performed as part of the test process, to be taken as a reference to the basic population. A schematic overview of the 4 layers and the integrity as a cross-section layer is shown in Figure 2.
Example

Following the above introduction of the PEGASUS Safety Argumentation, there now follows a demonstration of how it could be applied. This example is given in a graphic form and draws from proposed formalisations. It must be considered here that this example does not claim to show a complete or sufficiently detailed argumentation chain. Instead it is intended to illustrate how the layers – described in theory – could be applied in practice.

Layer 1 (Figure 2) shows a model based on UTAUT [5], but is slightly altered and depicted without moderating effects for the sake of simplicity. The general safety goals G_02 to G_13 in layer 2 (Figures 2 and 3) are derived from the design principles [13] proposed by the NHTSA and correspond to these partly word-for-word or have only been slightly reformulated, but do not reflect the points addressed by the design principles to their full extent. In this example, G_05 “Process and procedure for assessment, testing and validation of ADS functionality with the prescribed ODD is documented and considered” from the safety goals is broken down into partial goals and pieced together with the overall PEGASUS method (G_17 – G_20). In the specific safety goals G_21 – G_24 (Figures 2 and 3), the PEGASUS Safety Argumentation is applied to the highly automated AC driving function with the example of an effect field analysis. The layers 3 and 4 are depicted in Figure 4. An overview is given in Figure 5.
Figure 3 - Example safety argumentation (layer 1 and layer 2). The four-stage indicator expresses an estimation of the credibility of each element according to Table 2.
Goal: G_05 Process and procedure for assessment, testing and validation of ADS functionality with the prescribed ODD is documented and considered.

Strategy: S_02 Process and procedure for assessment, testing and validation with the prescribed ODD is documented and considered for different life cycle stages.

Goal: G_14 Process and procedure for assessment, testing and validation of ADS functionality with the prescribed ODD is documented and considered during research stage.

Goal: G_15 Process and procedure for assessment, testing and validation of ADS functionality with the prescribed ODD is documented and considered during development stage.

Goal: G_16 Process and procedure for assessment, testing and validation of ADS functionality with the prescribed ODD is documented and considered during operational stage.

Strategy: S_03 Process and procedure to be considered for assessment, testing and validation of ADS functionality with the prescribed ODD is split into sub processes and sub procedures and documented in each case.

Goal: G_17 Requirements are derived from laws, standards, directives and guidelines or comparable sources.

Goal: G_18 Information is gathered, preprocessed and analyzed in order to define logical scenarios as well as associated parameters respecting the prescribed ODD.

Goal: G_19 Logical scenarios are created systematically and the associated parameters are varied systematically considering the prescribed ODD and stored in a suitable form.

Goal: G_20 The ADS, including the human driver, is assessed, tested and validated in terms of nominal performance.

Strategy: S_04 Information is gathered, preprocessed and analyzed from different sources analysed in order to define logical scenarios as well as associated parameters respecting the prescribed ODD.

Goal: G_21 Information gathered from test drives with AC prototypes is preprocessed and analyzed in order to define logical scenarios as well as associated parameters respecting the prescribed ODD.

Goal: G_22 Information gathered from AC simulation is preprocessed and analyzed in order to define logical scenarios as well as associated parameters respecting the prescribed ODD.

Goal: G_23 Information gathered from relevant Simulator Studies, in particular with regard to human performance is preprocessed and analyzed in order to define logical scenarios as well as associated parameters respecting the prescribed ODD.

Goal: G_24 Information gathered from relevant Natural Driving Studies (NDS) is preprocessed and analyzed in order to define logical scenarios as well as associated parameters respecting the prescribed ODD.

Goal: G_24 Information gathered from relevant Accident Databases is preprocessed and analyzed in order to define logical scenarios as well as associated parameters respecting the prescribed ODD.

Figure 3 – Example safety argumentation (continuation of layer 2)
Goal: G.24 Information gathered from relevant Accident Databases is preprocessed and analyzed in order to define logical scenarios as well as associated parameters respecting the prescribed GDD.

Strategy: S.05 Data derived from German In-Depth Accident Study (GIDAS) is preprocessed and analyzed in order to get an understanding of accidents which might be mitigated or prevented by AC. Findings from this analysis support the identification of relevant parameter and the generation of logical scenarios.

Evidence: E.01 Analysis of accidents which might be mitigated or prevented by AC is carried out.

ID: E.01 “Analysis of accidents which might be mitigated or prevented by AC”

GIDAS data is used in version XY. A graphical representation of the filters applied to data is given below. The Analysis was carried out using R in version XY with additional packages A, B and C. Source code for the analysis is attached to this document.

ID: E.01 “Analysis of accidents which might be mitigated or prevented by AC”

Findings from this analysis support the identification of relevant parameter and the generation of logical scenarios:

Insert image GIDAS UTYP overview analysis here

Figure 3 – Example safety argumentation (continuation of layers 2, 3 and 4)
Figure 1 – Overview of the example safety argumentation, referring in particular to layer 2
Sources


Contact information

<table>
<thead>
<tr>
<th>General</th>
<th>Thematic</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEGASUS Project Office,</td>
<td>on behalf of Volkswagen Group Research:</td>
</tr>
<tr>
<td>DLR - Institute of Transportation Systems</td>
<td>Automotive Safety Technologies</td>
</tr>
<tr>
<td>Jens Mazzega</td>
<td>Alexander Maus</td>
</tr>
<tr>
<td>E-Mail: <a href="mailto:jens.mazzega@dlr.de">jens.mazzega@dlr.de</a></td>
<td>E-Mail: <a href="mailto:alexander.maus@astech-auto.de">alexander.maus@astech-auto.de</a></td>
</tr>
</tbody>
</table>