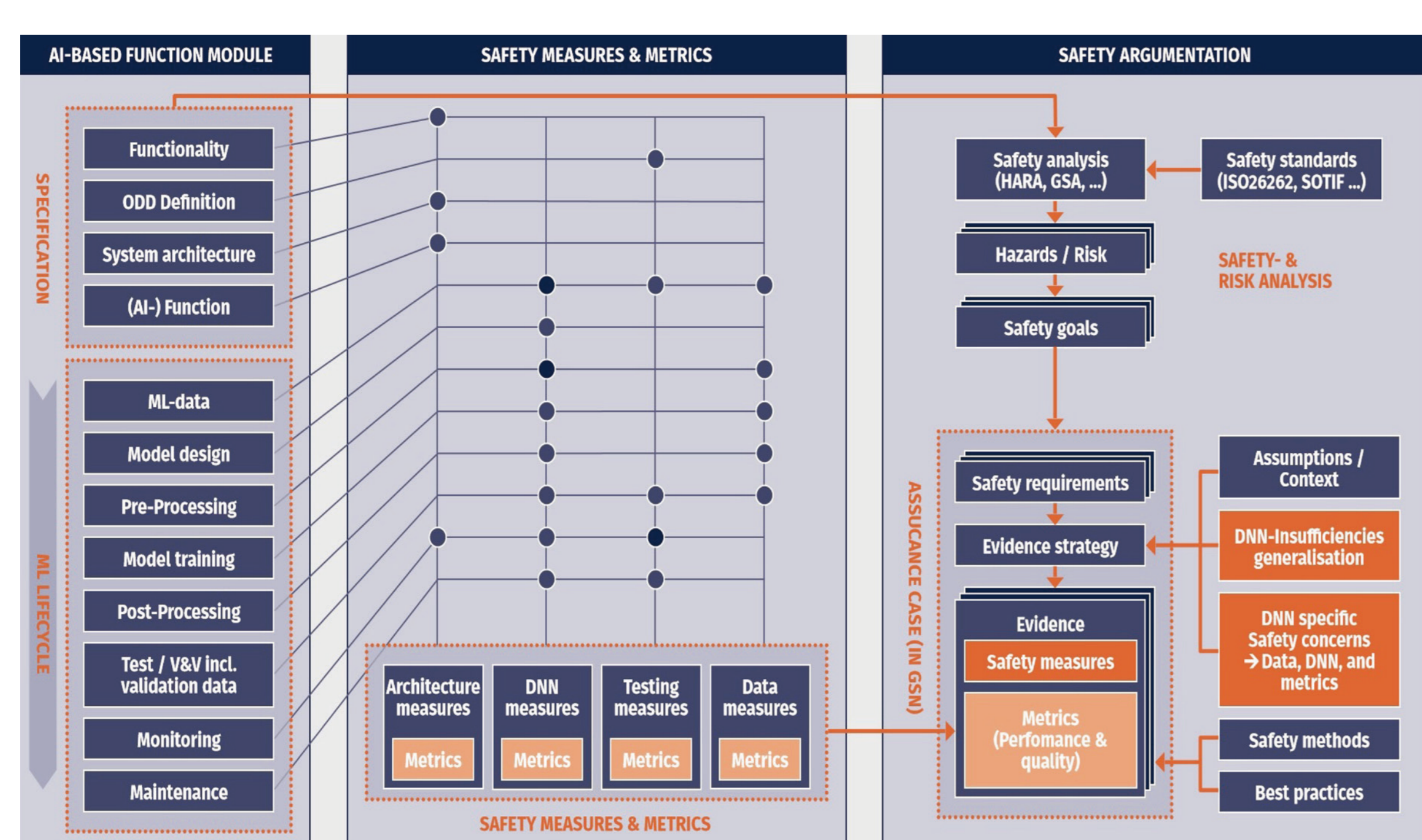
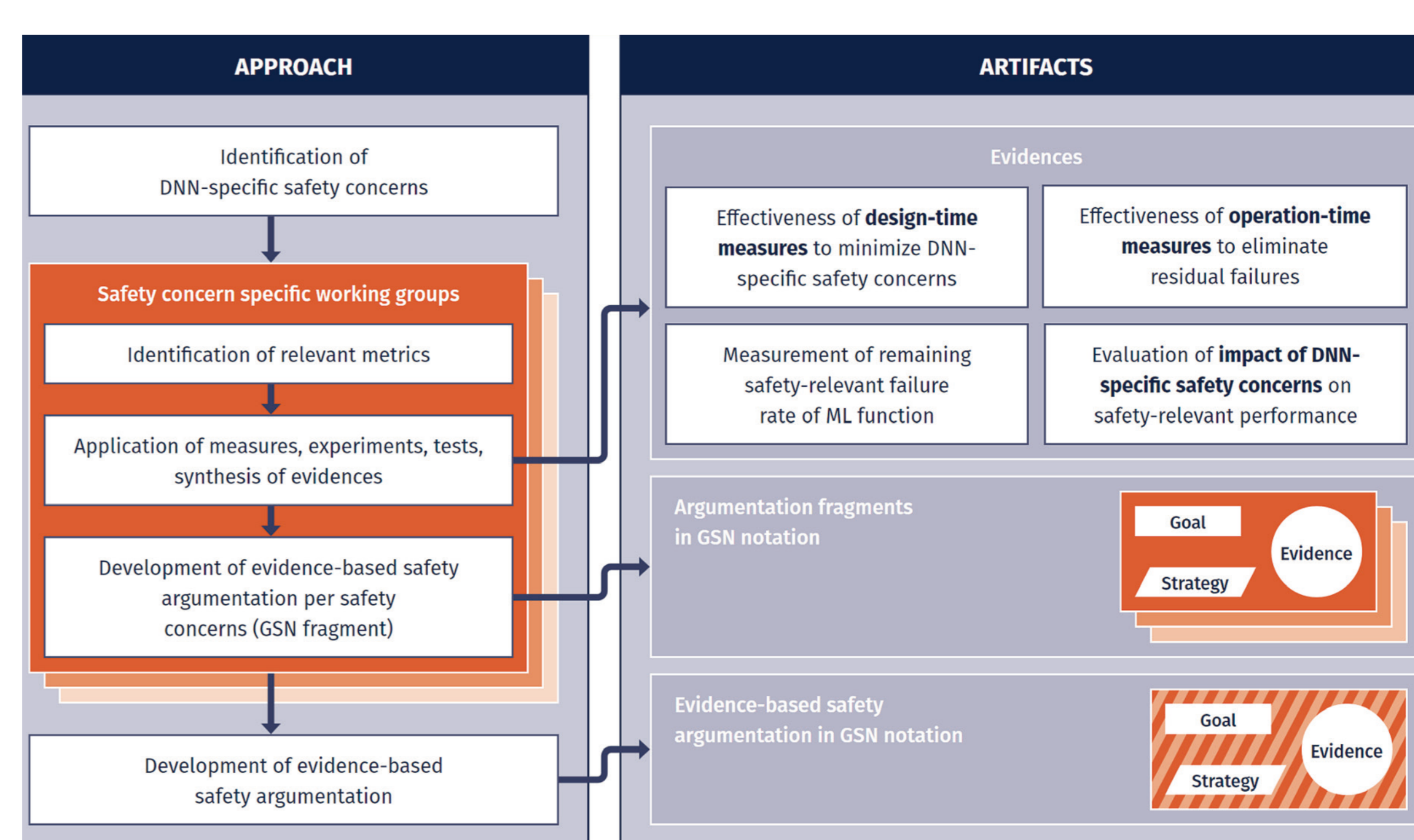


Figure 1: AI-based pedestrian detection



Safety Hypothesis:

Assurance of AI-based functions requires a methodology that can identify and systematically mitigate inherent weaknesses in AI functions. The methodology also includes a systematic approach for deriving a stringent evidence-based safety argumentation.



Assurance methodology for in-vehicle AI

The specification of the AI function is the elementary starting point, both for the construction of the safety argumentation and for the development of the function itself. In addition to the purely functional requirements, such as „recognition of persons on camera images“, the specification also includes the definition of the scope-of-use of the function, the so-called Operational Design Domain (ODD). When using Deep Neural Networks (DNN), the ODD specification also results in requirements for a systematic and representative selection of training and test data. The KI Absicherung project developed description languages and an ontology for the detailed specification of data and metadata. These are understandable for humans in order to be able to build up a comprehensible safety argumentation, as well as machine-readable in order to be able to carry out data analyses and test evaluations automatically.

Evidence-based safety argumentation

DNNs can be understood as complex black box approximation functions that are optimized by training data. As such, they may have insufficiencies in the generalization capability, which in the unfavorable case can lead to a

weakness of the software function. In order to systematically address these weaknesses, which can in particular be the cause of functional insufficiencies, a list of „DNN-specific safety concerns“ has been developed by AI and safety experts in KI Absicherung. These must be taken into account during the safety analysis and appropriate mitigation measures must be identified. By combining evidences in the safety argument it can be demonstrated that the DNN-specific safety concerns are adequately addressed.

References:

- Blank, F.; Hüger, F.; Mock, M.; Stauner, T.: Assurance Methodology for In-vehicle AI in Automobiltechnische Zeitschrift, #8, 2022.
- Mock, M.; Scholz, S.; Blank, F.; Hüger, F., Rohatschek, A.; Schwarz, L.; Stauner, T.: An Integrated Approach to a Safety Argumentation for AI-Based Perception Functions in Automated Driving. SAFECOMP Workshops, Springer, 2021
- <https://ki-absicherung-projekt.de/safety-concerns>
- <https://www.ki-absicherung-projekt.de>