

Problem Statement

Not all images are equally suitable for developing safe AI components. Various image properties can make the detection of pedestrians difficult. We focus on identifying performance-limiting features through information-theoretic analysis and developing effective mitigation strategies based on analytical insights.

Our Approach

In data engineering (Figure 1), we work with the other subprojects to define the part of the Safe AI methodology that handles data processing and links to the safety arguments through data-related safety properties. The focus is on the following overlapping areas:

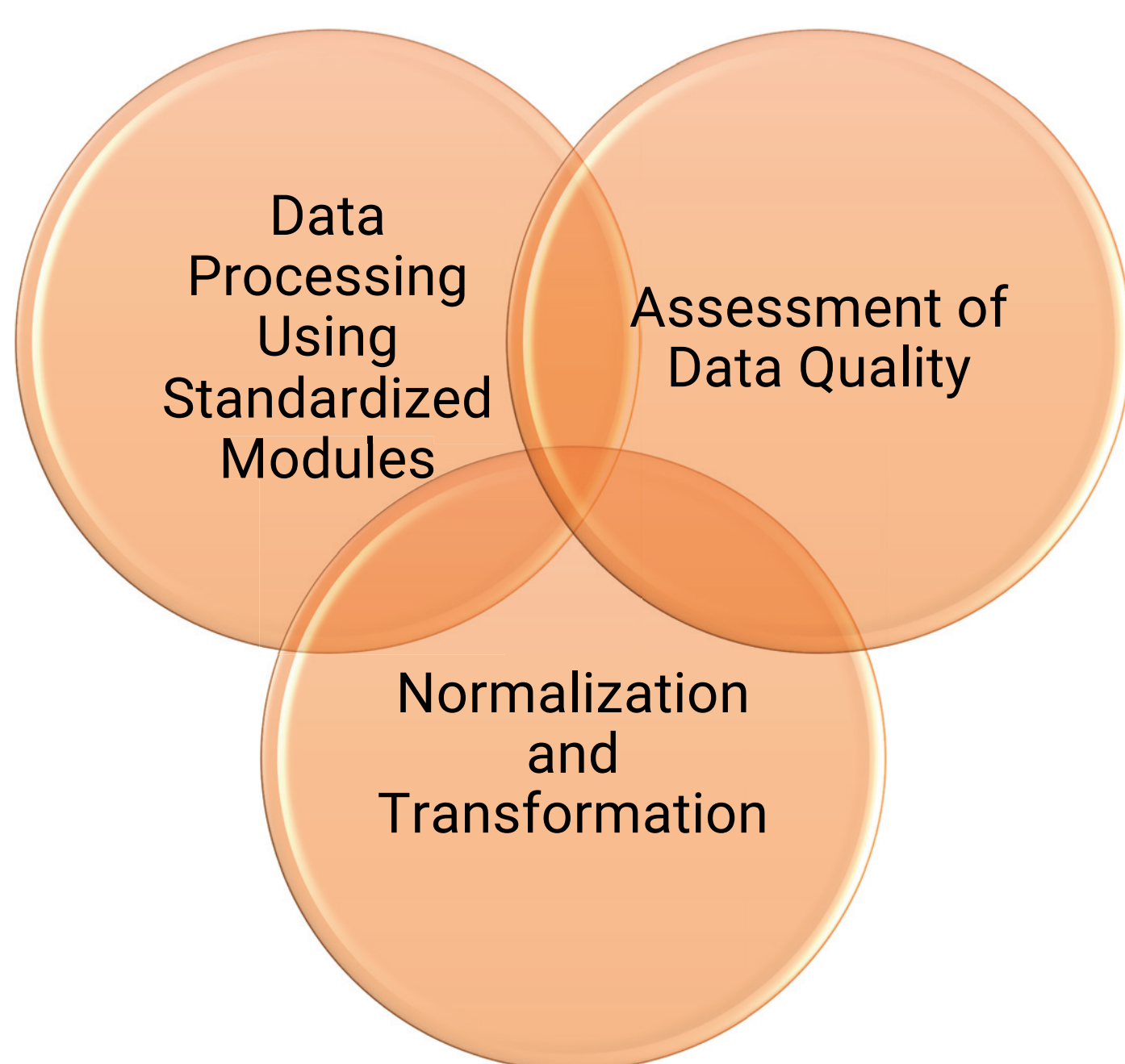


Figure 1: Data Engineering Process

Data Quality

Data quality describes the extent to which data is suitable for its intended purpose. In this project, the quality of real data is analyzed in terms of label correctness, and the realism of simulated data is also assessed.

Normalization

Normalization as a level of abstraction for incoming sensor data is considered at various levels. We use this to harmonize data from different sources into one overarching domain, which we call the top-level domain. It is defined through idealized operational parameters, including geographic boundaries, environmental conditions, and standardized sensor configurations. We consider normalization for

- Color spectrum and contrast
- Intrinsic camera parameters
- Extrinsic camera parameters

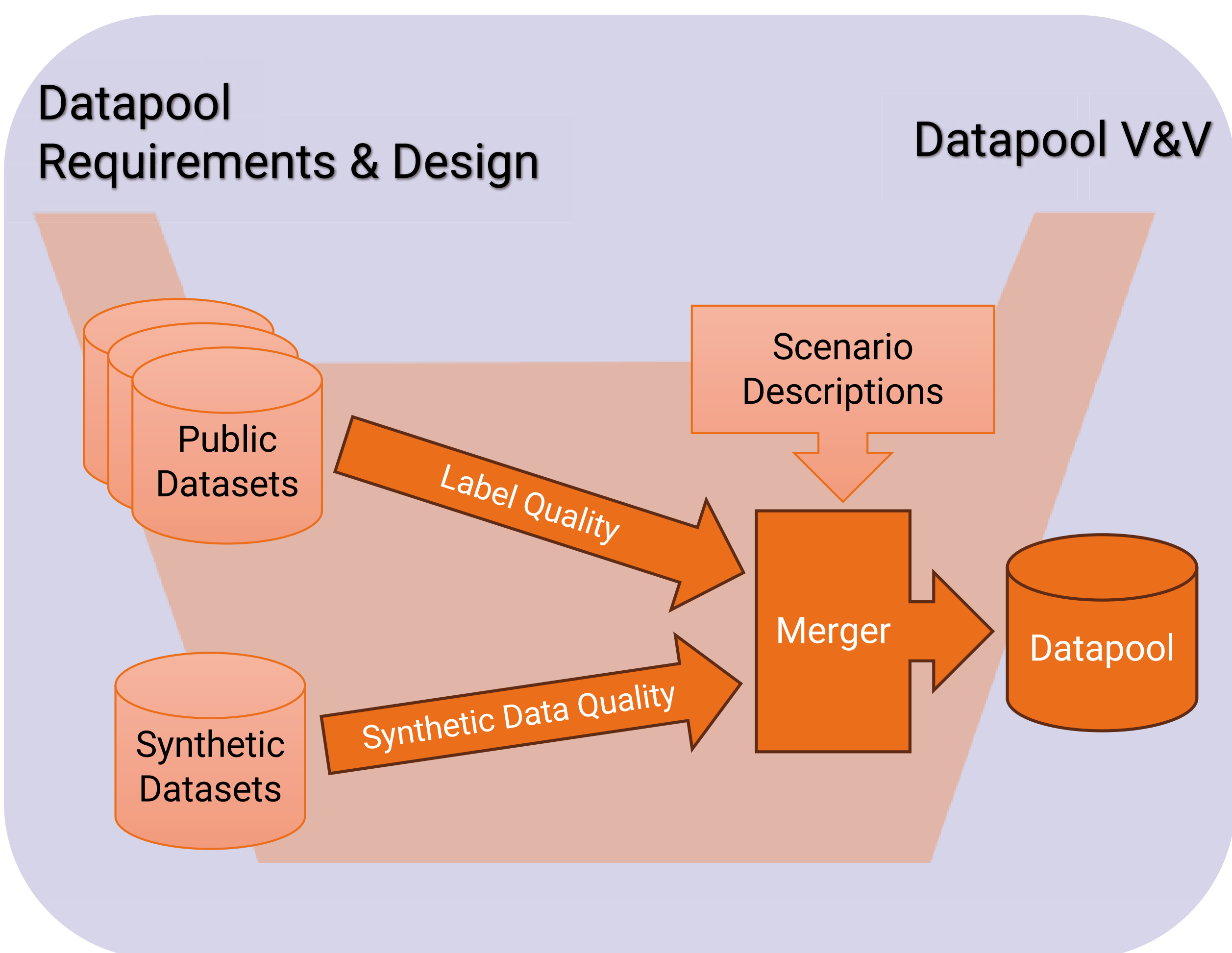


Figure 2: Data Engineering Process, our work focusses on the dark orange parts, light orange comes from others
(© DLR | Porsche | ZF | Luxoft | AVL | Bosch | BASt)

Reusable Data Engineering

Reusability analysis predicts software reuse in autonomous driving using white-box (code) and black-box (metadata) metrics. Ultimately, these insights support continuous metric monitoring, optimizing repository management, component adaptability, and long-term project efficiency beyond the project use-case.

Data Safety Analysis

The fundamental methodology of Safe AI Engineering is based on iterative development cycles. However, this means that, for a new iteration, it must be clearly defined at which point in the process the process can be resumed. To this end, the data-related analyses are combined with the model tests and monitoring results to identify which data features are responsible for performance-limiting behaviors. By identifying new triggering conditions, appropriate mitigation or improvement measures can be developed in the subsequent process.

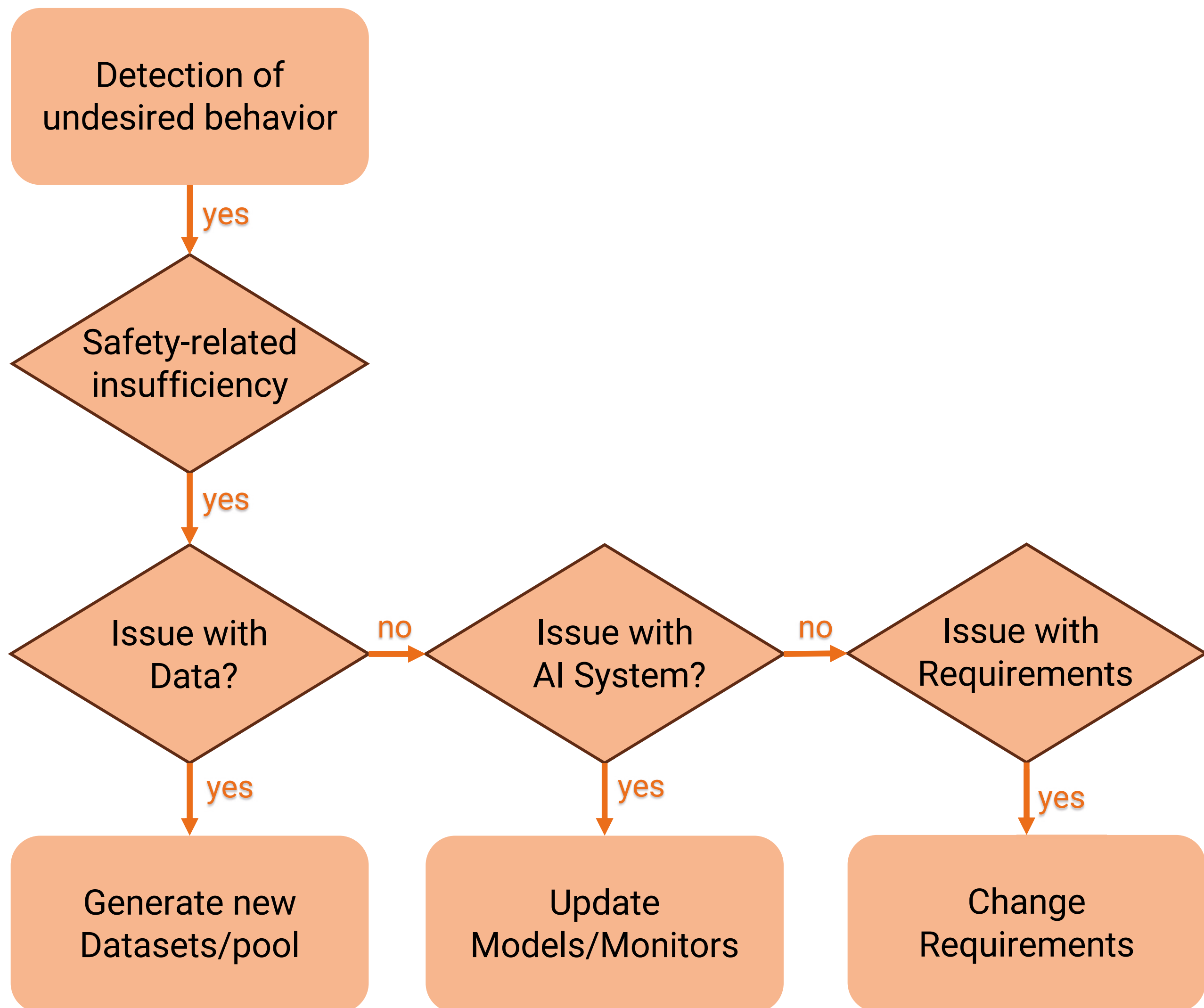
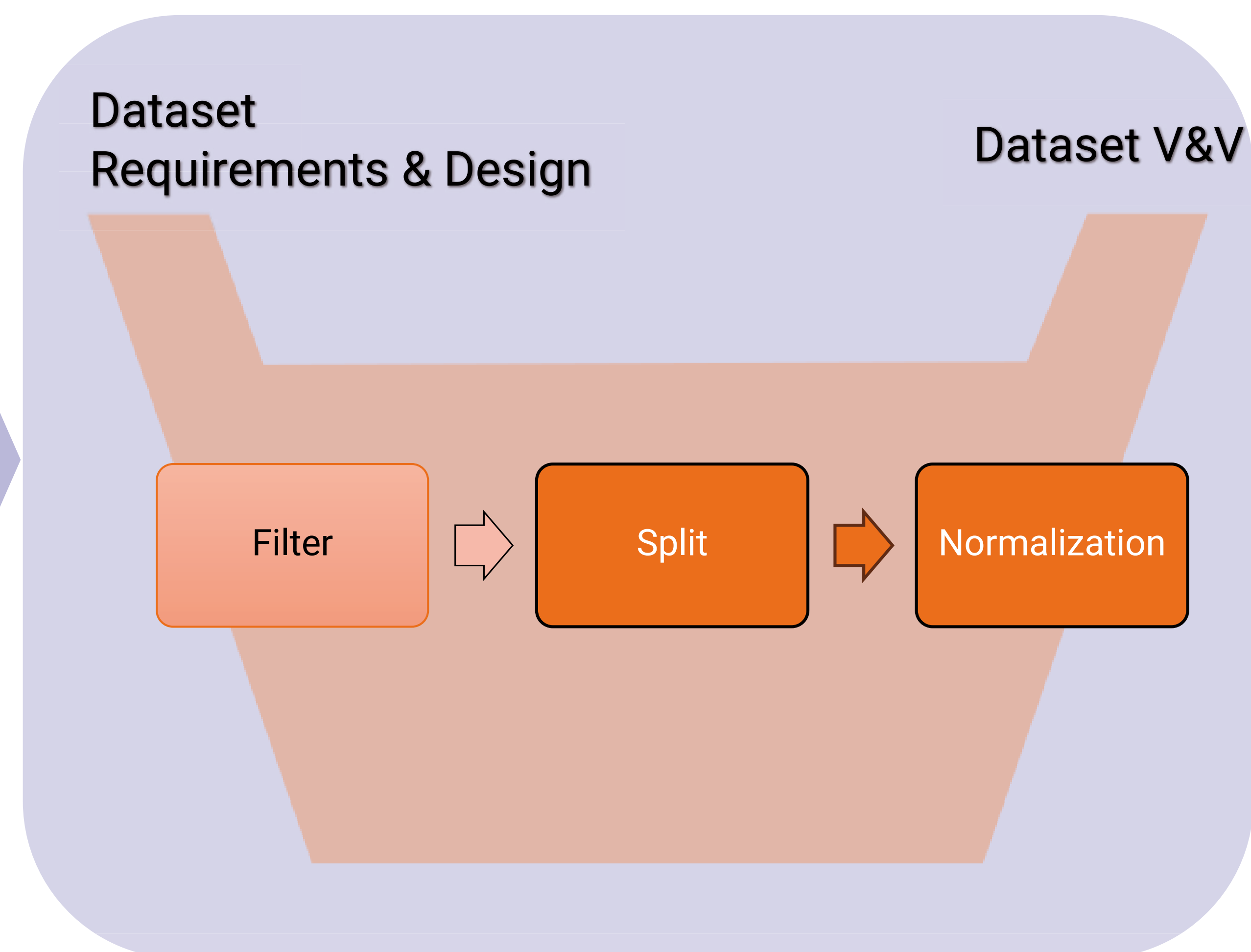


Figure 3: Data Safety Analysis (cf. ISO PAS 8800)

Publications:

- Andreas Albrecht, Istvan Bogнар, Himanshu Walia, Dinish K Karunakaran, Thomas Geipel A Data-Driven Engineering (DDE) Process for Verification and Validation of AI Systems Springer Communications in Computer and Information Science (CCIS) series, 2026
- Yannick Kees, Elena Hoemann, Frank Köster, Sven Hallerbach Image Quality Dependent Degradation for AI Systems, 2026 preprint arxiv:2607.25736v1



For more information regarding the poster, contact:
yannick.kees@dlr.de,
yannick.lingelbach@porsche.de



Safe AI Engineering is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Energy.



[linkedin.com/company/ki-familie](https://www.linkedin.com/company/ki-familie)
www.safe-ai-engineering.de