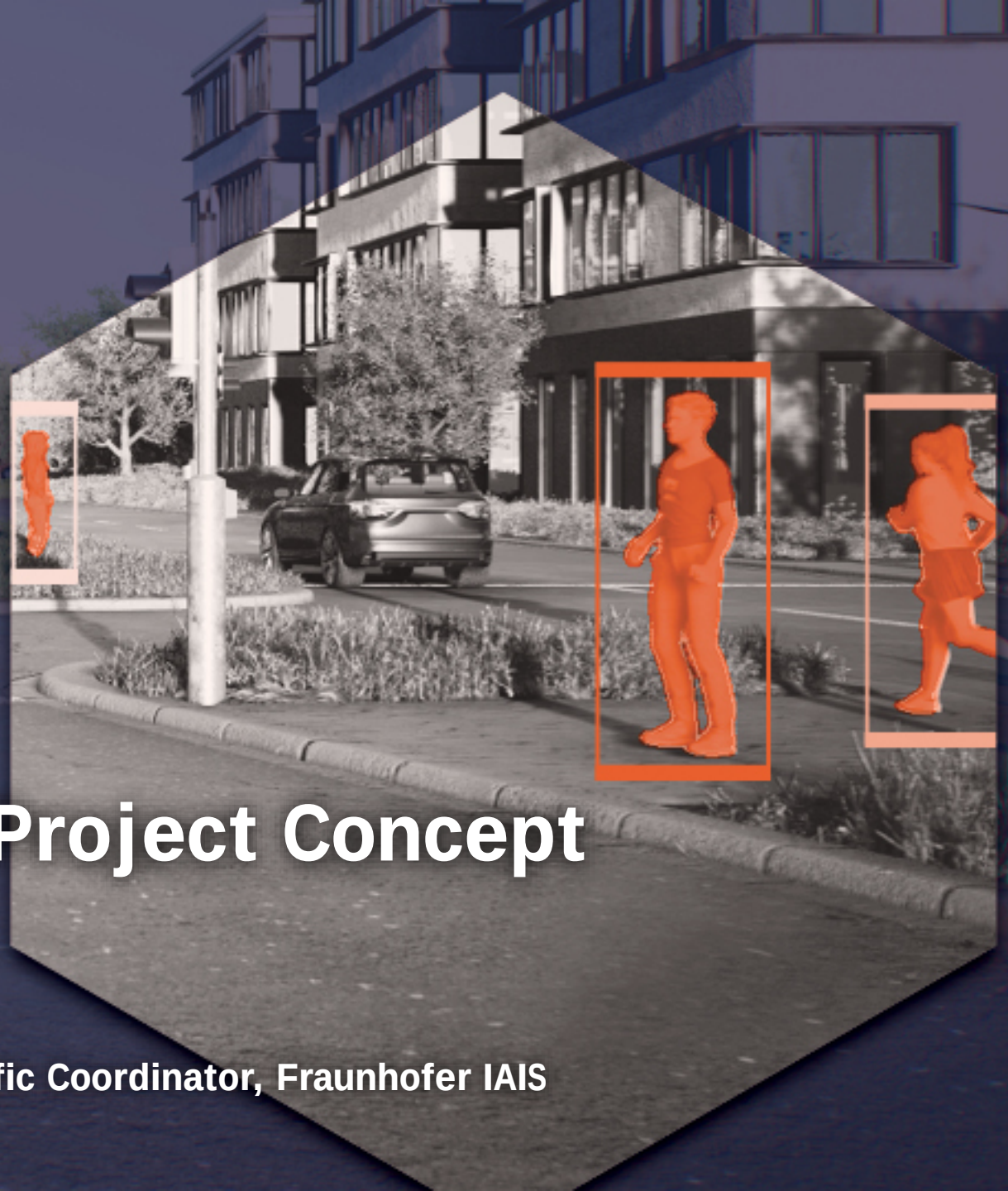
KI ABSICHERUNG

Safe AI for Automated Driving



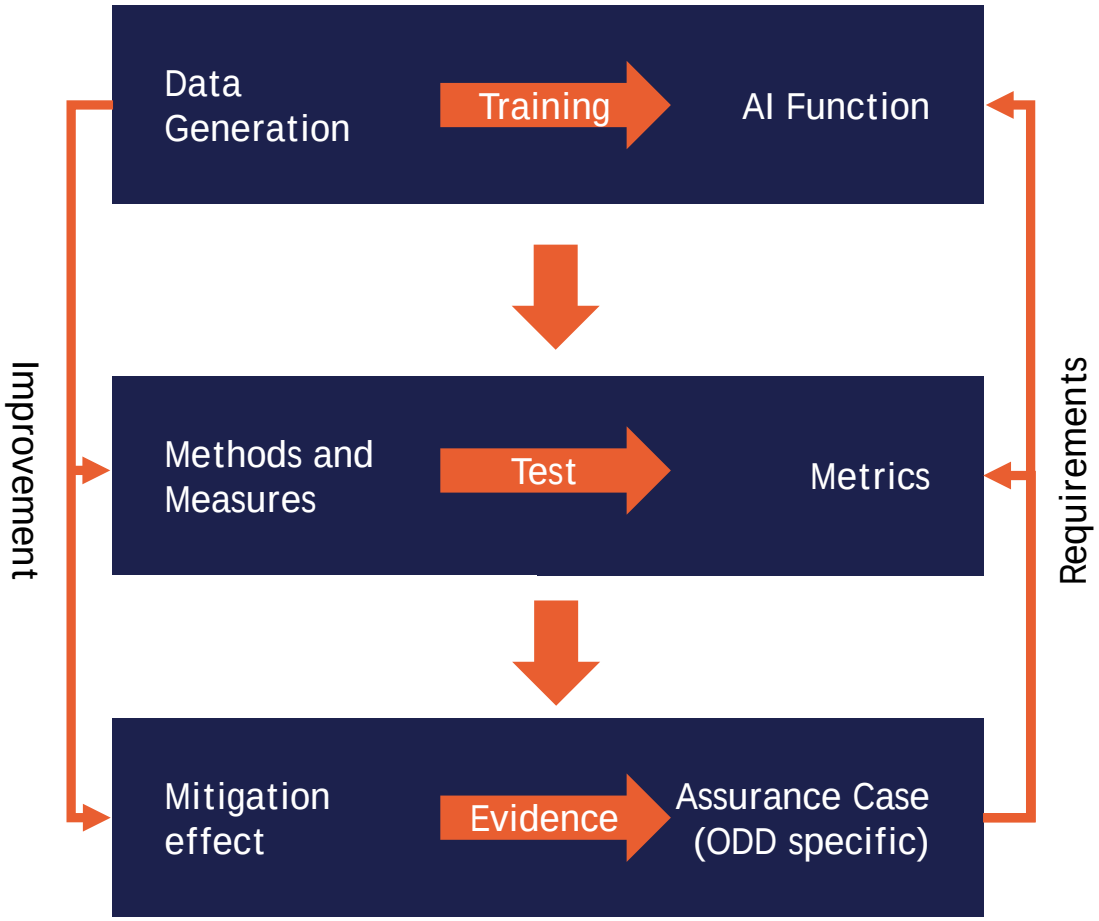
11.03.2021, Online, Interim Presentation

KI Absicherung: Proof of Project Concept

PD Dr Michael Mock, Consortium Co-Lead and Scientific Coordinator, Fraunhofer IAIS

From a data-driven AI function to an Assurance Case for the use case Pedestrian Detection

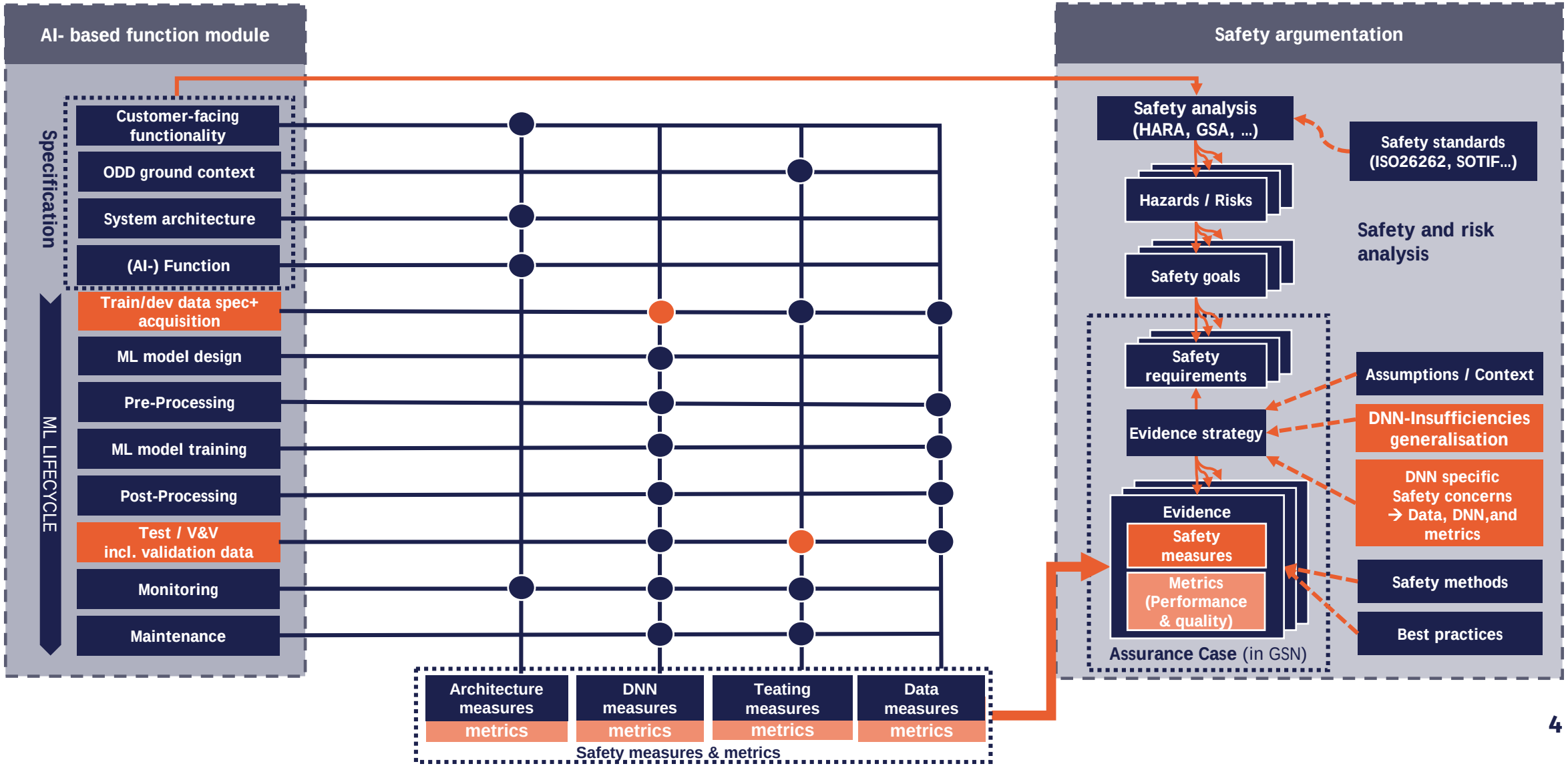
- Process-related generation of synthetic learning, testing and validation data.
- Development of measures and methods that improve the AI function over a wide array of metrics.
- Development and validation of testing methods for these metrics.
- Stringent safety argumentation for the AI function and its Operational Design Domain (ODD).



Proof of Project Concept (PoPC)

- **Define and implement the detailed technical workflow for developing a stringent safety-argumentation for AI-based functions in a minimalistic example**
- **Goals of the Proof of Project Concept**
  - Develop **an exemplary Mini-Safety argumentation** (represented as GSN)
  - Define consistent **terminology and workflow**
  - Document and implement as a **Blueprint** for the complete project
  - Cover all required project activities
    - beginning with safety requirement as starting point, also defining and generating data, and going through DNN insufficiencies, mitigating them by Methods and Measures, and measuring the success by metrics which lead to providing evidences
- **Definition in core team**
- **Implementation in PM and operational team**

Project Approach to Safety Argumentation for AI-based Functions (Big Picture)



PoPC Minimalistic Example

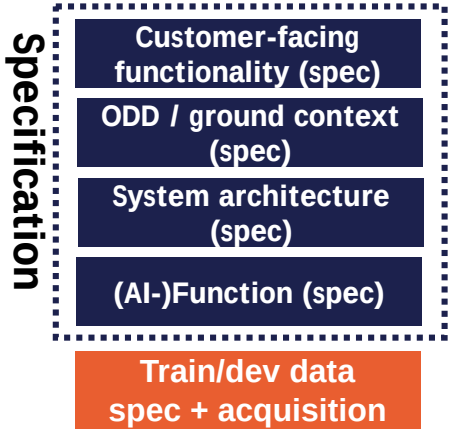
- Synthetic Data generated for one ground context (simple crossing, few assets and pedestrians)
- DNN insufficiency: **Insufficient generalisation capability**
- Measure: **Variational Autoencoder**
- Safety Requirement: **Inadmissible application of the AI-Function outside of the specified Operation Design Domain should be avoided**
- **Two-dimensional** ODD Definition: ODD is defined over one variation domain only
- Safety Argumenation: only **Mini-GSN** providing initial evidences

IN ODD (Dark)

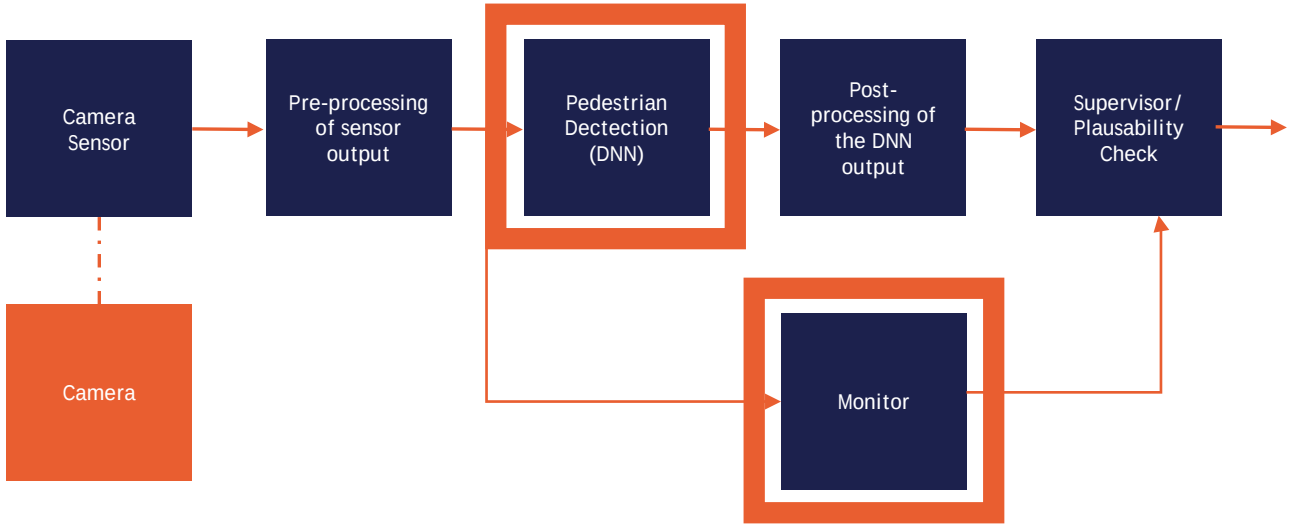
OUT ODD (Bright)

Images: KI-Absicherung, Tranche 3, BIT Technology Solutions

Specifications (Functions, Architecture)

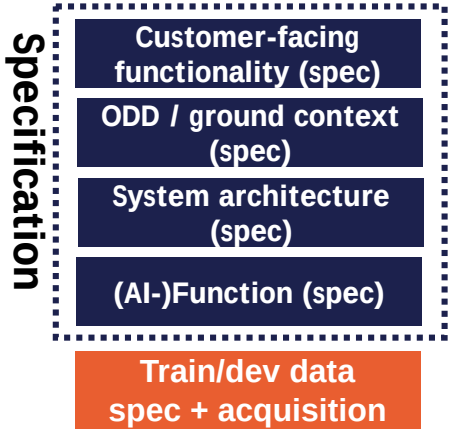


- **Customer Facing Functionality**
  - Automated driving function, that once activated by the driver ahead of an urban intersection in "sample village", takes control of the vehicle and drives it safely and smoothly through a specific type of intersection.
- **Architecture (Perception)**

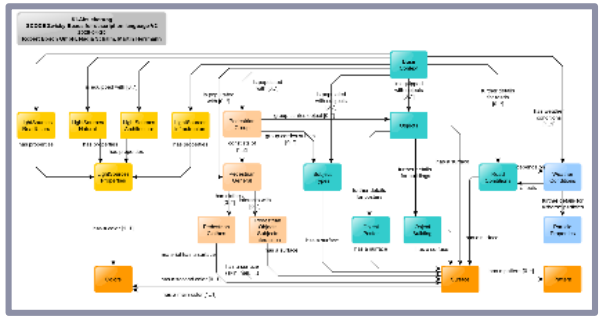


- **AI Function**
  - Among others, this function contains an AI function "**Pedestrian Detection**" whose task is to detect all visible vulnerable road users …

Specification of the Operational Design Domain (ODD)



- The ODD is specified by an **ontology** that describes the relationships of objects that can vary in a defined ground context.
- Derived **Zwicky Boxes** define dimensions and **value** ranges, e.g. for Natural Light Sources
- **Metadata specifications** serve as basis for data generation and testing



Source: Bosch

Images: P. Bhank, Bosch

PoPC two-dimensional ODD Definition

Dimension		IN of ODD
sun elevation	medium (6°- 20°) | Day (>20°)	twilight/sunrise (-6 to 0°) | low (0 - 6°)
sky	clear | high clouds	low partly clouded | low | completely clouded

Safety and Risk Analysis

Safety Goal

No pedestrian is harmed or injured by the vehicle during the automated ride through the intersection in "sample village" unless the accident is physically unavoidable.

Decomposition (simplified)

The derived safety requirement for perception is that within the ODD every relevant pedestrian is detected in all cases. Simplified: A pedestrian is relevant if she is visible/not occluded to more than X% and not more than Y meter away from the vehicle.

Derived safety requirements (PoPC only)

1. The bounding box detection works correctly in the ODD - this requirement is not considered further in the following
2. Input data out of the ODD is detected at runtime

Image: Bosch

Training of AI Functions

ML LIFECYCLE

- Train/dev data spec + acquisition
- ML model design
- Preprocessing
- ML model training
- Postprocessing
- Test / V&V (incl. validation data)
- Monitoring
- Maintenance

Synthetic Data

- Full meta-data annotations
- PoPC Data Split based on project data

AI Function SSD

- Single Shot Detector
- Multiple Bounding Boxes
- Fast Inference

AI Function Deeplab V3+

- Semantic Segmentation
- Advanced prediction performance

Images: BIT Technology Solutions, Opel, Intel

DNN safety measures and metrics

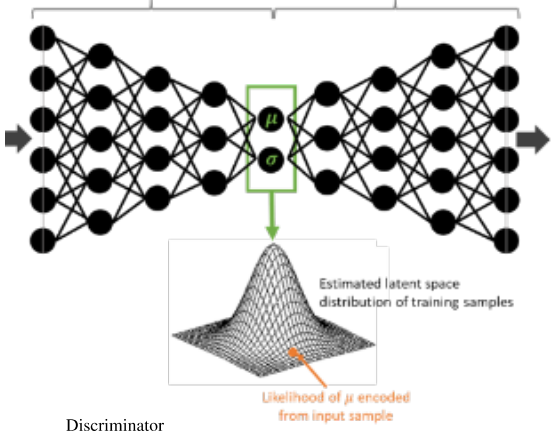
- **Autoencoders are trained to „reconstruct“ their input**
  - **„Reconstruction error“** measures the
  - distance beetwen input and output
  - High reconstruction errors imply that the input is „far away“ from the training data distribution

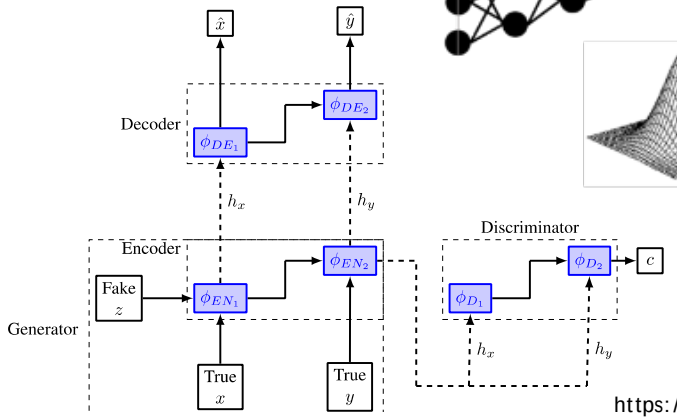
- **Generated Safety Evidence**
  - The Autoencoder is used as **an online monitor**
  - Inputs with high reconstruction errors are likely to **produce wrong outputs** in the AI function

Example for „IN ODD“ reconstruction

Images: BIT Technology Solutions, Bosch

Variational Autencoder





https://arxiv.org/pdf/1901.02514.pdf

Testing and Deriving Evidences

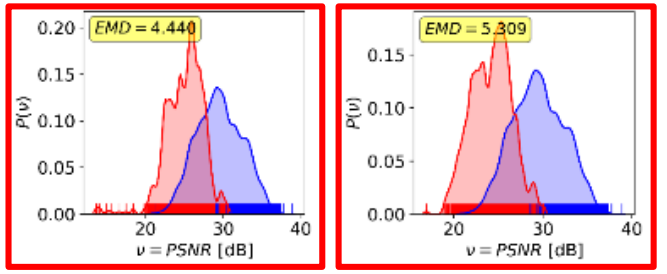
- PoPC splits defines IN ODD / OUT ODD test data
- **Reconstruction error is higher OUT ODD** (blue) than IN ODD (orange) on average
- A clear distinction on based on a single image is not possible
- **Statistical significance** of OUT ODD detection can be shown when testing sequences of frames
- A simple light based baseline test outpferforms the Autoencoder

- **Evidences can be derived**, but in the simplistic dark/bright ODD, simple algorithms work better

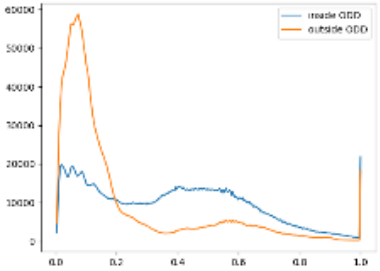
Architecture Measures	DNN Measures	Testing Measures	Data Measures
Metrics	Metrics	Metrics	Metrics

Safety Measures & Metrics

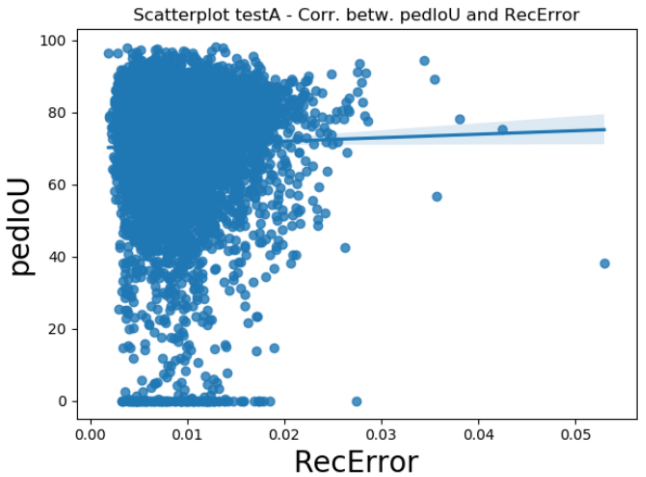
Reconstruction Error Tests



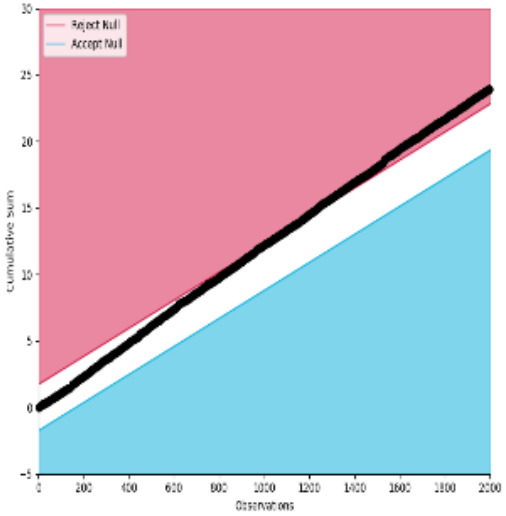
Light based Tests



Correlation Tests



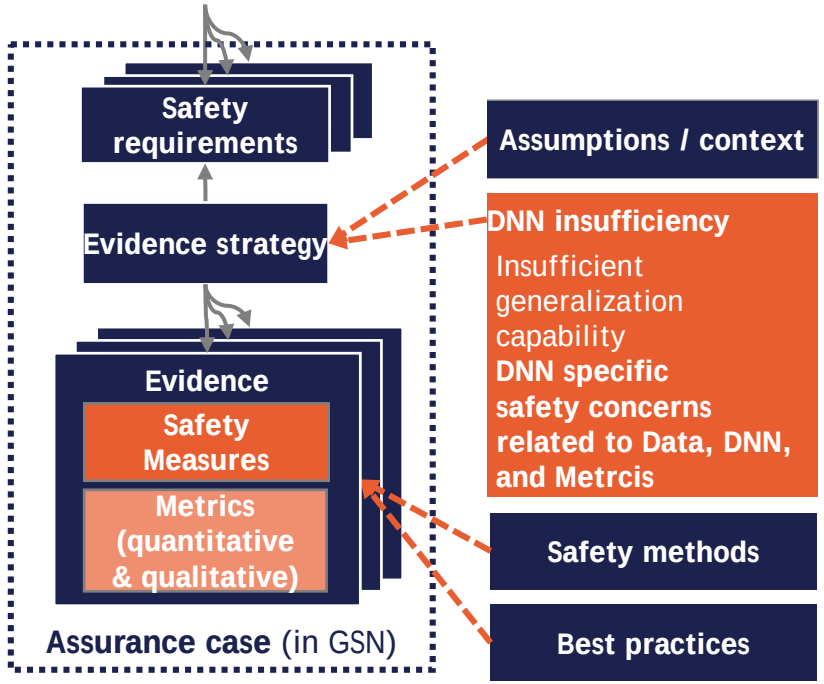
Significance Sequence Tests



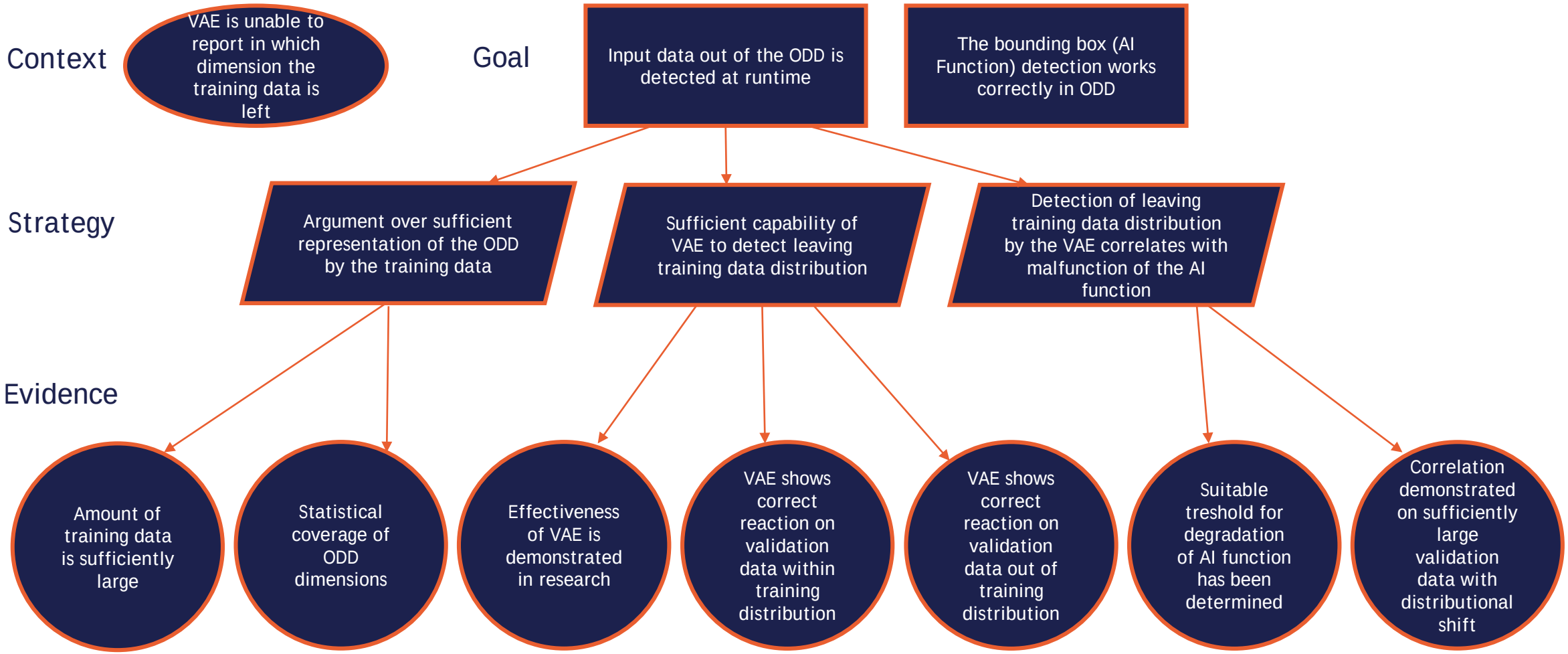
Images: TUBS, Continental, Bosch, Fraunhofer IAIS

Safety Argumetation (Method)

- An **assurance case** is used to **provide confidence that a system is safe to operate in a defined environment**
- **Assurance case strategy** is the **evidence based safety argumentation**.
- **Mitigation of DNN insuffiency**
  - **Identify**
  - **Define**
  - **Mitigate**
  - **Argue**
- Formal approach: "**Goal Structuring Notation**" (**GSN**)
- The **GSN visualizes** the **evidence based safety argumentation**



Resulting Assurance Case – Mini-GSN (simplified, PoPC only)



PoPC – Summary, Implications and Next Steps

- The PoPC has been **developed** in the core team
- The PoPC **exemplifies** the agreed overall project approach
- The PoPC has been **implemented** and used as a blue-print for the complete project
- **Workshop series on DNN specific safety concerns**
  - Harmonized and agreed for achieving project consensus
- **Series of evidence workshops**
  - Specific DNN safety mechanisms and methods have been analyzed
  - Method specific Mini-GSNs developed together with “test and safety buddies”
- The PoPC will be **extended** to cover a **multi-dimensional** ODD definition and include **multiple DNN safety mechanisms**, providing a **safety-argumentation** that takes the inherent **multi factorial nature** of DNN failures into account.

KI ABSICHERUNG

Safe AI for Automated Driving

PD Dr. Michael Mock, Fraunhofer IAIS

Consortium Co-Lead and Scientific Coordinator

michael.mock@iais.fraunhofer.de

KI Absicherung ist ein Projekt der KI Familie und wurde aus der VDA Leitinitiative autonomes und vernetztes Fahren heraus entwickelt.

KI FAMILIE

Gefördert durch:

Bundesministerium für Wirtschaft und Energie

aufgrund eines Beschlusses des Deutschen Bundestages

www.ki-absicherung.vdali.de @KI_Familie KI Familie