

30 Years

ECTRI TG Safety, October 2026

Division »Dependable Systems«

Dr. Daniel Schneider

Fraunhofer Gesellschaft

Fraunhofer-Gesellschaft

At a glance

Applied research focusing on key future-relevant technologies and the commercialization of findings in business and industry. A trailblazer and trendsetter in innovative developments.



Approx. 32,000
employees

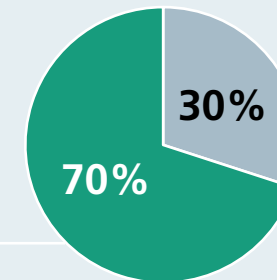


75 institutes
and research units

Business volume of €3.6 billion
Contract research totaling
€3.1 billion



Industrial contracts
and publicly funded
research projects

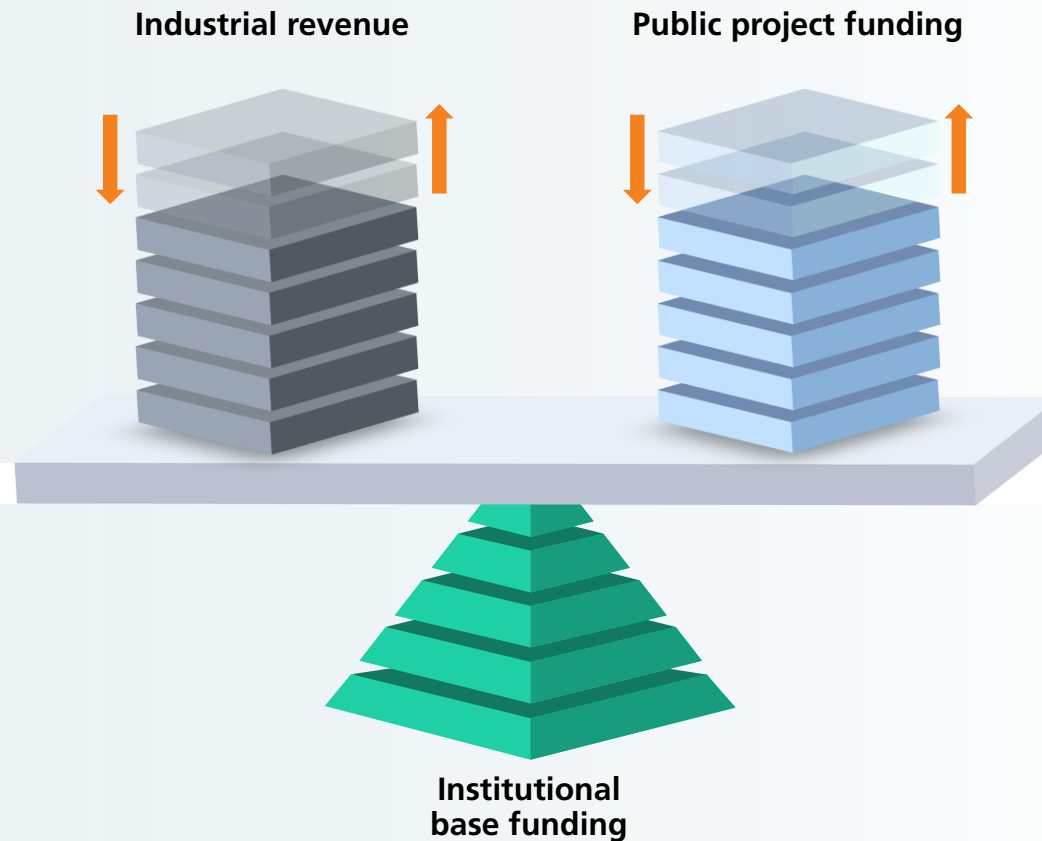


Base funding from
Germany's federal
and state governments

Fraunhofer-Gesellschaft

The Fraunhofer model involves generating at least two thirds of revenue from external sources

- Together, these revenue sources should account for at least two thirds of contract research revenue
- Target: **a balanced ratio**



- Ensure stable institutional funding (after adjusting for inflation)
- Annual increase of 3% (Pact for Research and Innovation)

Fraunhofer-Gesellschaft

The Fraunhofer-Gesellschaft's global network

8 legally independent foreign affiliates
1 opening soon

Active collaboration with partners in roughly **90** countries

5 representative offices build bridges to local markets around the world

- | | |
|-------------------------|-----------------------|
| ● Foreign affiliate | ● ICON project |
| ● Center | ● R&D (miscellaneous) |
| ● Representative office | ○ opening soon |
| ● Innovation platform | |



#WHATSNEXT

Applied research and technology transfer

Building a bridge to industry

- **Application and transfer of new technologies**
– transferring innovative concepts into practice
- **Development of innovative solutions, contract research**
in customer-individual projects
- **Licensing of methods and tools**
on the basis of innovative Fraunhofer technologies
- **Potentials analyses, studies, and expert reports**
from a neutral, independent perspective
- **Seminars, presentations, and coaching**
customized to your needs
- **Strategic collaborations on key topics**
along our competencies and focus topics



Fraunhofer IESE

Fraunhofer IESE

The Institute for Software & Systems Engineering



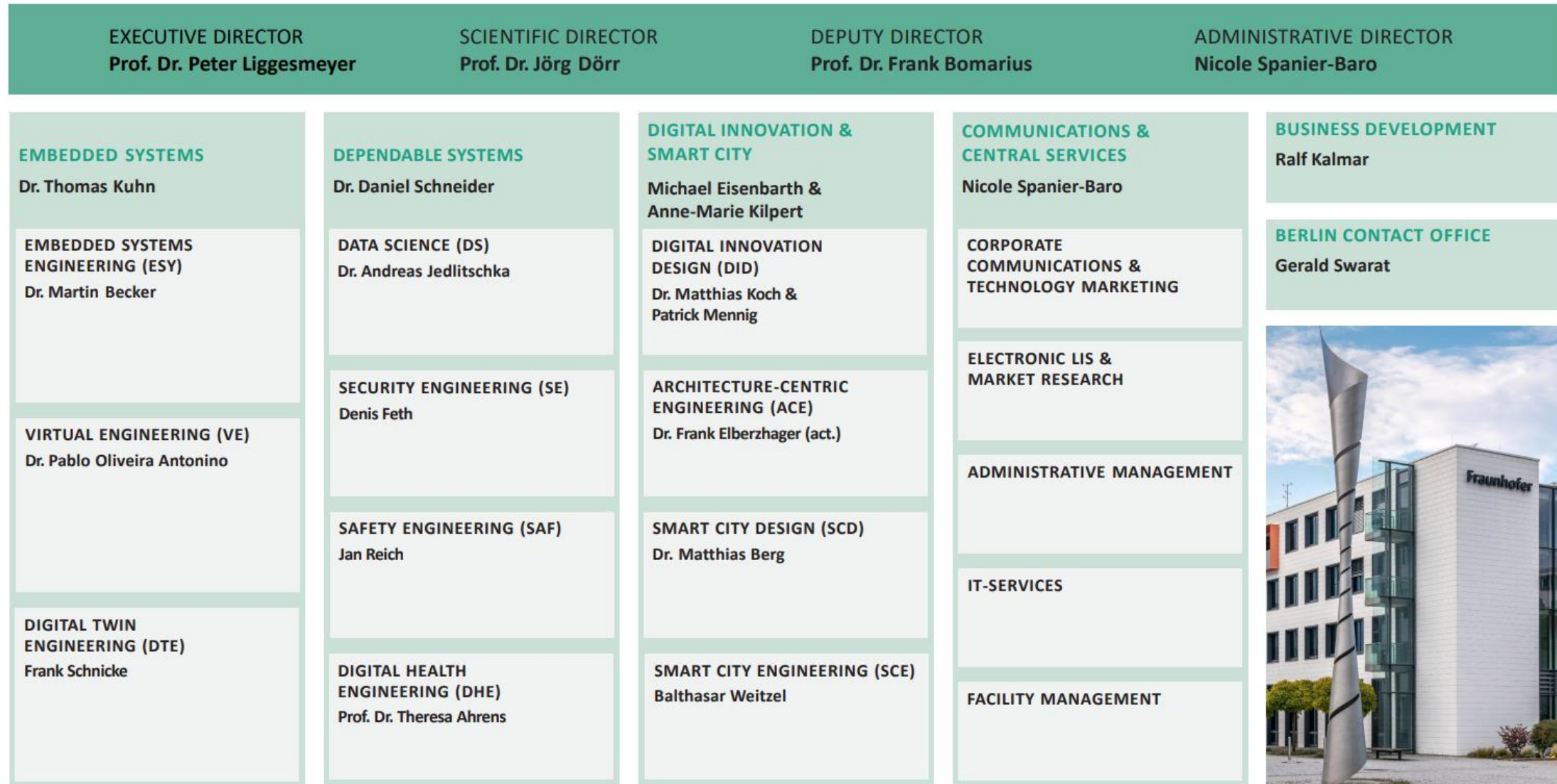
Our vision:

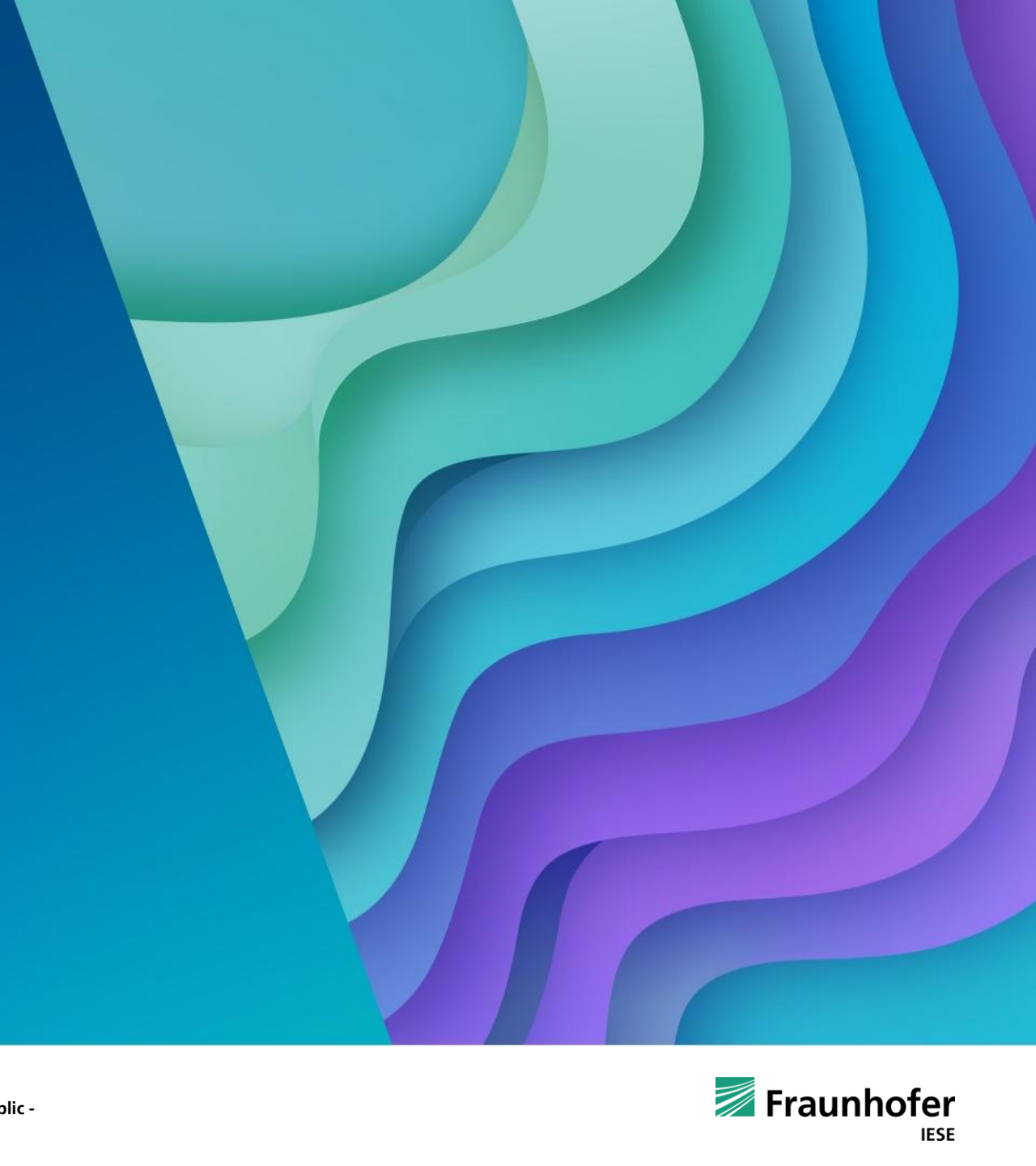
Dependable digital Systems.

Our mission:

Applied research for excellent, efficient software and systems engineering. We develop leading-edge, dependable, and resilient systems that increase the success of our partners.

FRAUNHOFER IESE ORGANIZATIONAL CHART





Division »Dependable Systems« Overview

DSYS at a Glance

Four complementary departments with a shared dependability focus

Division DSYS key features

- **DSYS combines four complementary departments:** three addressing key dimensions of dependability and Digital Health Engineering applying them in a critical domain.
- A **distinctive competence portfolio** builds on a **strong research track record** across these fields..
- **Strong research, industry and policy networks** support scientific excellence, standardization and transfer.
- **Dependability is becoming essential** for future critical systems, while major research challenges remain unresolved.
- This creates a focused and coherent division that contributes directly to the IESE mission and strategic priorities.

Data Science:

Dependable and certifiable AI;
AI for Software & Dependability Engineering.

Safety Engineering:

Model-based AI-supported Safety Engineering;
Safety assurance for connected automated systems.

Security Engineering:

Cybersecurity and data sovereignty;
AI-supported Security Engineering + AI-supported Compliance

Digital Health Engineering:

Digital Health solutions with a focus on Dependable AI and Interoperability

Collaboration within DSYS Creates Distinctive Strengths

Integrating Safety, Security, Data Science, and Digital Health

AI-powered Dependability Engineering with special focus on Safety and Security Engineering

- AI4SE is a natural topic for us as SE institute and now becoming a prime objective with the focus shift towards efficient software engineering;
- DSYS has a strong portfolio of collaborative projects, including MedSafe, TITAN, AI4SDV, MOD2SDV, and upcoming IPCEI AI projects.

Dependable AI, developing technologies for dealing with uncertainties

- Multiple departments jointly contribute to dependable AI through UncertAInty, FHIRStarter, PETRA, Safe AI Engineering, MedSafe, TITAN, IPCEI AI, Q-AI, and FCAS.
- We are heavily engaged in standardization wrt AI uncertainty (ISO/IEC TS 25223 in particular)
- The strategic objective is a shared AI Uncertainty Management toolkit applicable for both, AI4SE as well as assuring AI in autonomous systems

Safety-Security Co-Engineering

- SmS conference at IESE & series of projects with significant volume in research and industry
- Presently writing a Whitepaper with an influential standardization group (SmS, Safe AI and DE-AISI)

Cross-Divisional Collaboration Creates Greater Impact

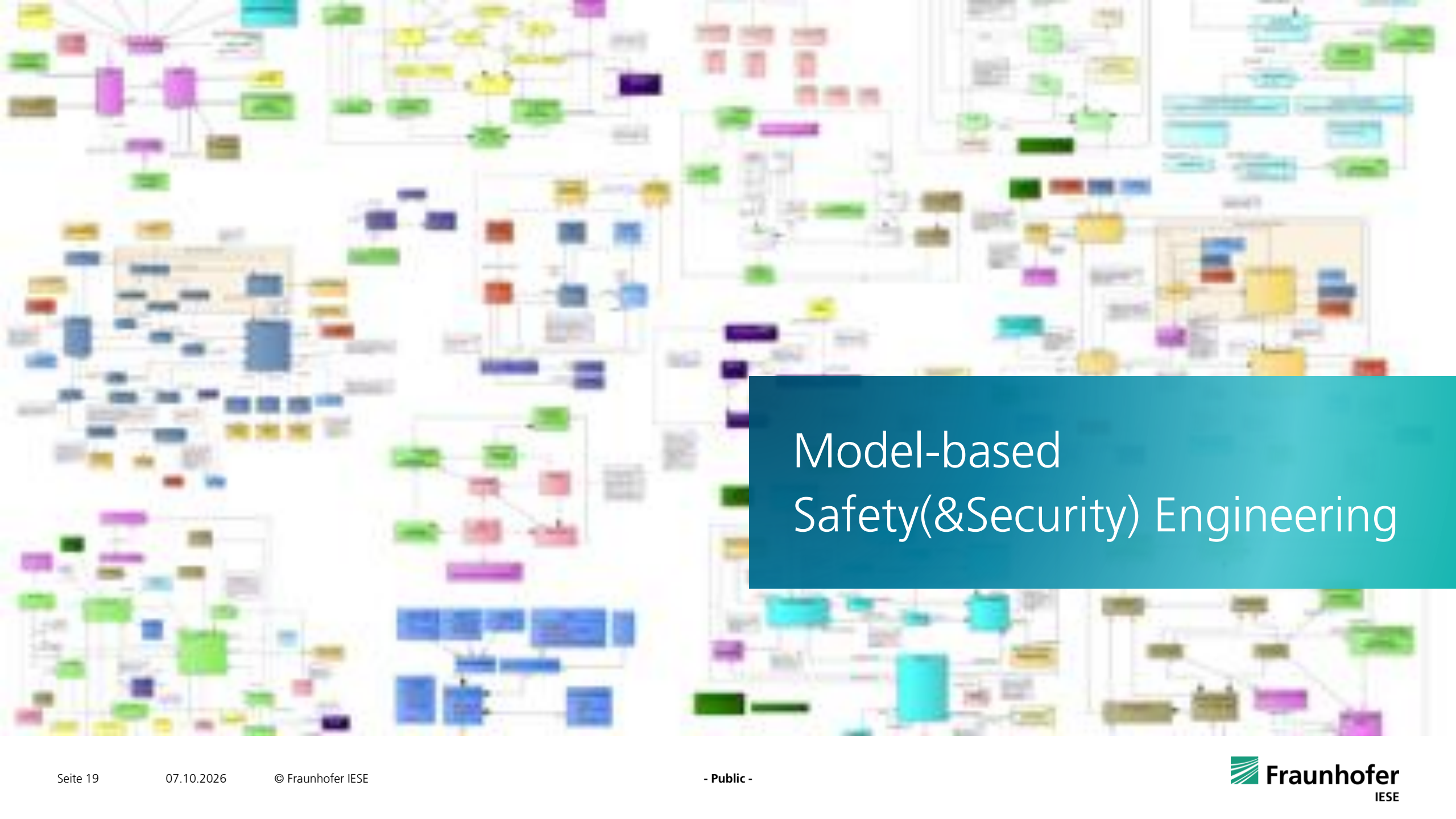
Combining complementary capabilities in strategic application domains

Autonomous Systems, Autonomics, and Robotics

- IESE has a long track record in dependable autonomous systems, complemented by several current industry projects such as current projects with Huawei, TIER IV, ARTC, Hitachi
- Joint activities across divisions combine Safety, Security, Data Science, architectures, digital twins and virtual engineering.
- Projects such as MOD2SDV demonstrate the value of combining these capabilities around shared system challenges.

Pharma 4.0

- A growing project portfolio includes ATMP, RNAuto, Vimoprop and, presently the IHI project Easygen.
- The joint focus is on higher levels of automation in pharmaceutical production, particularly quality assurance and quality management.
- Collaboration combines Safety Engineering and Digital Health expertise from DSYS with Virtual Engineering and digital-twin capabilities from other divisions.



Model-based Safety(&Security) Engineering

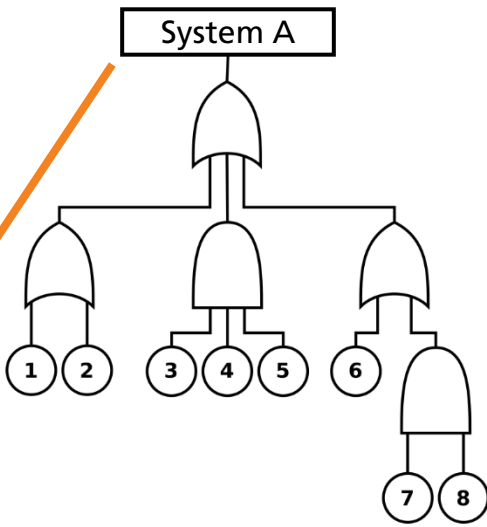
Typical Safety Engineering Tool Landscape

Hazard analysis and risk assessment (HARA)
Usually table-based



ID	Variant/Condition	Funktion	Funktions-Block-Definition	Sicherheit	Letzte Modif. of. Funktion	Informations-System-Status	Notizen
101	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
102	Start zu Stop	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
103	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
104	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
105	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
106	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
107	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
108	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
109	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
110	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
111	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
112	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
113	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
114	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
115	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
116	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
117	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
118	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
119	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion
120	Start zu Start	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion	Startfunktion

Fault Tree analysis
e.g. Isograph FaultTree+



Safety Case



Functional and Safety requirements

Safety concept (Safety measures)

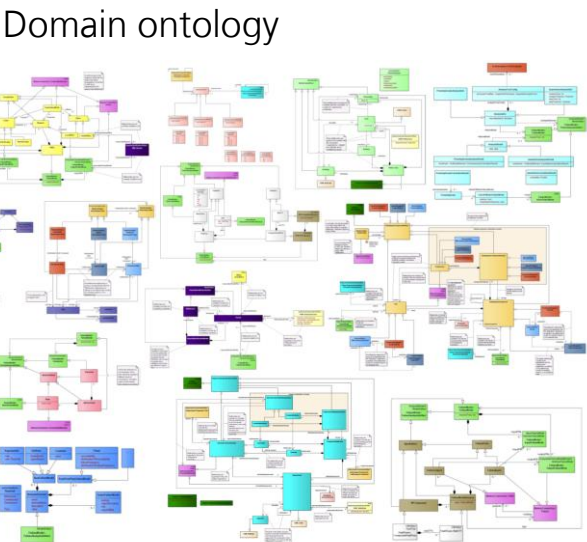
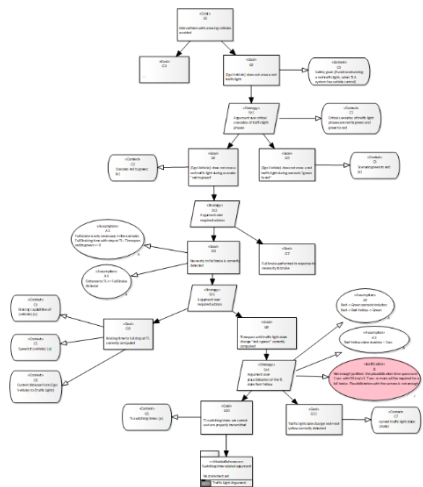
Architecture Design with UML Diagrams



Dependencies between contents of different documents

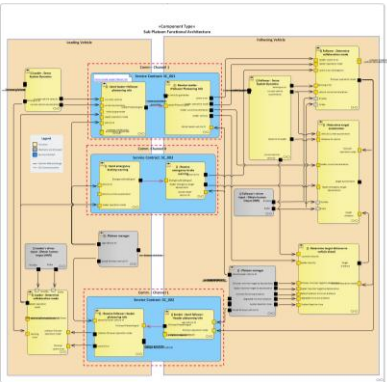
Integrated model-based Safety Engineering

Safety concept & safety case
Documentation through (GSN)

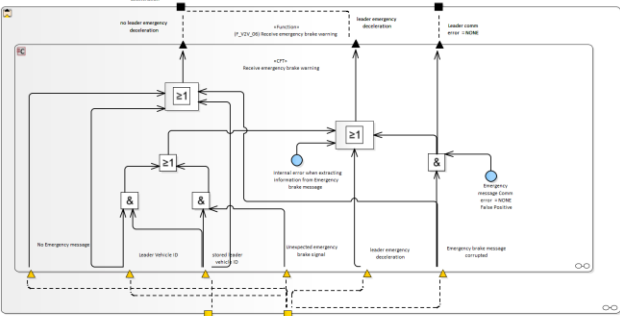


Hazard and Risk Assessment (HARA)

Functional/ logical/ technical system design



Modular component-integrated failure modeling with Component Fault Trees (CFT)



- Integration of the safety artifacts in a model repository
- Formal traceability
- Reusability of elements
- View tailoring

Safety Engineering with safeTbox

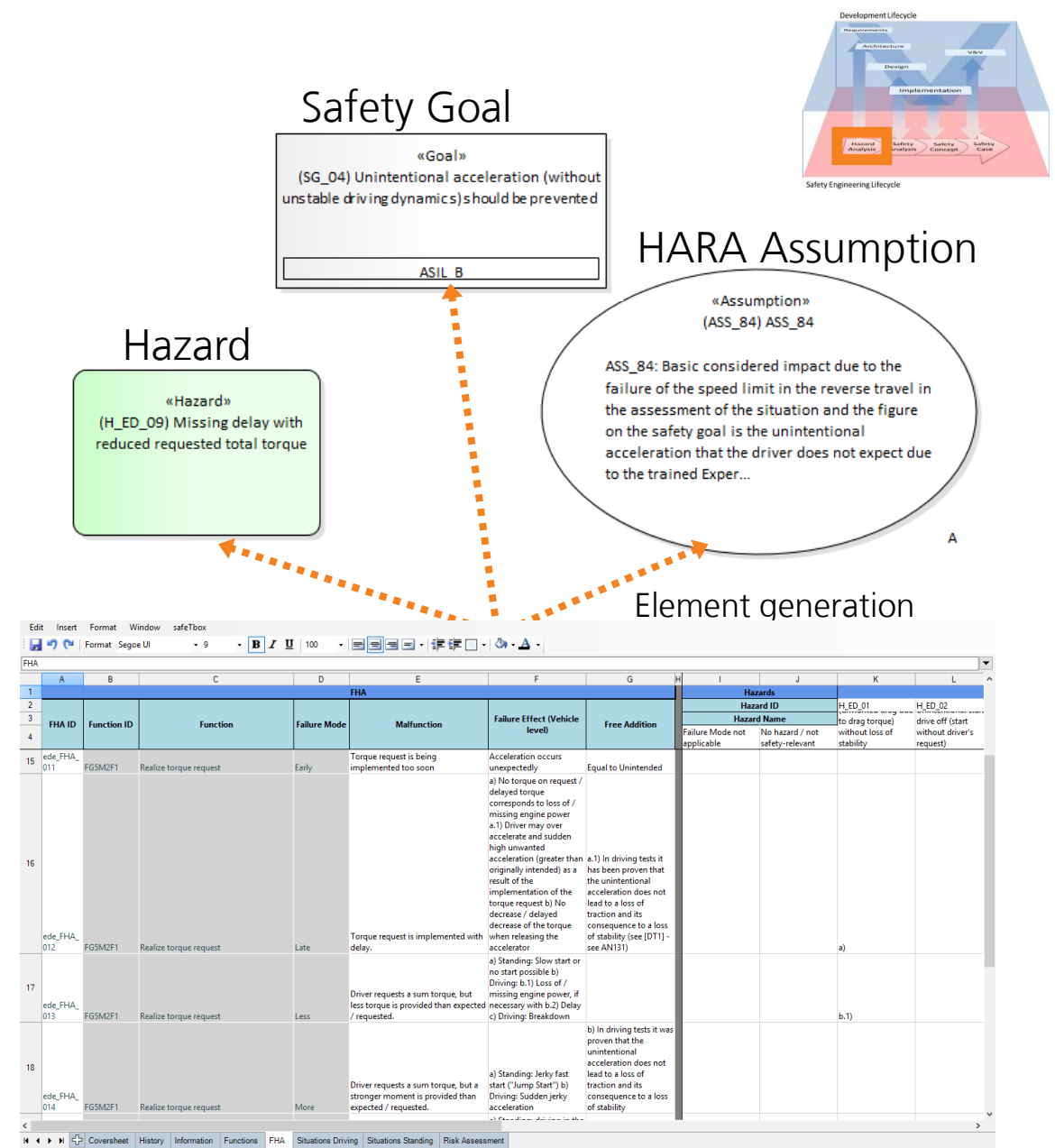
Hazard and Risk Assessment (HARA)

Supports

- Predefined HARA Template for Automotive Domain according to ISO 26262
- Functional Hazard Analysis (FHA)
- Catalog definition for HAZOP guide words and Situations (Driving and standing)
- Risk Assessment

Key Features

- Table-based UI (intuitive)
- Modeling elements can be reused in UML or safeTbox diagrams
- Validation check (for completeness)



Safety Engineering with safeTbox

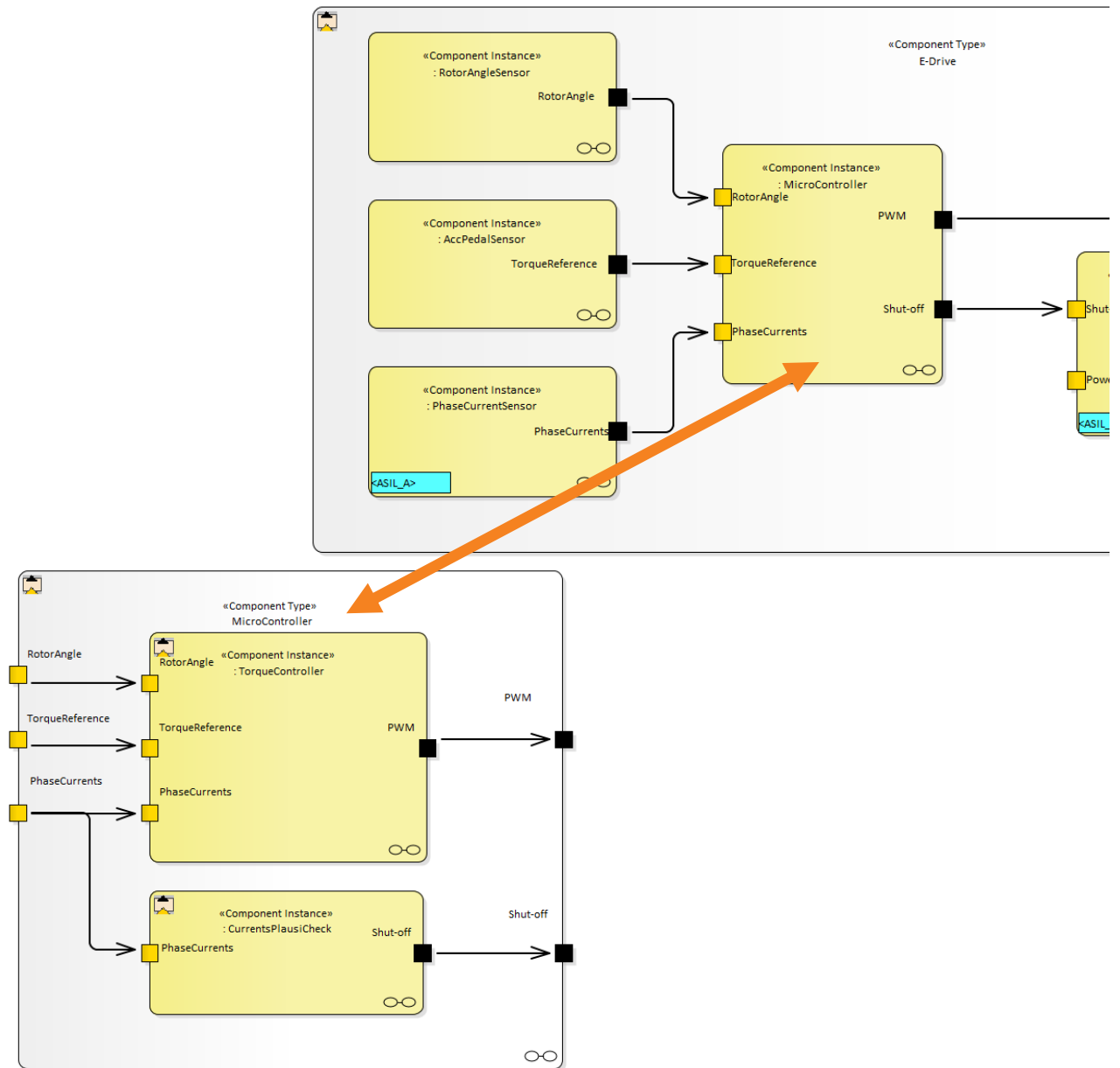
Architecture design

Supports

- Generic representation
- Functional, logical & technical architecture
- Focus in the interaction with safety engineers
- Complexity handling through model hierarchization
- Variability through realization views
- Reuse through instantiation

Key Features

- Type-instance consistency
- Formal allocation to failure models



Safety Engineering with safeTbox

Fault analysis with (CFTs), we use this A LOT

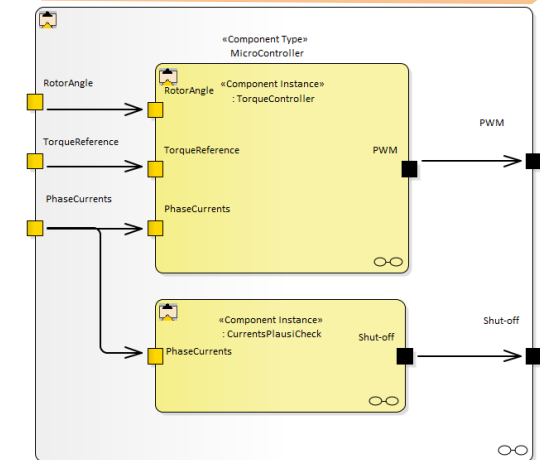
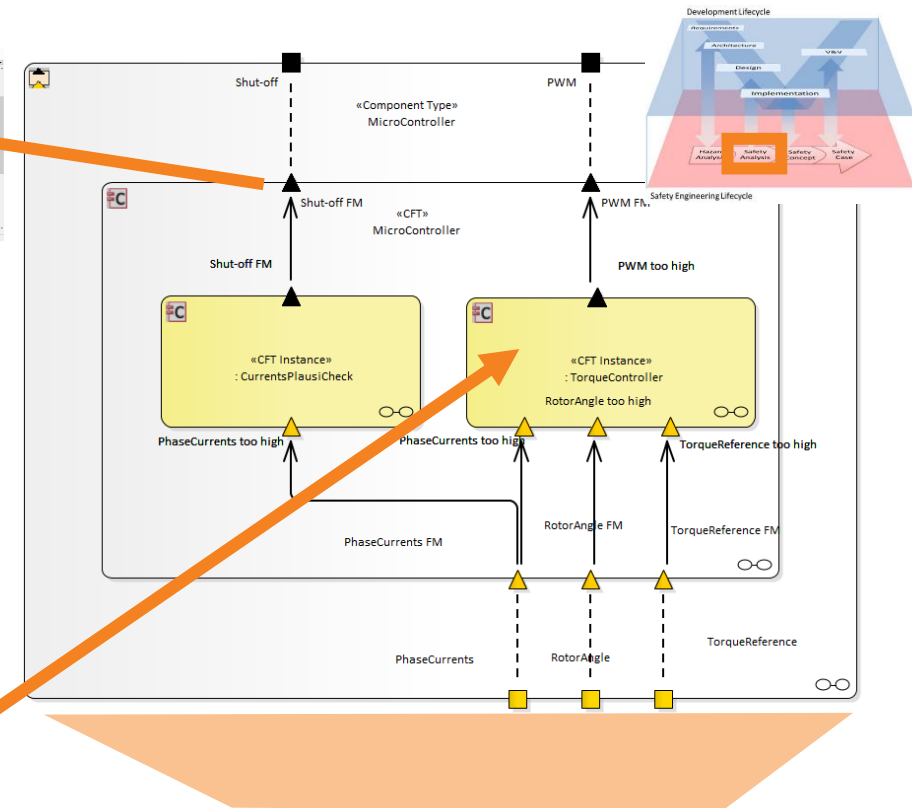
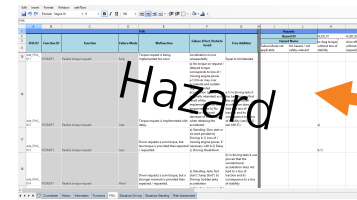
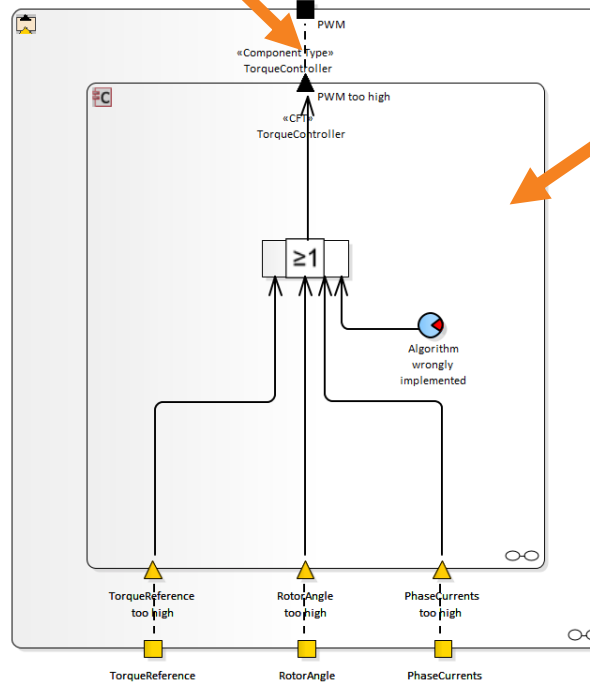
Supports

- Component Fault Trees
- Qualitative + Quantitative analysis
- Variability, reuse as in the architecture design
- SysML-Support CFT <-> SysML Block

Key Features

- Interface consistency between Architecture and Fault model
- IESE analysis engine

Mapping between functional ports and failure modes



Safety Analysis features

- Qualitative Analysis
 - Minimal Cut Sets Analysis
- Quantitative Analysis
 - Top Event probability
 - Event importance analysis
 - Common Cause Failure Analysis
- Integrated fault tree analysis with IESE's own engine or with the certified FTA tool Isograph FaultTree+
- Results as excel export/modeling element

The screenshot displays the 'CFT Analysis Result Viewer' window. At the top, the 'Element of interest' is 'Acceleration FM : Value too high'. Below it, 'Occurrence probability' is 3.15E-04 and 'Mission time' is 1 Hour. There are checkboxes for 'Show negative implicants' (checked) and 'Show Identifier' (unchecked). The main section is titled 'Prime implicants' and contains a table with 4 rows. The second row is highlighted in blue. Below this is a section for 'Individual implicant events' with a table of 2 rows. The final section is 'Event importance' with a table of 2 rows. At the bottom, there are buttons for 'Create project backup', 'Create result artifact(s)', 'Export to Excel', and 'Close'.

Element of interest: Acceleration FM : Value too high
Occurrence probability: 3.15E-04
Mission time: 1 Hour
☒ Show negative implicants
☐ Show Identifier

Prime implicants

#	Order	Probability	Importance	Prime implicant
1	1	4.63E-05	0.09639	Internal Failure
2	2	1.482E-04	0.3085	Algorithm wrongly implemented, Emergency Shut-Off Omission
3	2	1.482E-04	0.3085	Emergency Shut-Off Omission, RotorAngle too high
4	2	2.048E-06	4.264E-03	Emergency Shut-Off Omission, InternalSensorFailure (CCF Event)

Individual implicant events

#	ID	Positive	Name	Parent path	Failure Distribution
1		True	Algorithm wrongly implemented	E-Drive :: MicroController :: TorqueController	Fixed (Occurrence Probability : 0.463)
2		True	Emergency Shut-Off Omission	E-Drive :: Emergency Shut-off	Fixed (Occurrence Probability : 0.00032)

Event importance

#	Event name	Fussell-Vesely	Birnbaum	Parent path
1	Emergency Shut-Off Omission	0.8925	1.34	E-Drive :: Emergency Shut-off
2	Algorithm wrongly implemented	0.3124	3.241E-04	E-Drive :: MicroController :: TorqueController

☐ Create project backup

Safety Engineering with safeTbox

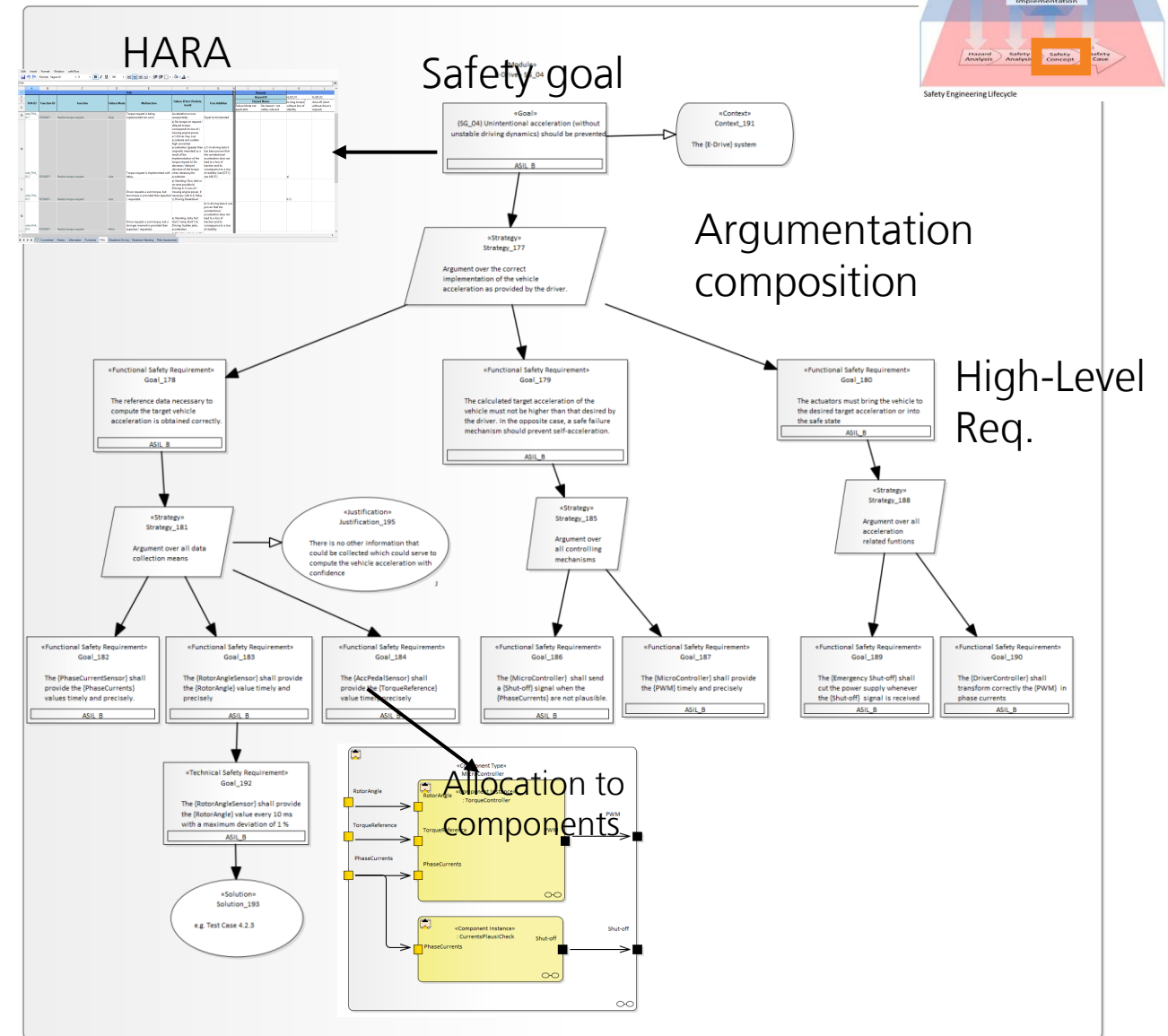
Safety Concept / Case

Supports

- Safety Concept and Safety Case documentation based on the Goal Structuring Notation (GSN) [1]
- Integrity level allocation

Key Feature: Formal Traceability to existing modeling artefacts; e.g.

- HARA
- Architecture
- Fault model
- ConSert



[1] <http://www.goalstructuringnotation.info>

Curious?

Free safeTbox download at <https://www.safetbox.de>

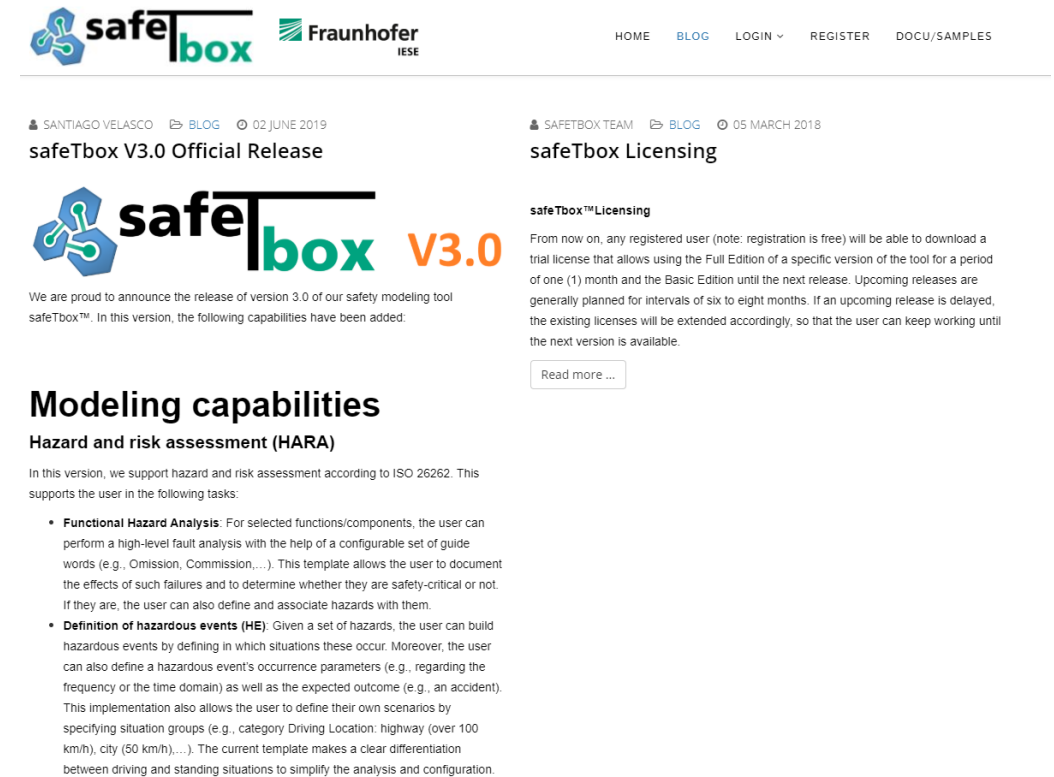
Visit our Blog

Please contact us under

safeTbox@iese.fraunhofer.de

Services

- Integration of safeTbox into your toolchain
- Trainings for safeTbox
- Customization
- Methodical safety engineering support



The screenshot displays the safeTbox website interface. At the top, the safeTbox logo and the Fraunhofer IES logo are visible, along with navigation links for HOME, BLOG, LOGIN, REGISTER, and DOCU/SAMPLES. The main content area features two blog posts. The first post, titled 'safeTbox V3.0 Official Release' by SANTIAGO VELASCO dated 02 JUNE 2019, includes the safeTbox V3.0 logo and a paragraph announcing the release of version 3.0, highlighting new capabilities. The second post, titled 'safeTbox Licensing' by SAFETBOX TEAM dated 05 MARCH 2018, includes the safeTbox™ Licensing logo and a paragraph explaining the trial license and upcoming releases. A 'Read more ...' button is present at the bottom of the licensing post. The third post, titled 'Modeling capabilities' by SAFETBOX TEAM dated 05 MARCH 2018, includes the Modeling capabilities logo and a paragraph explaining the hazard and risk assessment (HARA) capabilities, followed by a bulleted list of features: Functional Hazard Analysis and Definition of hazardous events (HE).

safeTbox V3.0 Official Release
SANTIAGO VELASCO | BLOG | 02 JUNE 2019

We are proud to announce the release of version 3.0 of our safety modeling tool safeTbox™. In this version, the following capabilities have been added:

safeTbox™ Licensing
SAFETBOX TEAM | BLOG | 05 MARCH 2018

From now on, any registered user (note: registration is free) will be able to download a trial license that allows using the Full Edition of a specific version of the tool for a period of one (1) month and the Basic Edition until the next release. Upcoming releases are generally planned for intervals of six to eight months. If an upcoming release is delayed, the existing licenses will be extended accordingly, so that the user can keep working until the next version is available.

[Read more ...](#)

Modeling capabilities
SAFETBOX TEAM | BLOG | 05 MARCH 2018

Hazard and risk assessment (HARA)

In this version, we support hazard and risk assessment according to ISO 26262. This supports the user in the following tasks:

- **Functional Hazard Analysis:** For selected functions/components, the user can perform a high-level fault analysis with the help of a configurable set of guide words (e.g., Omission, Commission,...). This template allows the user to document the effects of such failures and to determine whether they are safety-critical or not. If they are, the user can also define and associate hazards with them.
- **Definition of hazardous events (HE):** Given a set of hazards, the user can build hazardous events by defining in which situations these occur. Moreover, the user can also define a hazardous event's occurrence parameters (e.g., regarding the frequency or the time domain) as well as the expected outcome (e.g., an accident). This implementation also allows the user to define their own scenarios by specifying situation groups (e.g., category Driving Location: highway (over 100 km/h), city (50 km/h),...). The current template makes a clear differentiation between driving and standing situations to simplify the analysis and configuration.

Project Examples

Model-based safety engineering e.GO Life

Challenge

- Conduct the required safety engineering activities as per ISO26262 for a newly developed electric car

Contributions of SAF

- Item definition
- Hazard and Risk Analysis
- Creation of a functional architecture
- Safety analysis on the functional architecture
- Development of the functional safety concept
- Partial development of the technical safety
- Validation goals
- Safety case

Customer Benefits / Results

- Creation of safety engineering work products for the e.GO Life



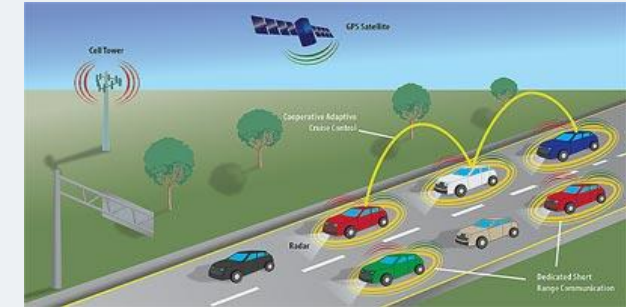
Project Examples

Safety-Security for Major Automotive Industry Client

Challenges

- Development for recent ADAS and planning ahead for Autonomous Vehicles
- Safety and Security co-engineering
- Comply with international automotive standards incl. ISO 26262, 21434
- Integrate with EAST-ADL modeling language

ADAS Use Case: CACC Platooning

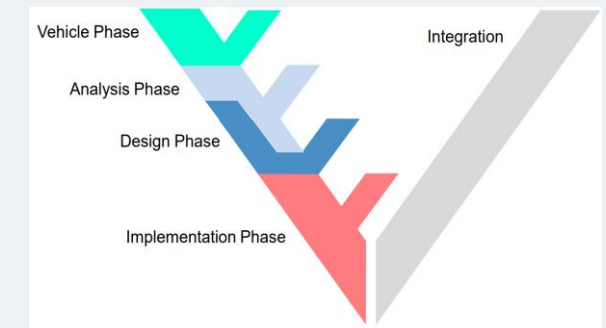


<https://www.fhwa.dot.gov/publications/research/operations/15032/index.cfm>

ISO 26262, ISO 21434



V-Model Lifecycle



Automotive System
Modeling Language

EAST-ADL

Project Examples

Safety-Security for Major Automotive Industry Client

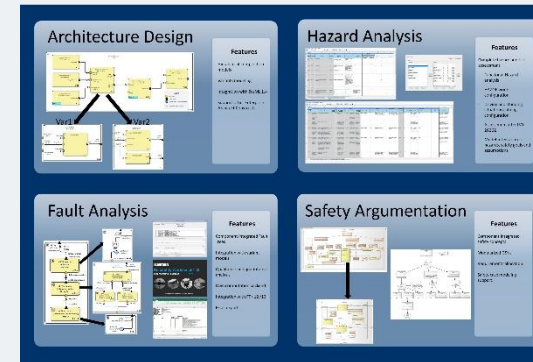
Key Results

- Prototypical model-based tool support in SafeTbox
- Integrated Safety and Security Co-engineering
- HARA + TARA
- Systems-Theoretic Process Analysis
- FMEA + FMVEA
- Component Fault Trees
- Assurance Cases via Goal Structuring Notation
- Support from item definition up to technical design phase

Prototype Model-Based Tool Support



HARA, TARA, CFTs, FM(V)EA, GSN



STPA

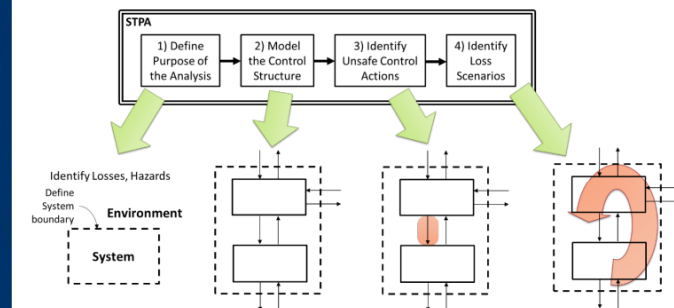
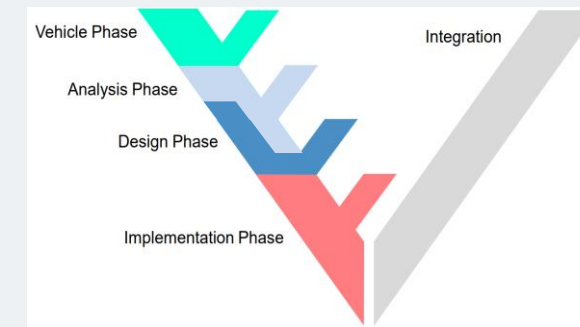


Figure 2.1: Overview of the basic STPA Method

V-Model Lifecycle



Implementation of ISO/SAE 21434 “Road Vehicles – Cybersecurity Engineering”

Service / consulting / training

Target group

- Primary: Automotive manufacturers and automotive suppliers
- Secondary: Companies that supply products and services with corresponding safety and security requirements

Customer challenges

- Implementation of the ISO 21434 standard

Offerings

- Advising the customer on the implementation of the new standard ISO/SAE 21434 “Road vehicles – Cybersecurity engineering” (from August 2021)
- Integration of cybersecurity engineering with safety engineering according to the companion standard ISO 26262 “Road vehicles – Functional safety”
- In cooperation between Div1 and Div2, we offer tool support based on the existing tool safeTBox

References

- IESE-Report 009.22/E “Threat Analysis and Security Requirements Elicitation”
- Huawei (TR 1.7 “Safety-Security-Co-Engineering”)
- NewTec
- Japan Manned Space Systems Corporation (JAMSS)

Network-based Penetration Testing Service



Target group

- Companies that are not sure whether they are vulnerable to external attack and whether the security measures they have taken are effective

Customer challenges

- **Initial situation:** Requirement from various standards and possibly also from the company's own ISMS
- **Issues:**
 - Lack of expertise, staff, and tooling to perform penetration tests professionally
 - Lack of expertise and staff to remediate the identified findings

Offerings

- Independent and neutral determination of the current, externally accessible attack surface
- Verification of the effectiveness of the implemented measures
- Proof of implementation and, if necessary, follow-up of such scans incl. development
- Recommendations for implementation (guidance)

References

- Fraunhofer audits
- MBCC Group
- DAMOVO

AI-assisted Engineering

AI for Software Engineering

Challenges



Human-AI Co-Engineering with Quality in Mind

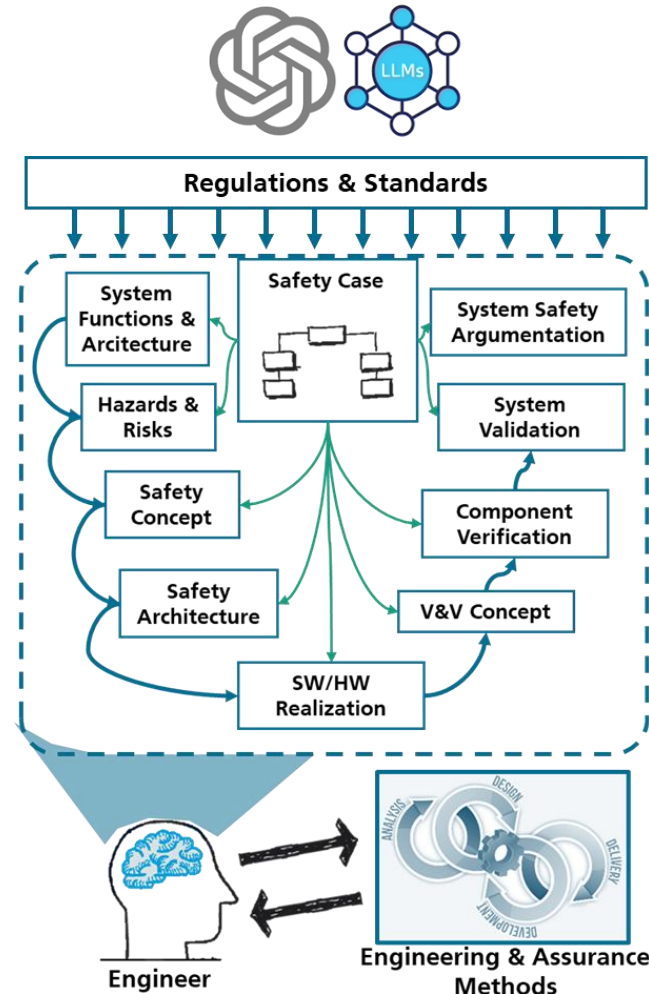
- **Performance, Quality & Security:** A primary challenge is managing GenAI's output, which can lead to efficiency loss, poor software quality, and the introduction of subtle but critical security vulnerabilities that are difficult for developers to detect.
- **Compliance & Safety Certification:** Flaws in GenAI work products complicate adherence to strict industry guidelines like ASPICE or CMMI and create significant hurdles in the approval and certification process for safety-critical systems.
- **Accountability & Tool Qualification:** A key governance question is “who is responsible when AI-generated software fails?”. This is linked to the lack of reliable methods to formally validate and qualify AI-based tools for use in professional, high-risk environments.
- **Data Privacy & Intellectual Property (IP):** There is a major business risk in protecting proprietary source code from being sent to third-party cloud services. This is compounded by the legal uncertainty surrounding the ownership and licensing of AI-generated code. In addition, models trained on “unknown” data provide opportunities for attacks.
- **Human Expertise & Skill Evolution:** A critical task is redefining the essential competencies for professionals. This includes finding ways to teach and maintain irreplaceable, foundational skills as the experience of building systems “from scratch” becomes less common due to over-reliance on AI. Technical debt can become a major issue.
- **Organizational & Financial Overhead:** Beyond individual productivity, companies face a high total cost of ownership from subscriptions and hardware. Furthermore, there is significant effort required to integrate, continuously monitor, and update AI tools effectively within established development pipelines, corporate processes, and training programs.

AI-assisted Safety Engineering

Use Cases and Research Areas

Use Cases

- **Smart Standards:** From standards / regulations to project-specific requirements
- **Smart HARA/TARA:** From hazards to critical scenarios
- **Smart Test Cases:** From safety requirements to test cases / code assertions
- **Smart Traceability:** From unstructured legacy safety artifacts (e.g. Excel) to tool-overarching traceability analyses

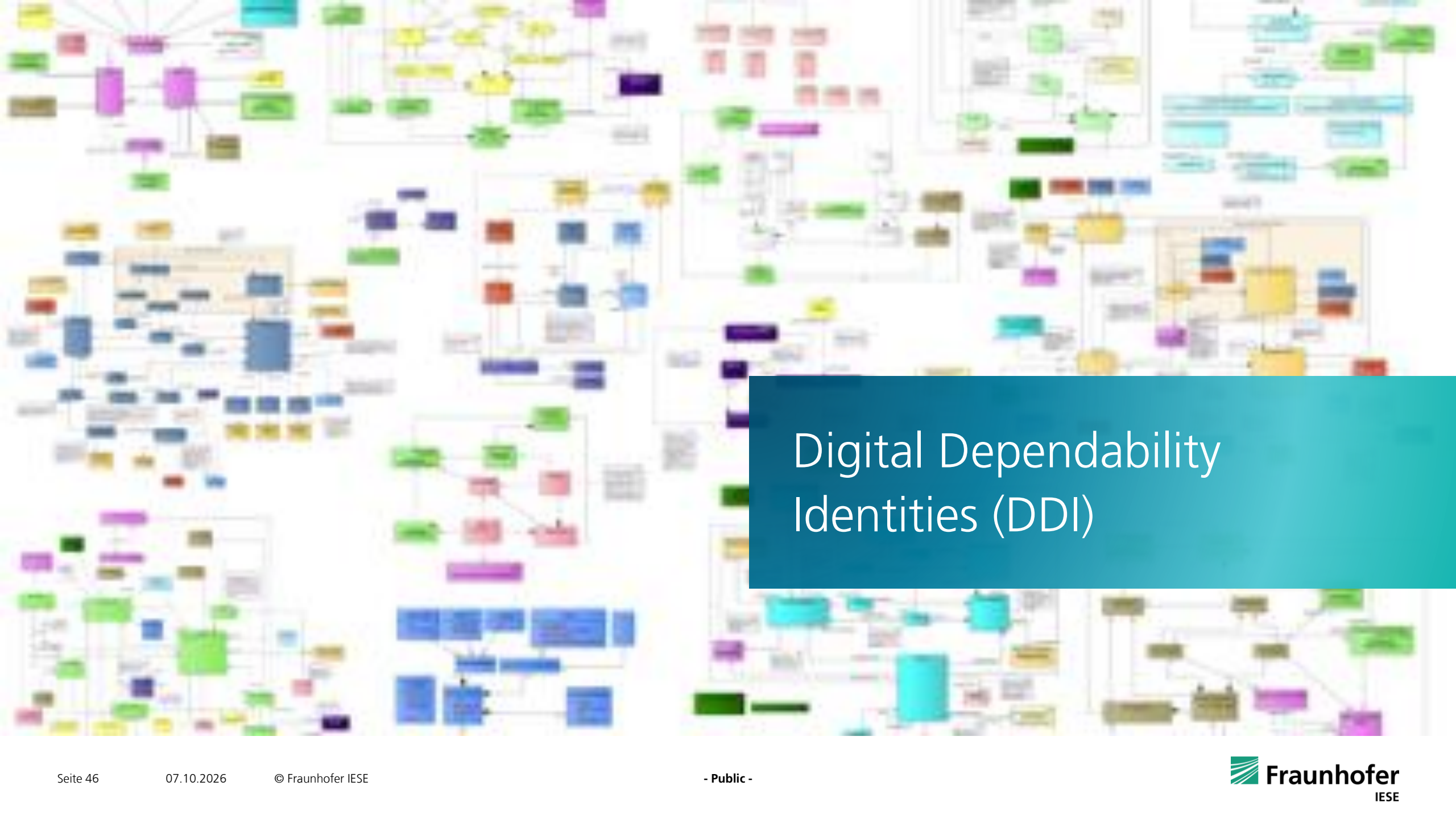


Research Areas

- More efficient and effective safety engineering processes with AI
- „Safe“ Engineer-AI Collaboration Architectures
- Uncertainty Quantification of LLM output
- Risk-based qualification of AI tools considering applicable standards like ISO/IEC CD TS 22440-1

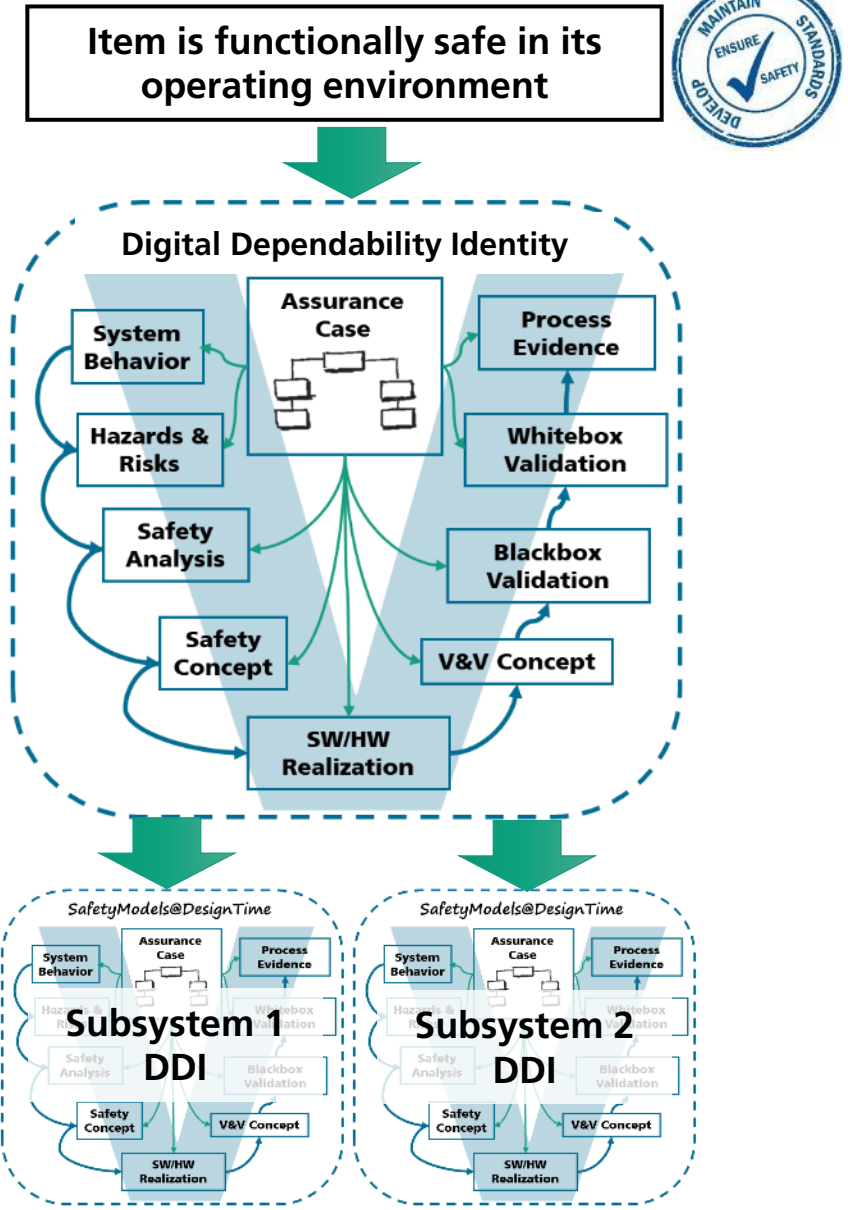
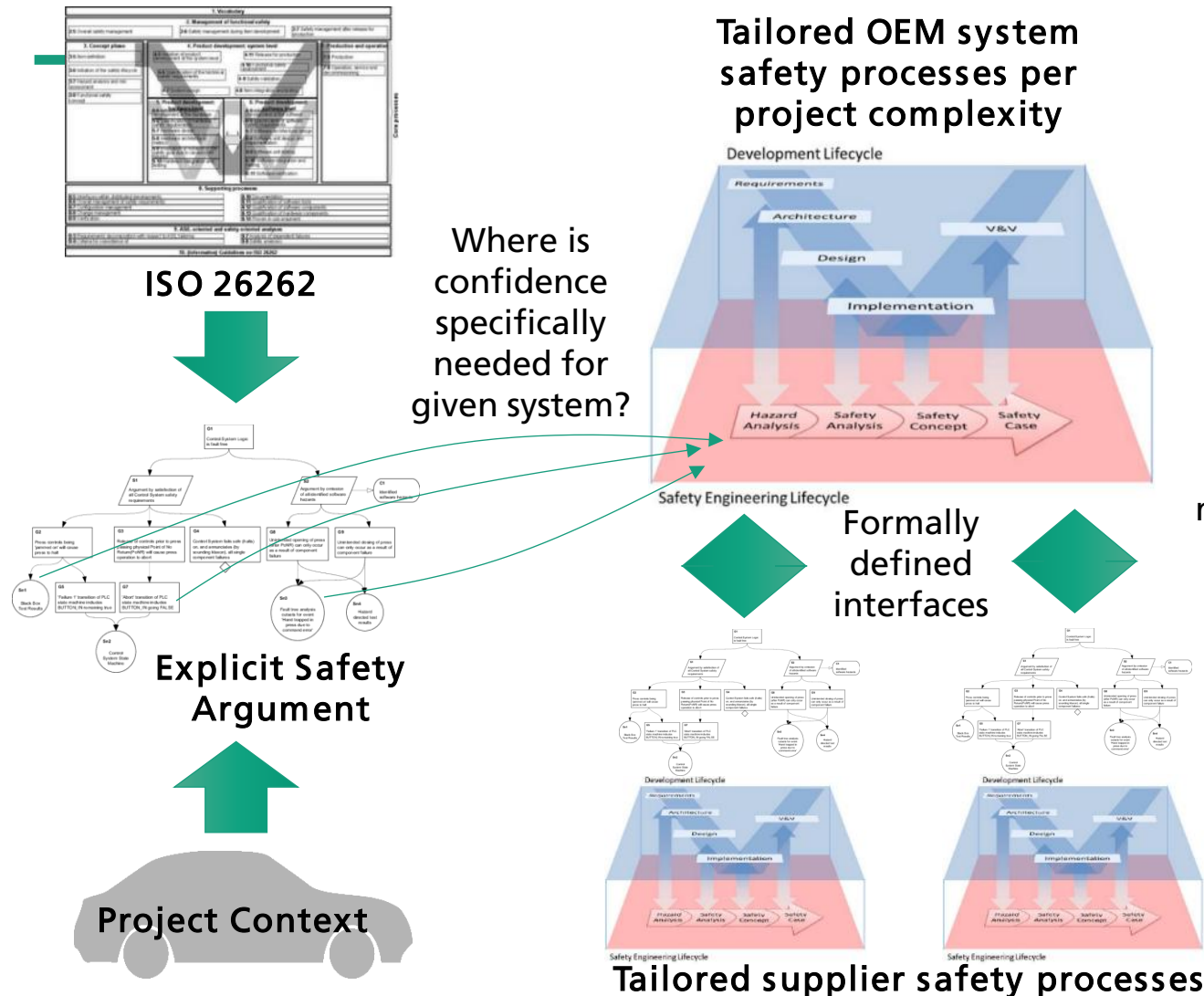
Experience.
Efficiency.
Excellence.

3 Years

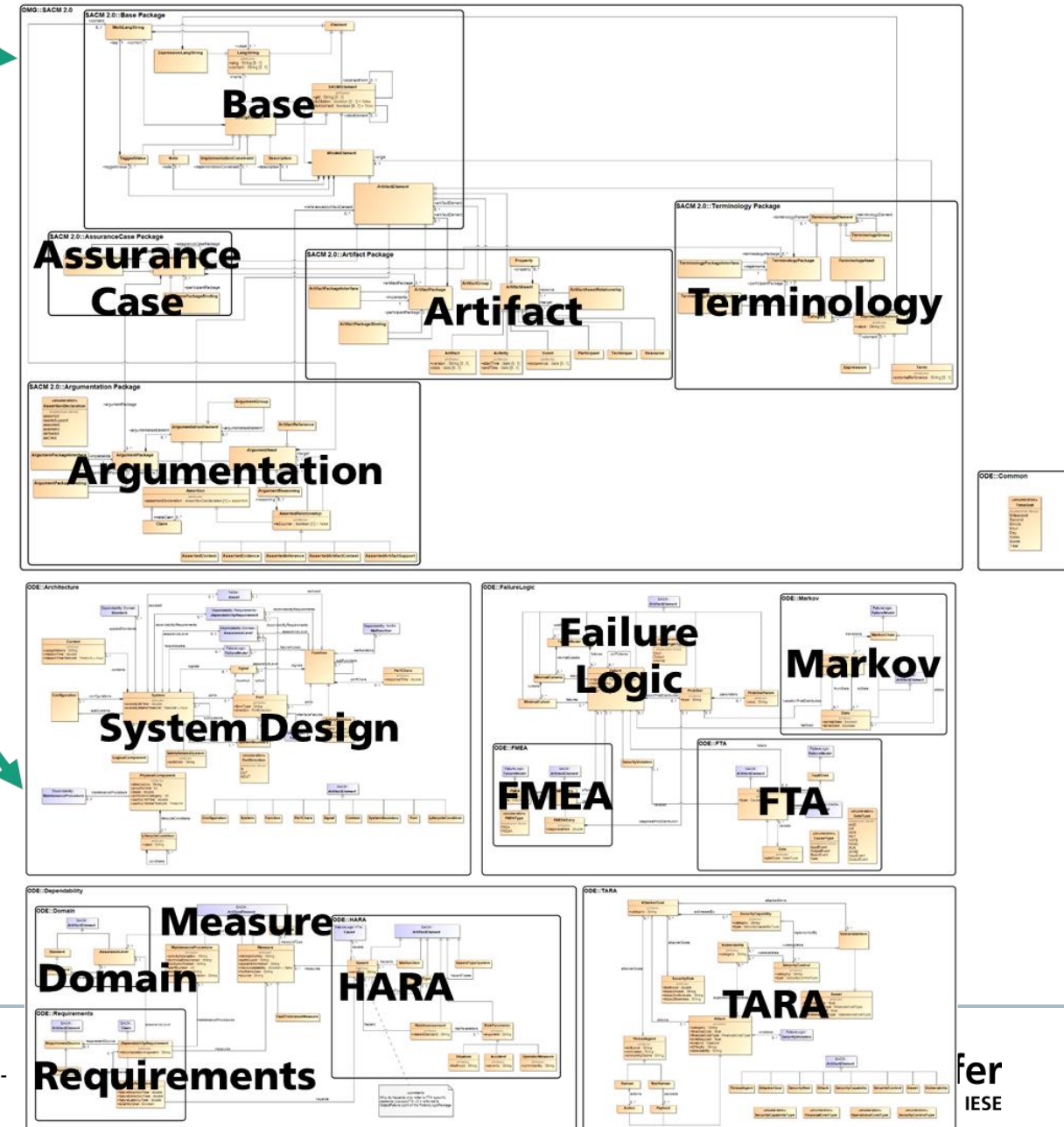
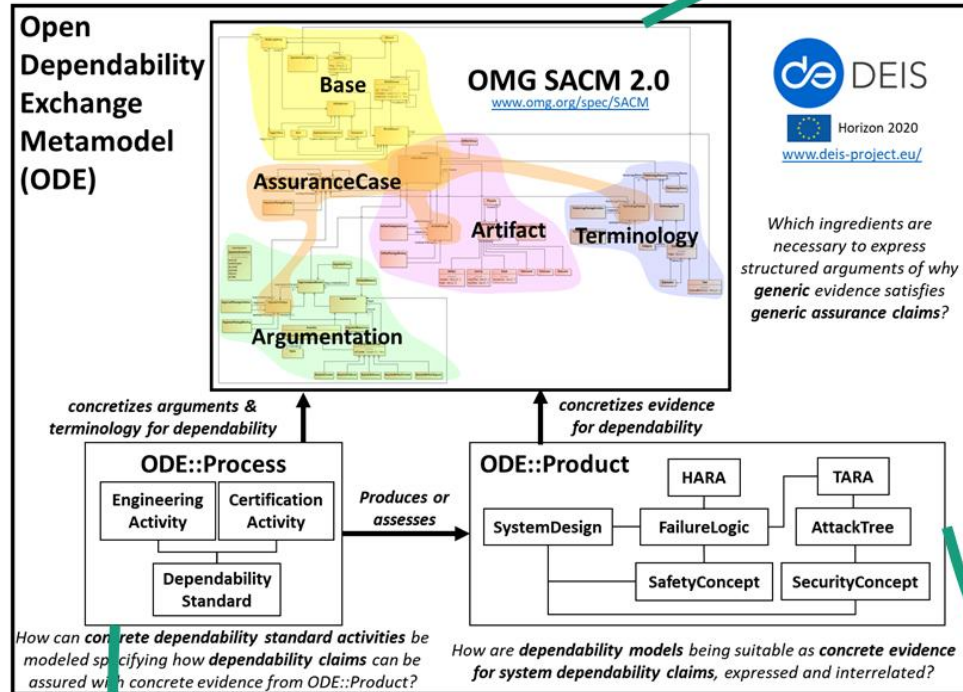


Digital Dependability Identities (DDI)

(Automotive) safety engineering in the future



DDI (and ODE) Details



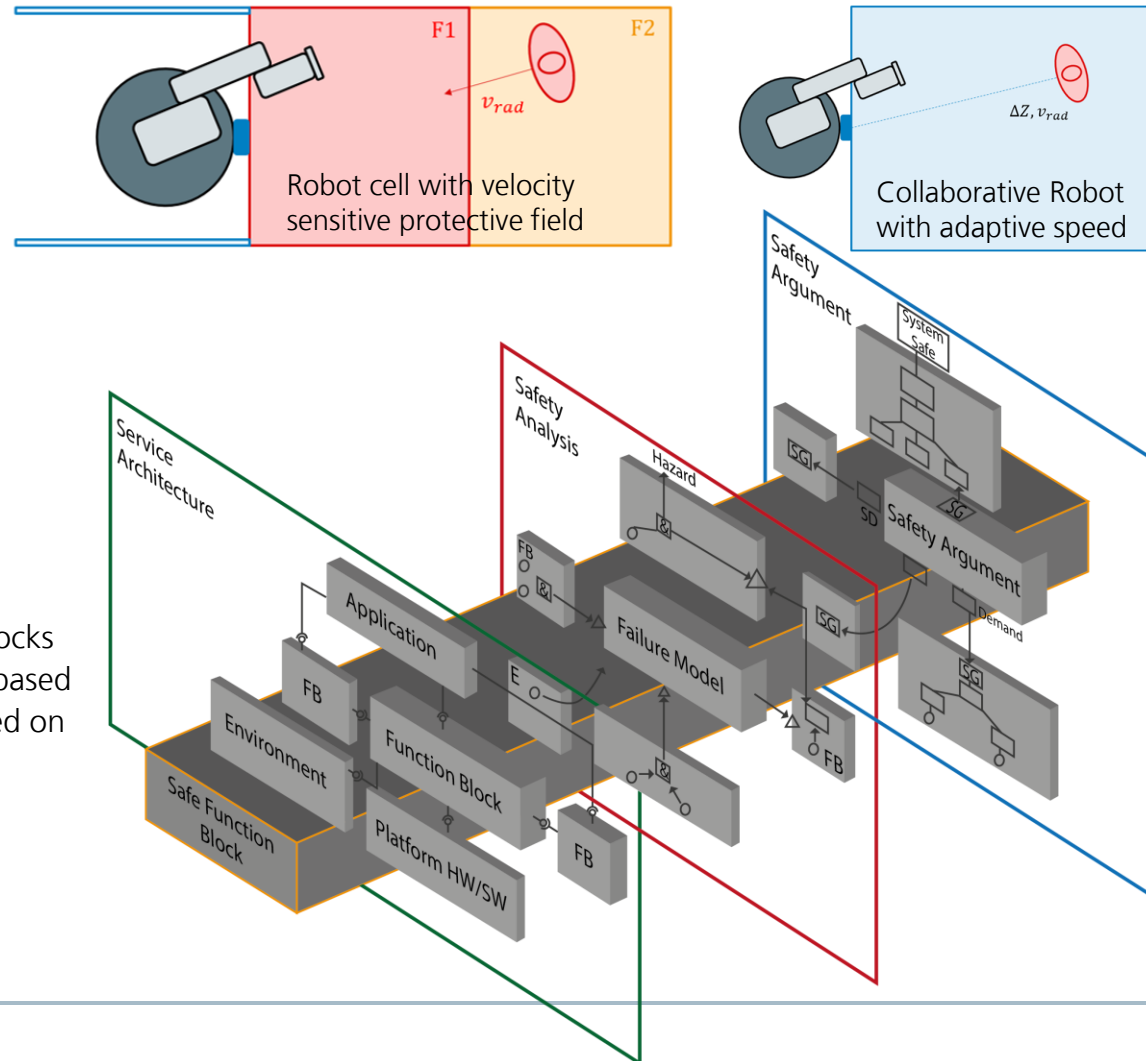
SICK: Modular and safe function blocks in CPS scenarios

Reuse of modular function blocks in different application contexts and on different sensor hardware

External: Use Case Description

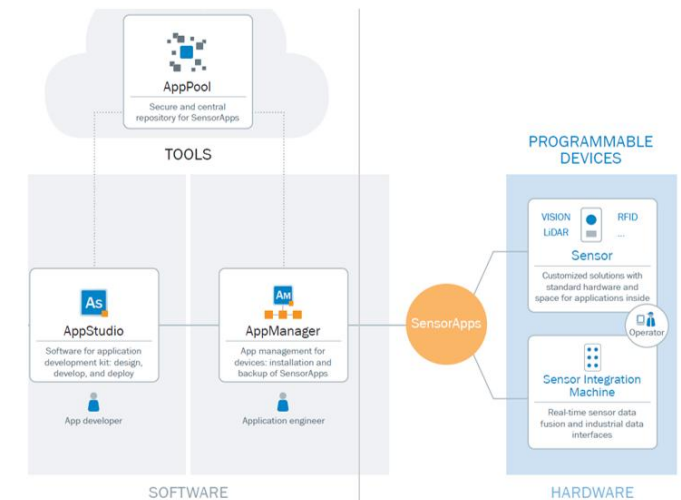
ConSert Know-How for modular safety architectures

MBSA Know-How



Consert Engineering Case Study for safe modular function blocks for sensors in the industrial automation context

safeTbox models: Service Architecture, CFTs, Modular Safety Concept, GSN Argument Fragment



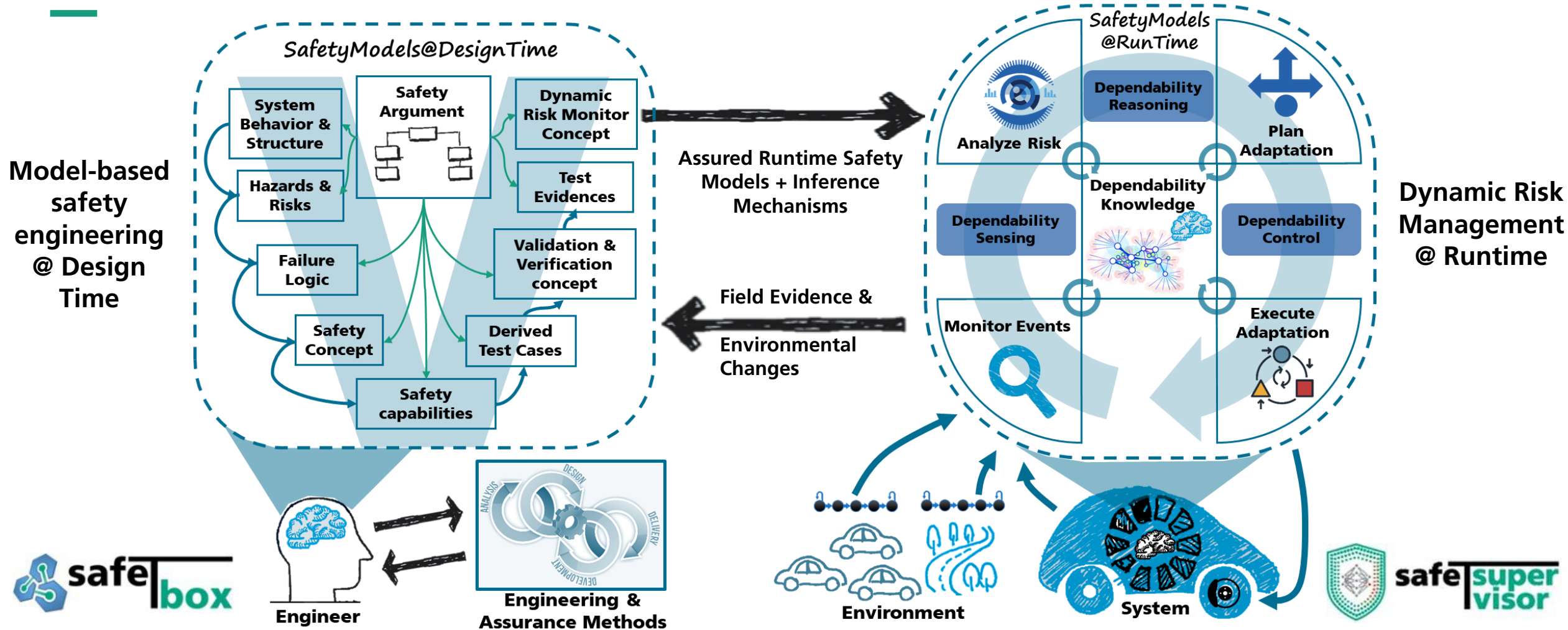
- Goal: Facilitate reuse of function blocks through modularizing architecture based on services and safety artifacts based on ConSert engineering methodology



Dynamic Risk Management

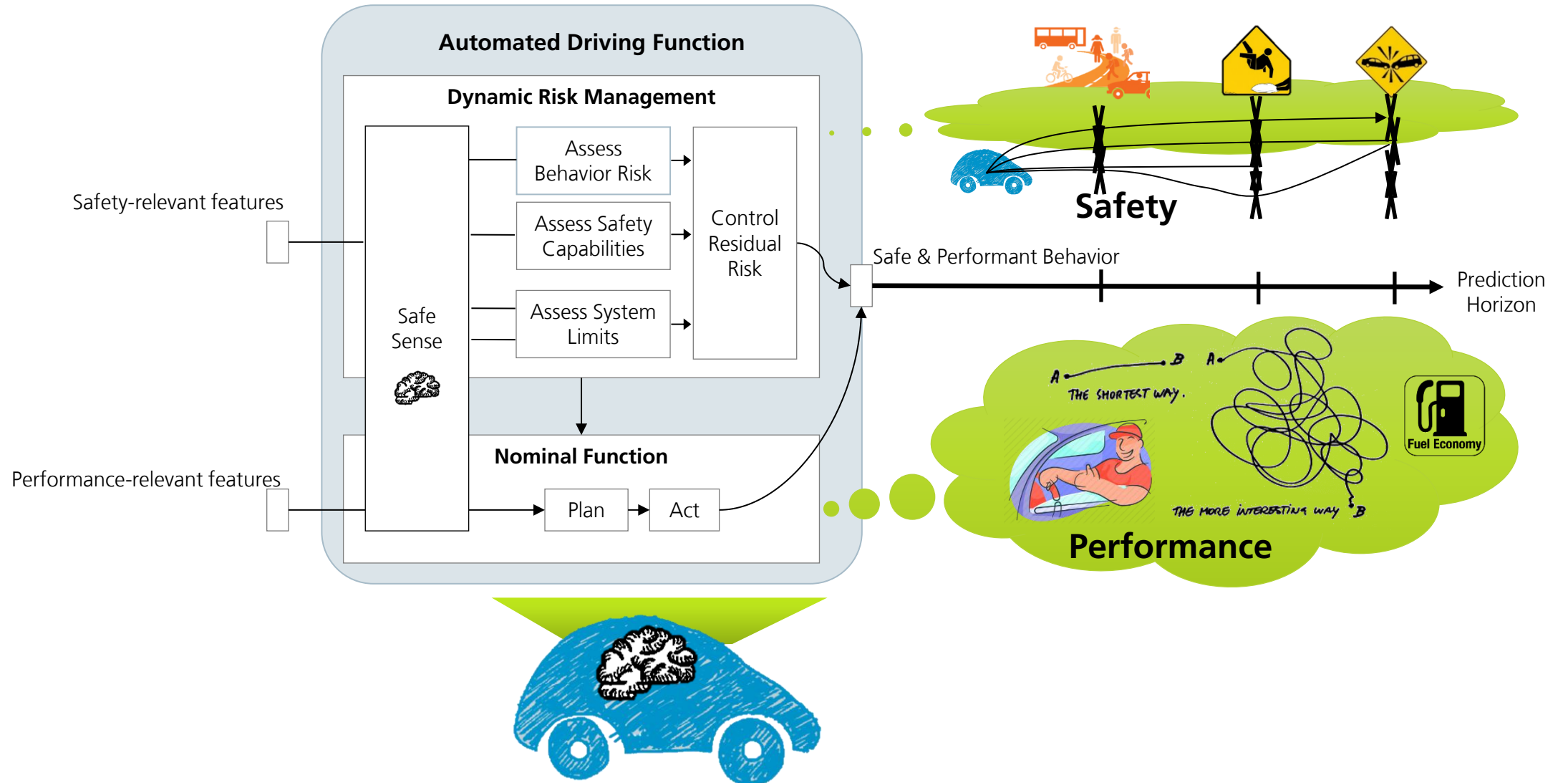
Dynamic Risk Management @ Fraunhofer IESE

15 years research trajectory in assurable simtime/runtime dynamic risk monitoring models

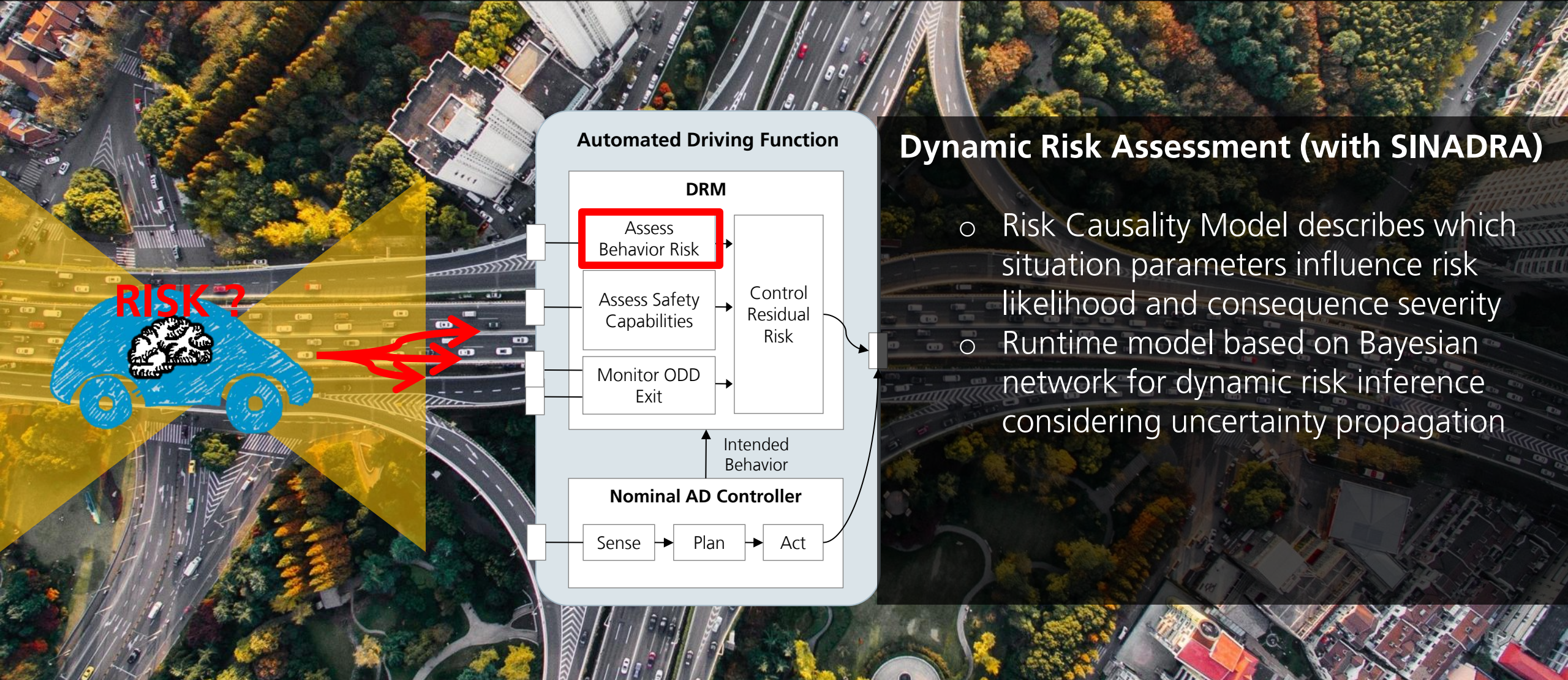


Schneider, Daniel, et al. "Dynamic risk management in cyber physical systems." Engineering Safe and Trustworthy Cyber Physical Systems: Essays Dedicated to Werner Damm on the Occasion of His 71st Birthday. Cham: Springer Nature Switzerland, 2025. 127-146.

Dynamic Risk Management – Making systems safety-aware



DYNAMIC RISK MANAGEMENT FOR AUTONOMOUS SYSTEMS



Autonomous shuttle operation in semi-structured pedestrian zones

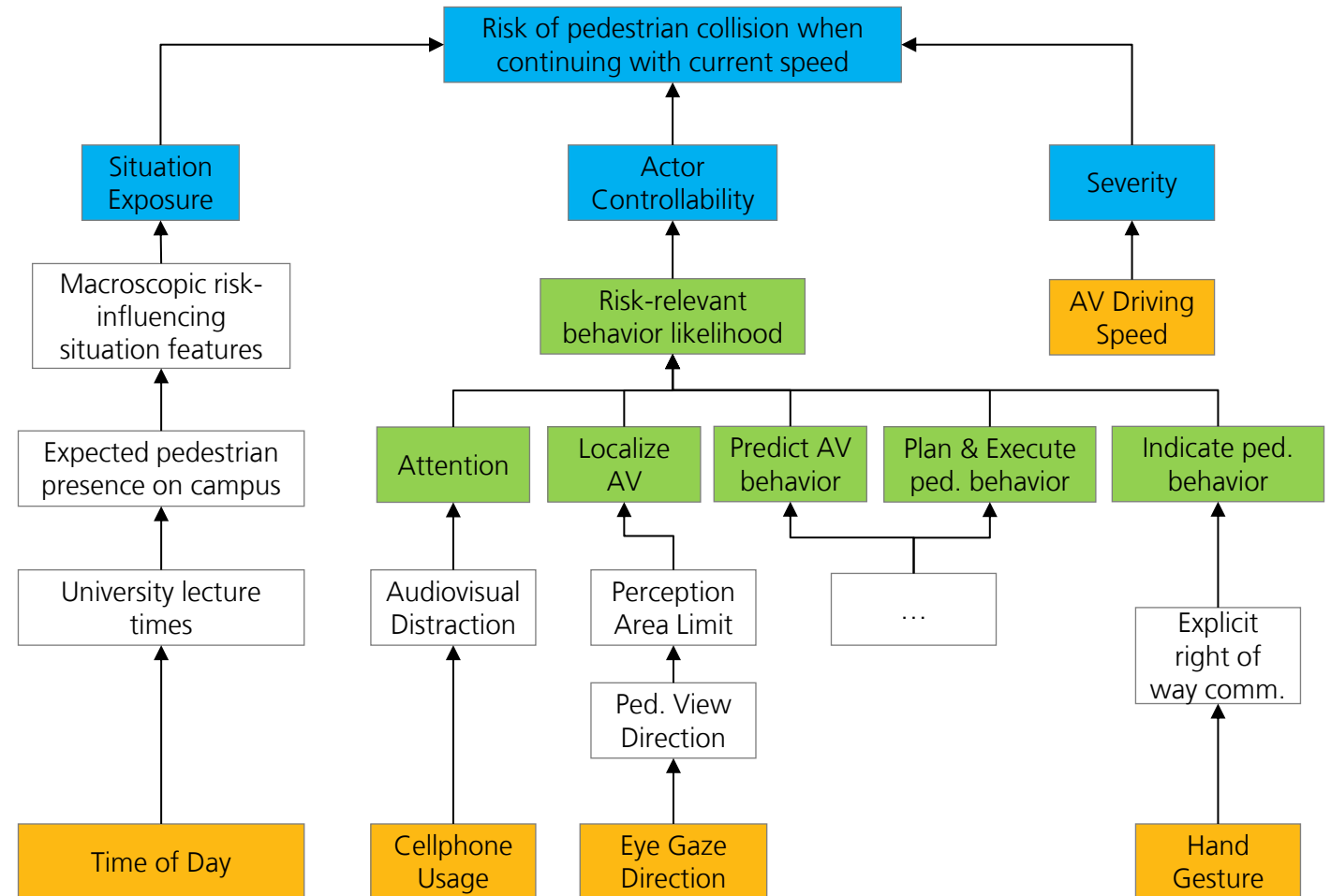
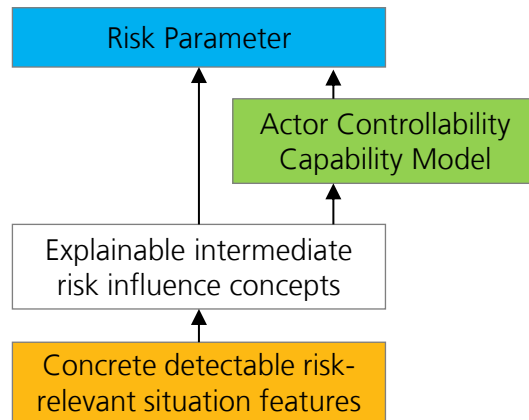
Dynamic Risk Assessment for a Campus Shuttle

- Operation on campus, city center zones, i.e. semi-structured environments
- Issue: Freezing robot problem
- Cause: Too conservative dynamic risk assessment
- Effect: False positive safety interventions
- Solution: Enable system to judge risk of pedestrian more intelligently by using improved situational awareness (=Situation-Aware Dynamic Risk Assessment)



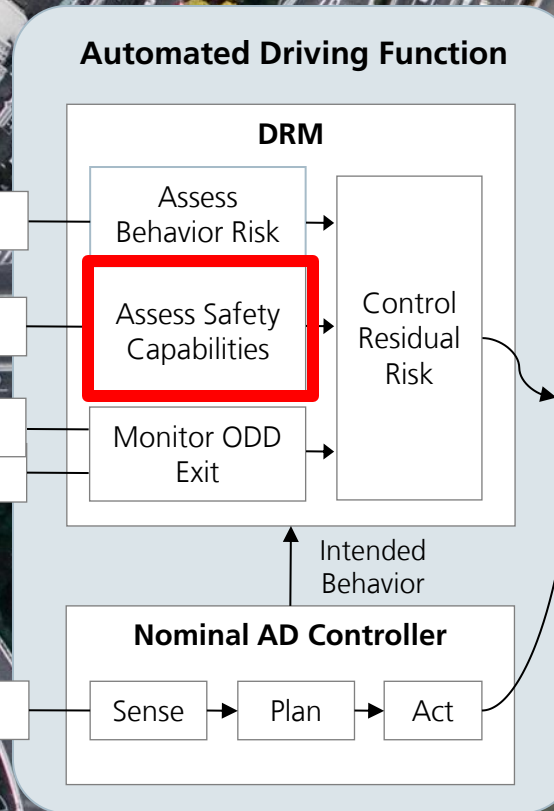
SINADRA Risk Decomposition Concept

Dynamic Risk Assessment for a Campus Shuttle



DYNAMIC RISK MANAGEMENT FOR AUTONOMOUS SYSTEMS

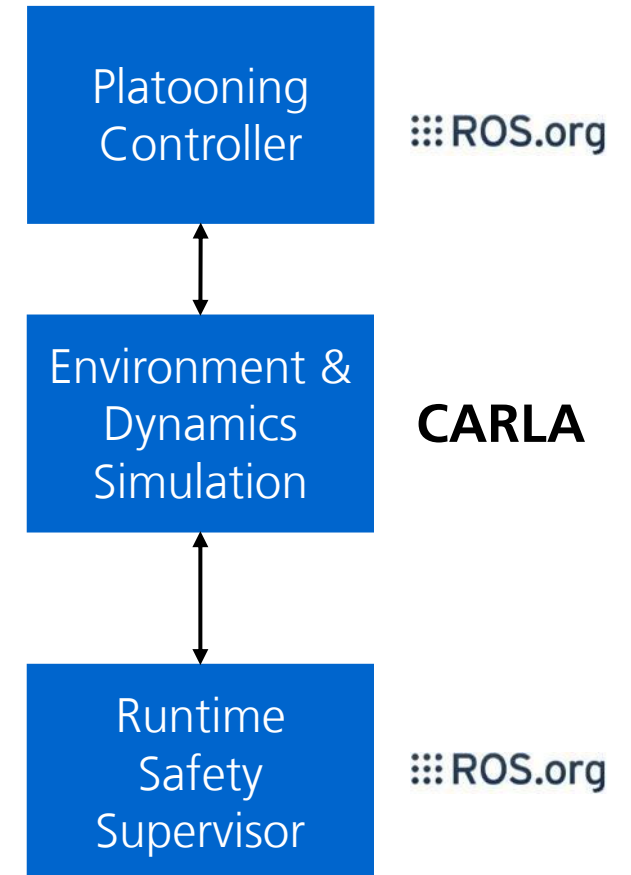
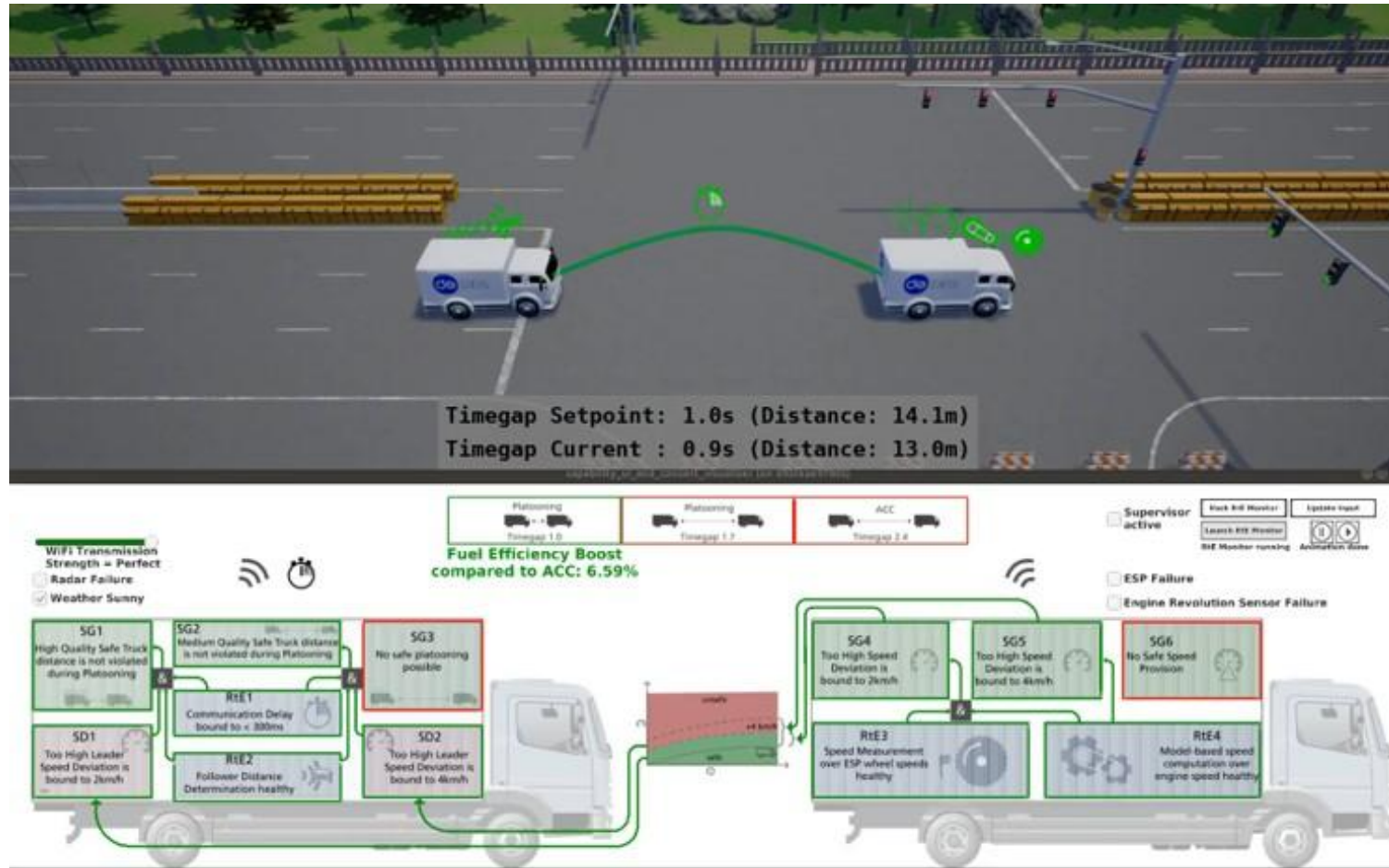
Capabilities?



Dynamic Capability Assessment (with Conditional Safety Certificates)

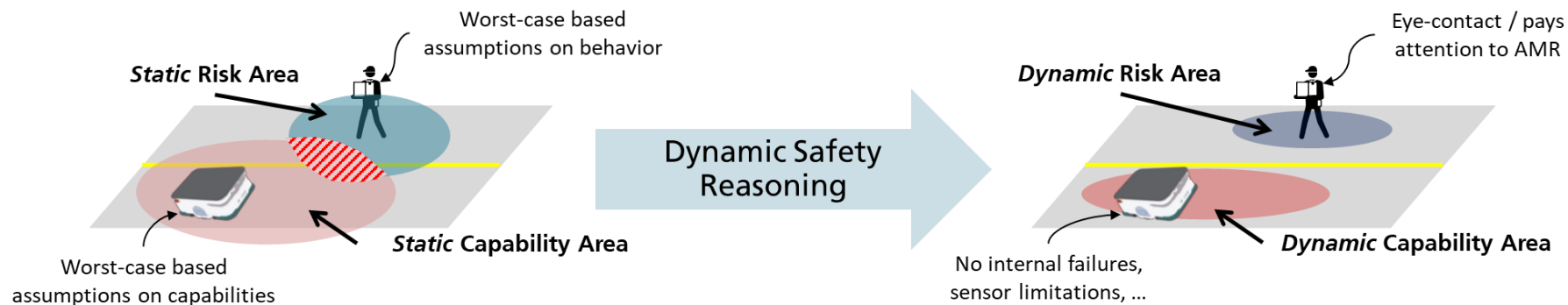
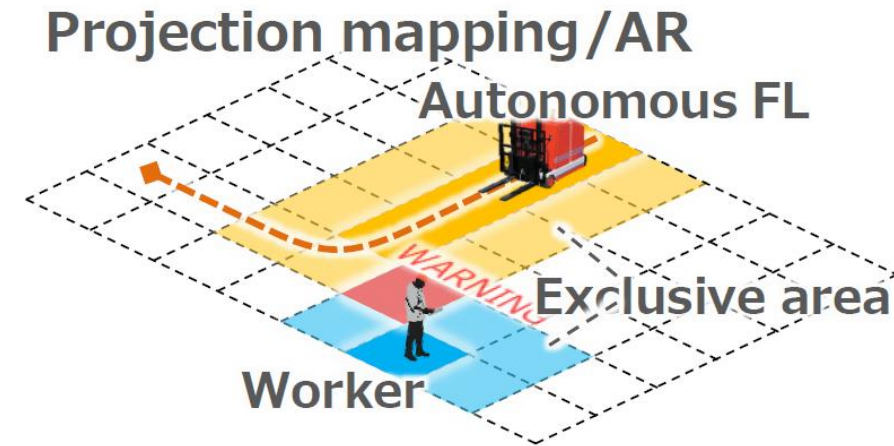
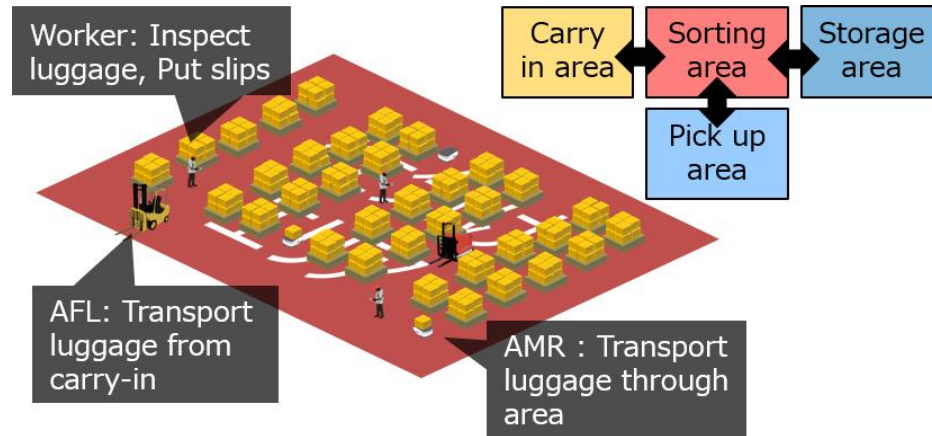
- Modular safety concept variant runtime models
- Algorithms for runtime composition and safety guarantee evaluation
- Supports heterarchically structured systems

Demonstration of Dynamic Capability Assessment for cooperative platooning

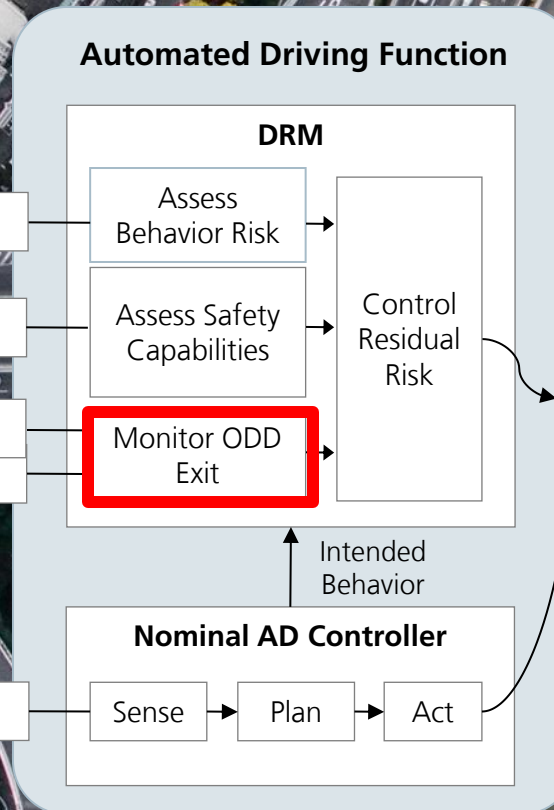
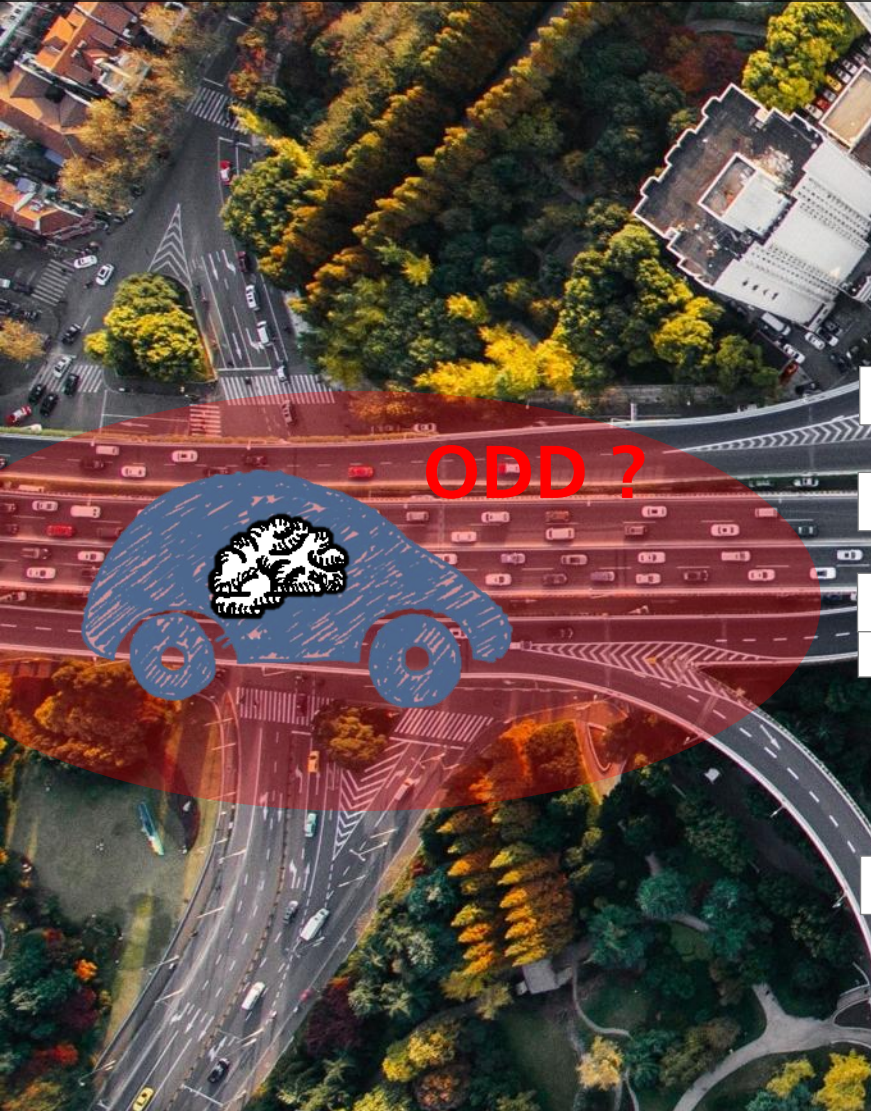


Hitachi 2021/22/23: Safety of Autonomous Mobile Robots in Smart Logistics

Method transfer for engineering dynamic CPS safety concepts for human-machine cooperation



DYNAMIC RISK MANAGEMENT FOR AUTONOMOUS SYSTEMS



Dynamic ODD Monitoring

- Definition: SAE J3016
- Language: OpenODD
- Content: BSI/PAS 1883
- Process: AVSC
- Runtime monitoring of ODD potentially integrated with DRA and DRC

Predictive Autonomy Lab – Overview

Empirical Causal Driver Behavior Research Platform

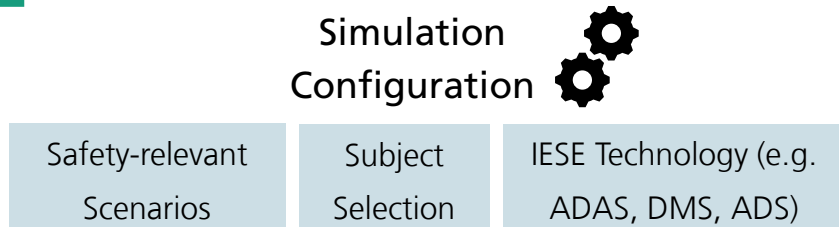
PREDICTIVE AUTONOMY LAB



Predictive Autonomy Lab @ Fraunhofer IESE

Experimental Workflow

Our Focus: Problem and Hypothesis Definition
motivated by Safety Engineering



DATA GENERATION

Research Hypothesis based on relevant aspects for safe ADAS supervision and ADS



PROBLEM

- Traffic law
- Safety engineering competence
- Use case + system + domain experts
- ...

Simulation Data



Human Subject Observation Data



DATA



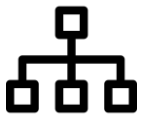
Researcher with tools

ANALYSIS



Research Hypothesis Evaluation

SOLUTION



Situation-specific Human Behavior Model

Intelligent DMS, DRM-Component, Environment Simulation Model

Predictive Autonomy Lab

Components



HTC VIVE Tracker



VR-Glasses HTC VIVE Focus Vision



Driving Simulator HW + SILAB SW



Eye-Tracker



Infrared-Marker



Hautwiderstand-Sensor



Physiological Sensors + SW



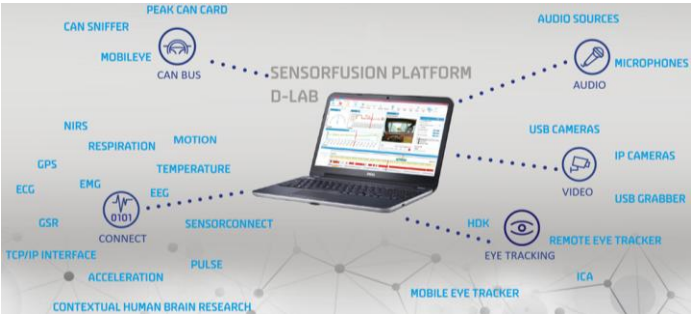
Pedestrian Platform Omni One



3xWebcam



Microphone



D-Lab Sensorfusion Platform



Ensuring safety of systems with ML components

Safety challenges

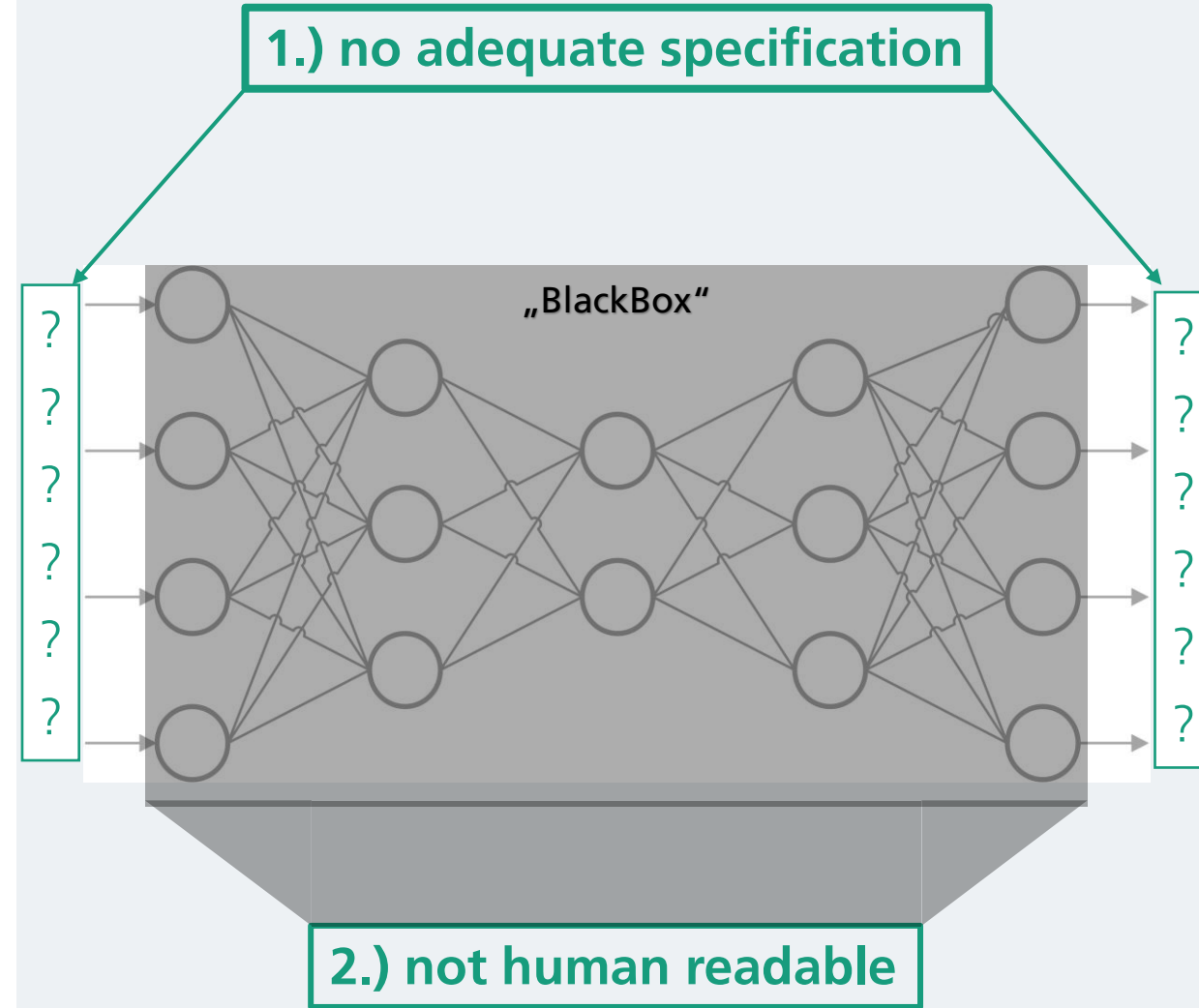
Two key issues

Typically, a sound requirements specification is missing

- There will be training data and maybe a partial requirements specification
- This is not very surprising, because ML is particularly attractive to address problems where it is hard to come up with a sound specification (e.g. camera-based object classification)
- This complicates V&V and the generation of sound evidence for a safety argument

In addition, proper analysis and verification is difficult due to a lack of explainability

- A Blackbox: Established WhiteBox Techniques (such as Inspections, Walkthroughs) not applicable
- Apparently insignificant changes at the inputs can lead to very significant changes at the output



Building blocks for a safety concept



ROBOT SYSTEM

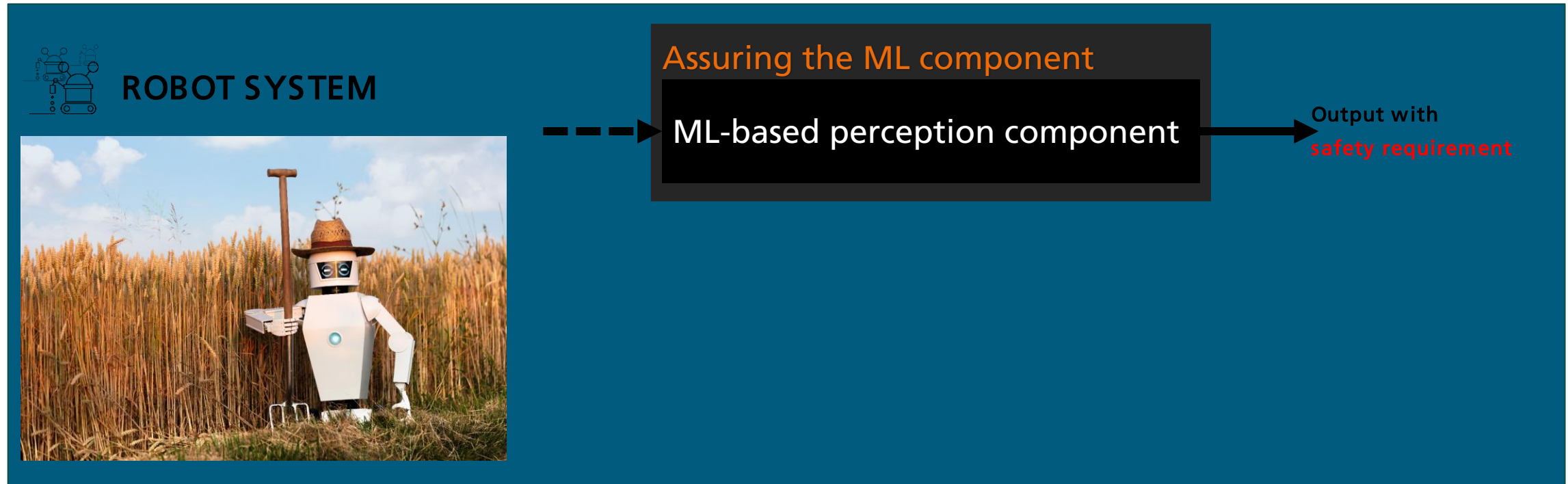


ML-based perception component



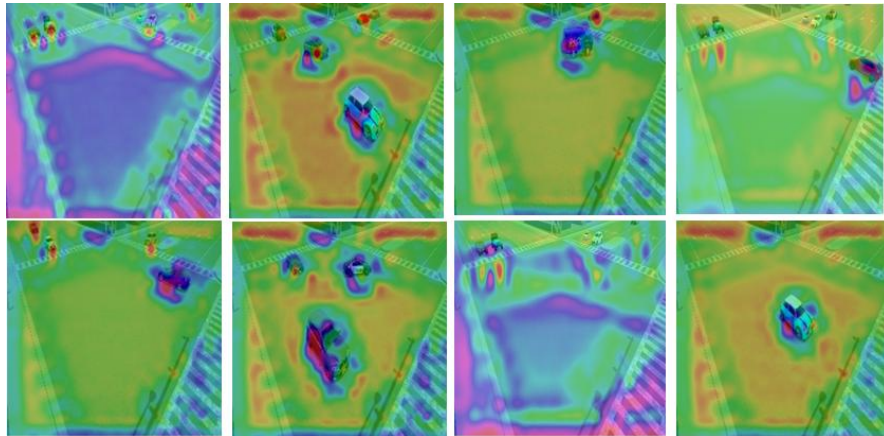
Output with
safety requirement

Building blocks for a safety concept

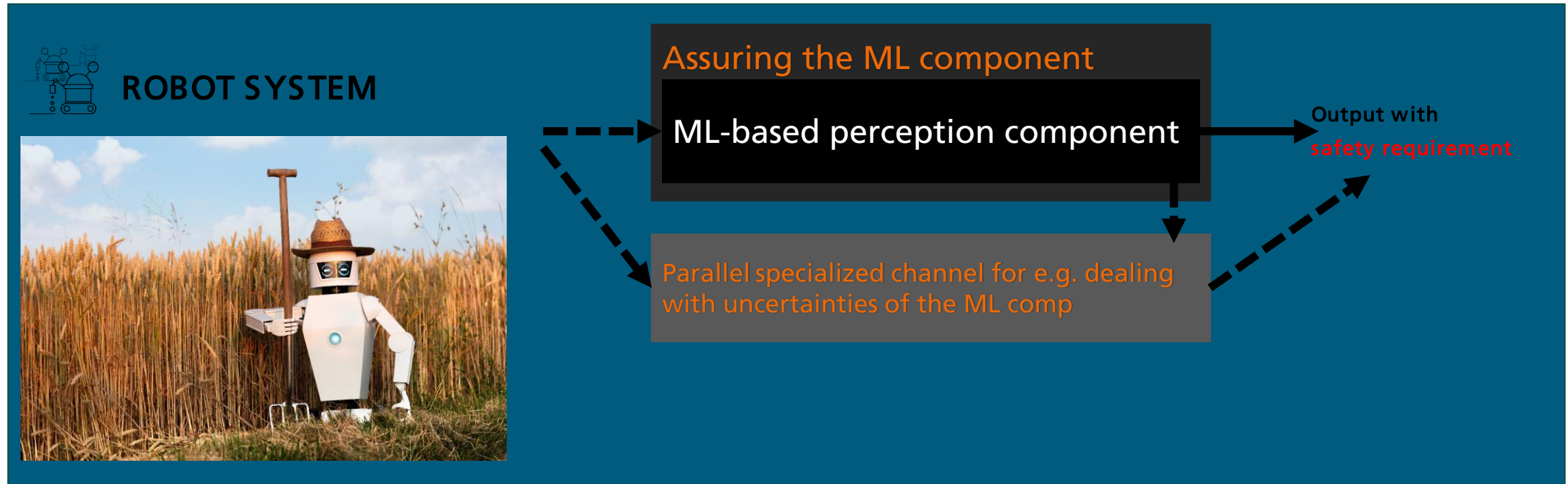


Starting points for assuring DL components

- Methods and techniques for analyzing and hardening (zB: XAI)
- For example, in case of object classification and conv nets, you can improve the performance of the NN by using techniques such as heatmapping or GradCAM, but you cannot assure it will always work
 - I.e. you cannot prove that every classification will be correct
- In general, guidance is required wrt. adequacy of techniques and generated evidence



Building blocks for a safety concept



Neuro-symbolic AI

Combining ML/neural AI with complementary approaches

- Combination of neuro AI and symbolic (AI) approaches to benefit from complementary strengths and weaknesses
- Endless possibilities to combine approaches
- From a safety perspective, e.g. using a combination of a neural network and a decision model or a FSM could be interesting (to better link to a safety argumentation)
- Open (research) questions:
 - Which approaches to combine for a specific purpose?
 - How to integrate specific neuro and symbolic approaches?
 - ...

NEURO

Context
/ Inputs

Neural Network / Learned Model

Outputs

SYMBOLIC

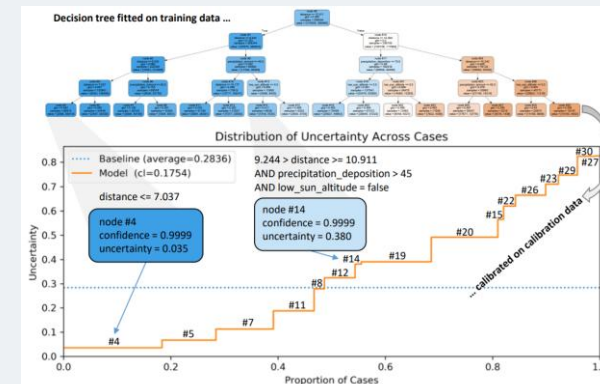
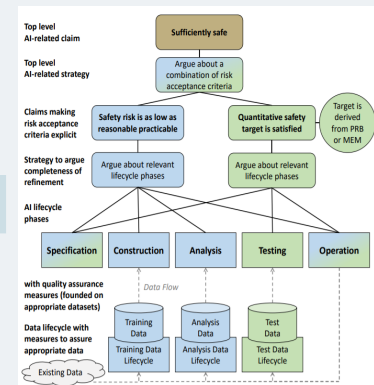


Fig. 3 A calibrated decision-tree-based quality impact model with confidence=0.9999 and its evaluation, visualizing uncertainty estimates and certainty loss (cl) in comparison to DDM baseline.

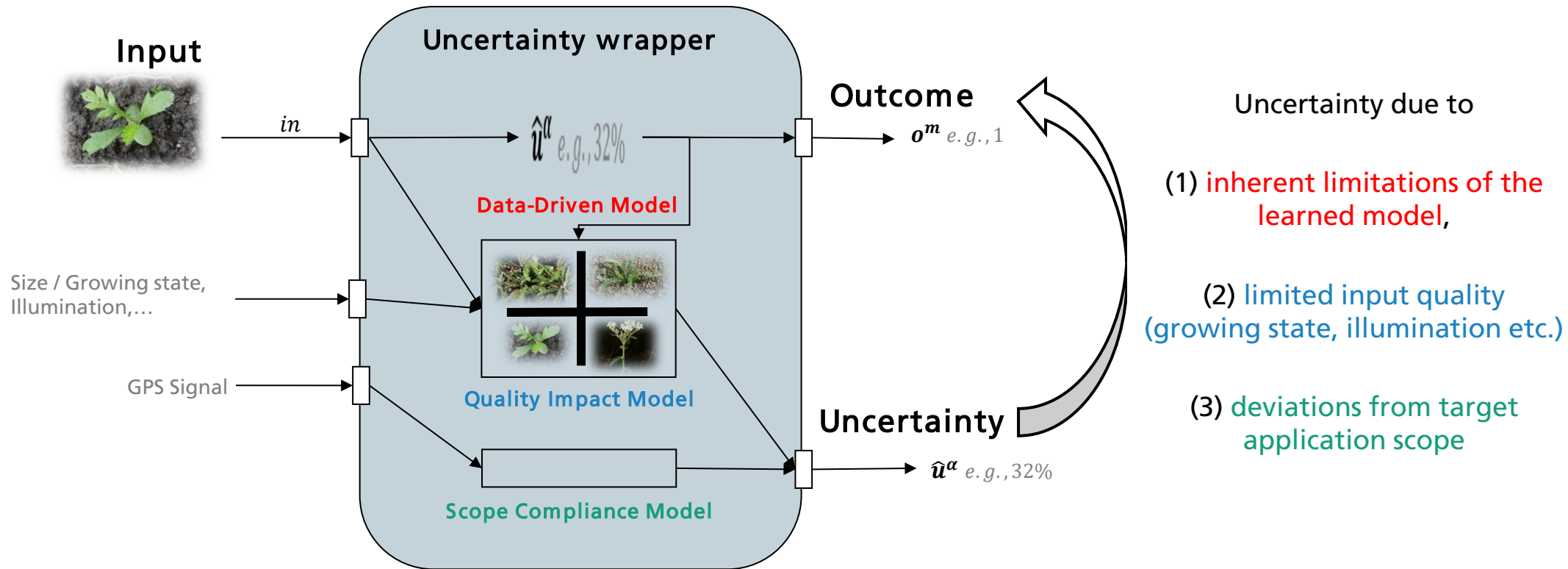
Symbolic Model



Assurance Case

IESE Approach: Uncertainty Wrapper

Making uncertainty explicit



Reference Project BOSCH – Friction analysis and distance control

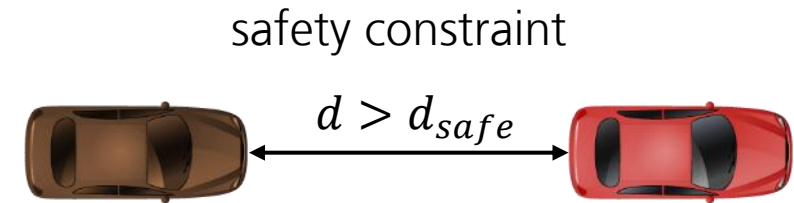
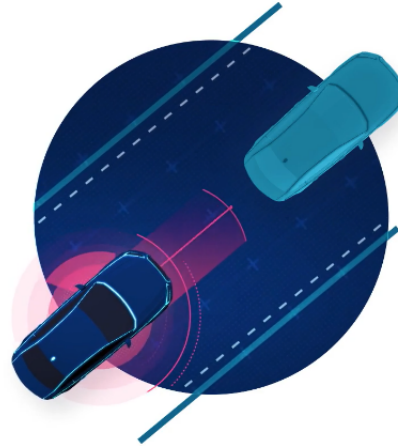
Don't hit the car in front of you

Based on common laws of physics and a century of driving experience, new human drivers are taught to leave enough space between themselves and the car in front of them in order to provide enough time to react if the car in front were to suddenly brake. The 'Two Second Rule' is an intuitive way for any human driver to leave a safe distance without needing to calculate the velocities of both cars, the driver's reaction time, and the front vehicle's braking capability.

The AV, on the other hand, can be programmed to create and operate within a safe following distance, based on the formula below.

$$d_{min} = \left[v_r \rho + \frac{1}{2} a_{max} \rho^2 + \frac{(v_r + \rho a_{max})^2}{2\beta_{min}} - \frac{v_f^2}{2\beta_{max}} \right]_+$$

The moment the distance between the two cars is less than d_{min} , the AV brakes until a safe following distance is restored, or until the vehicle comes to a complete stop, whichever occurs first.



$$d_{safe} = \left[v_F \rho + \frac{1}{2} a_{max,acc,F} \rho^2 + \frac{(v_F + \rho a_{max,acc,F})^2}{2a_{min,brake,F}} - \frac{v_L^2}{2a_{max,brake,L}} \right]_+$$

$$a_{max,brake,L} = \min \left(\frac{F_{b,traction,limit,L}}{m_L}, \frac{F_{b,brakesystem,limit,L}}{m_L} \right)$$

$$F_{b,traction,limit} = m_L \cdot g \cdot \mu_L$$

Those (und other) values of variables are affected by uncertainties

IESE Approach: "SafeML"

Measuring similarity of training data and field data

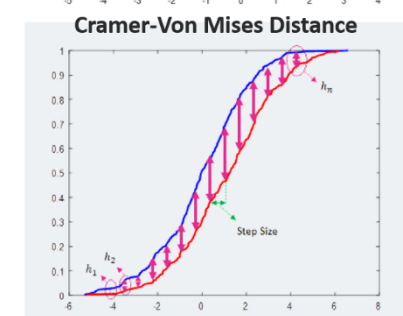
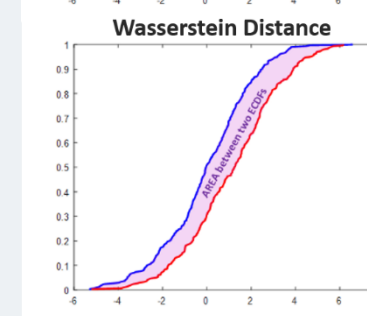
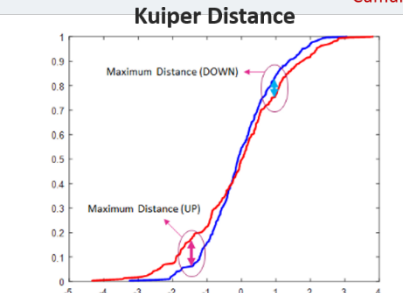
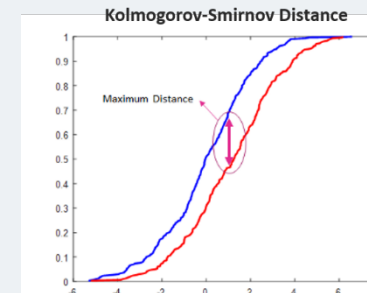
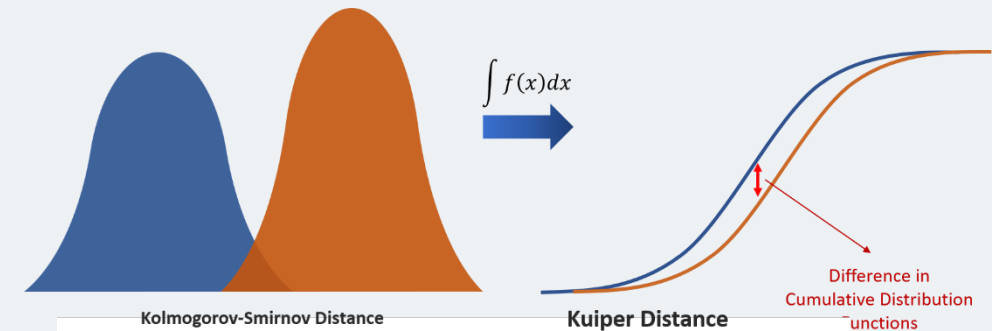
- Challenge: How to monitor scope compliance ? How do we know we're operating in the intended context ?
- Approach: SafeML uses statistical distance measures to evaluate 'how far' from our trained context are we currently operating in. If exceeding user-specified thresholds, alternative actions can then be employed.
- Benefits
 - Monitor uncertainty of operational context compliance
 - Maintain safe state by not trusting ML when out of intended context



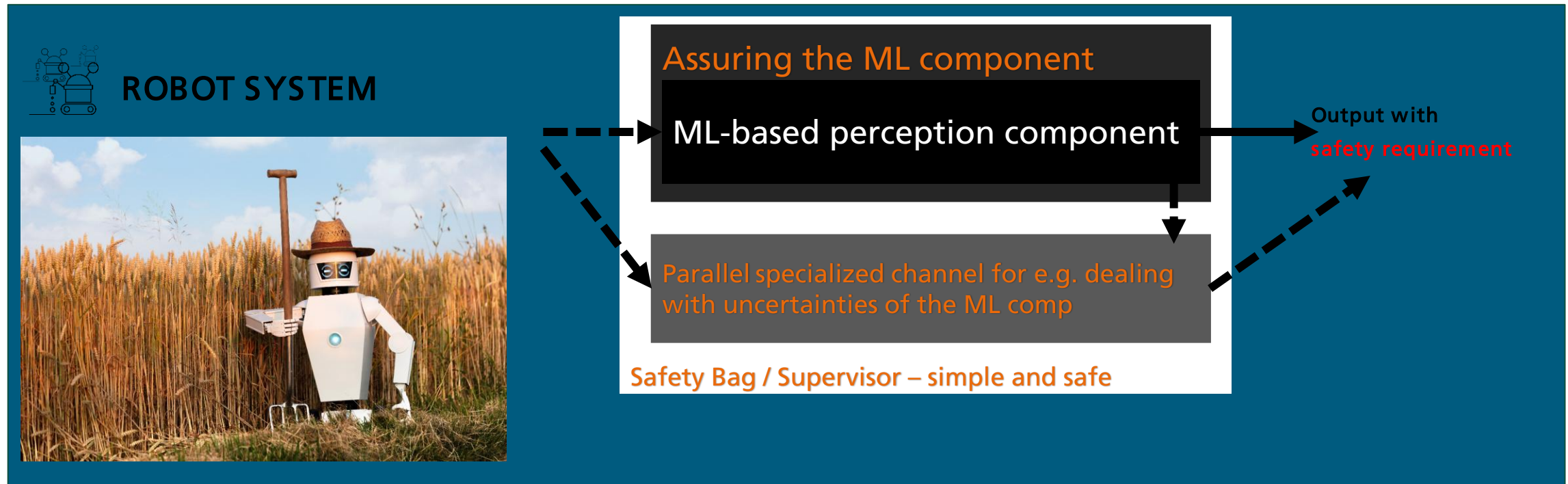
<https://github.com/ISorokos/SafeML>

UNIVERSITY
OF HULL

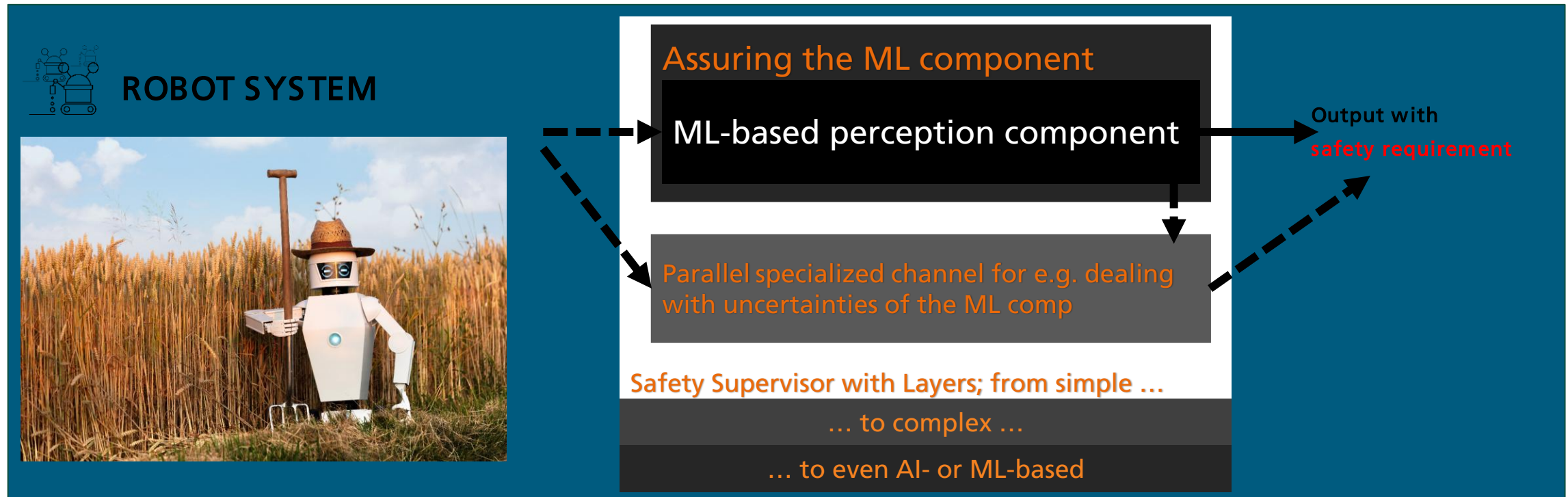
TECHNISCHE HOCHSCHULE NÜRNBERG
GEORG SIMON OHM



Building blocks for a safety concept

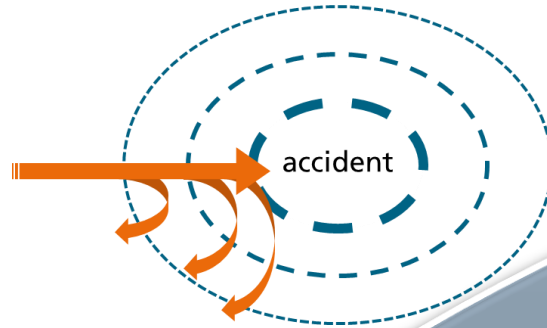


Building blocks for a safety concept



Layered dynamic risk assessment concepts

USING AI TO IMPROVE SAFETY



"Conventional" last layer of defense

Static rules, high integrity sensors. Examples: Avoid self-acceleration (during manual driving); Avoid differential torque higher than 1200 NM

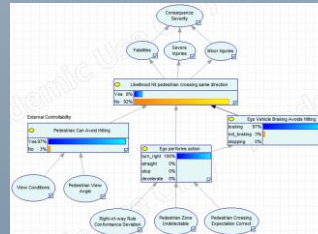
Rule-based dynamic risk management using vehicle sensors

Means to specify and implement safe behavior without considering every situation explicitly

Example: Risk metrics like time-to-collision etc.



(AI-based) situation aware dynamic risk assessment



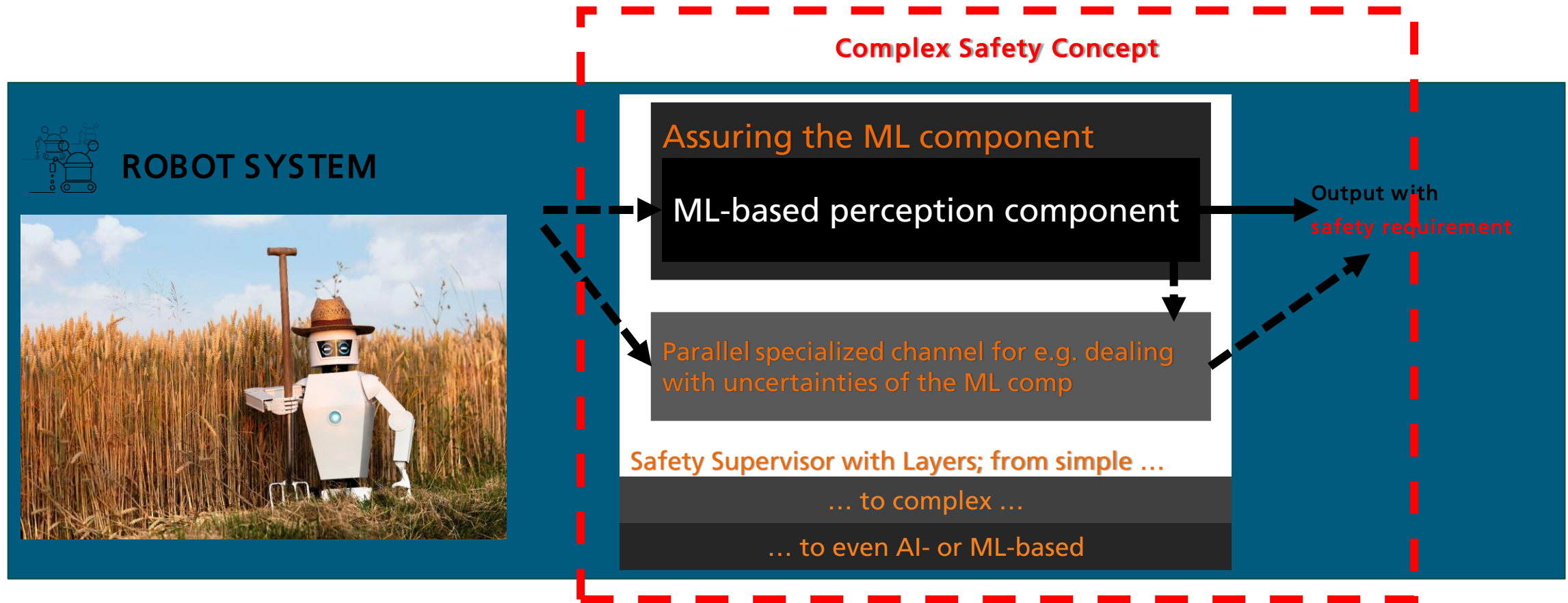
ANN-based dynamic risk assessment

E.g. resembling an experienced human driver wrt the capability to identify potentially dangerous situations very early.



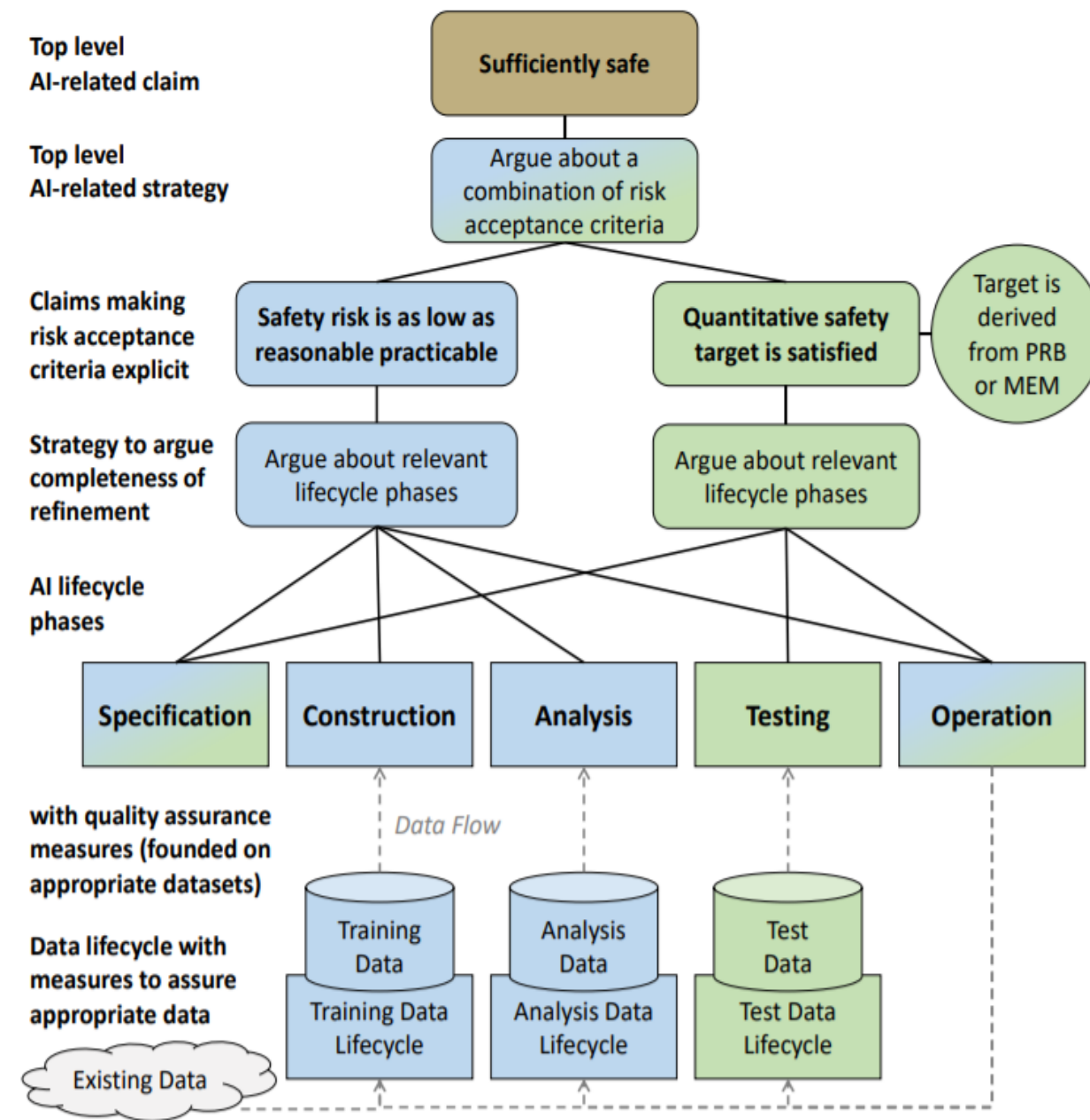
Fig. 4. Left: Uncritical scene (0) — Right: Critical scene (0.21)

Complexity requires systematic approach



ML Risk Acceptance Criteria and ML Assurance Case

- ML assurance approaches offer only partial answers, varying and incomplete coverage
- Top-level argument based on arguing over combined risk acceptance criteria
- Challenge: Traditional systems focus on As Low As Reasonably Practicable (ALARP) risk acceptance criterion
 - Likely not sufficient for ML
- Solution: Combine with
 - Positive Risk Balance (PRB): risk of hazard < benefit of application, or,
 - Minimum Endogenous Mortality (MEM): risk of death by hazards
- Argue over system and ML development lifecycle
- UWs can be applied to bound uncertainty given confidence levels e.g. during Testing



Contact

Dr. Daniel Schneider
Division Dependable Systems
Tel. +49 151 649 530 70
daniel.schneider@iese.fraunhofer.de

Fraunhofer IESE
Fraunhofer-Platz 1
67661 Kaiserslautern
www.fraunhofer.de

Thank you for your attention!
