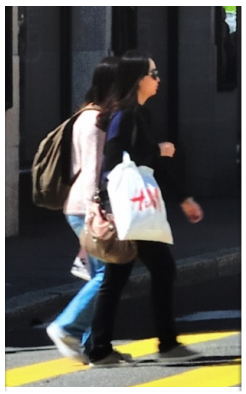
The GAP Between AI Development and Automotive Safety Criteria

Christian Müller

Background

The complexity of urban scenarios no longer permits classical, rule-based programming of algorithms, especially for scene understanding.
AI-based perception functions are required.

Intentions

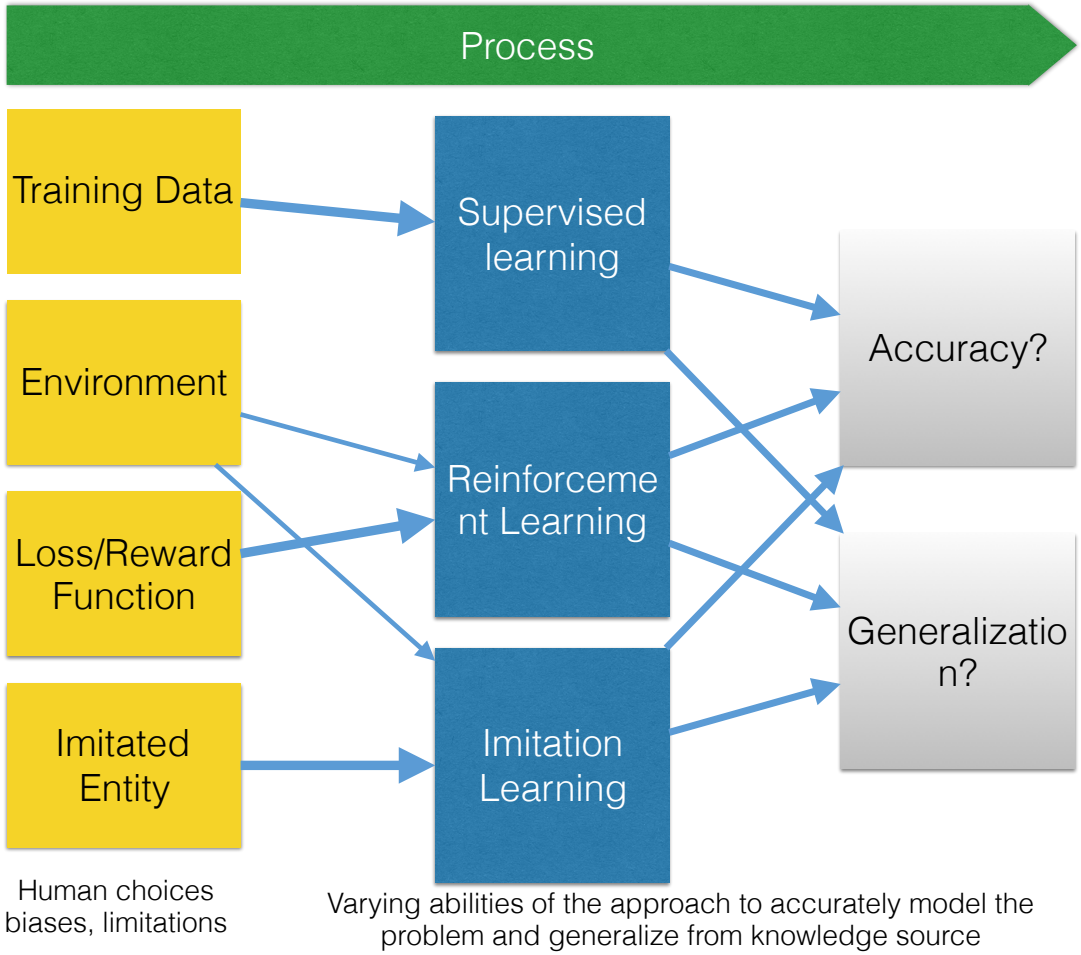


Social Connections

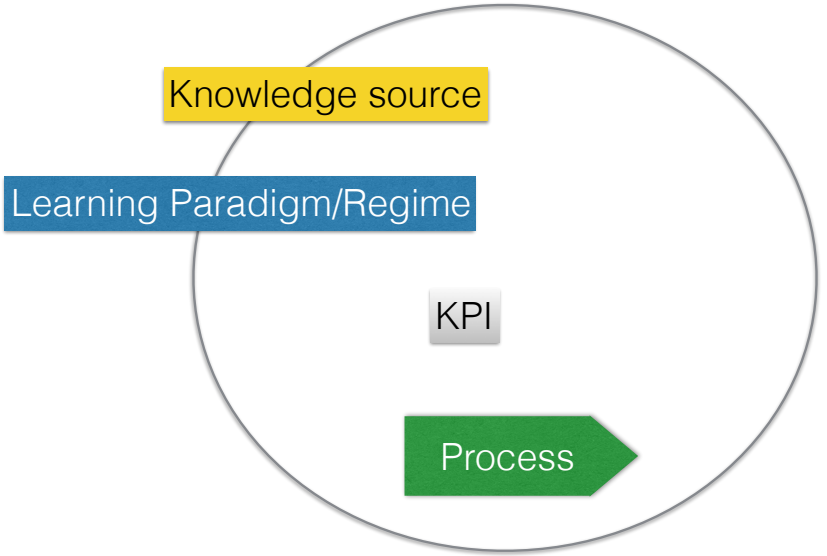
Interests

Swarms

Types of AI — Basic Validation Requirements

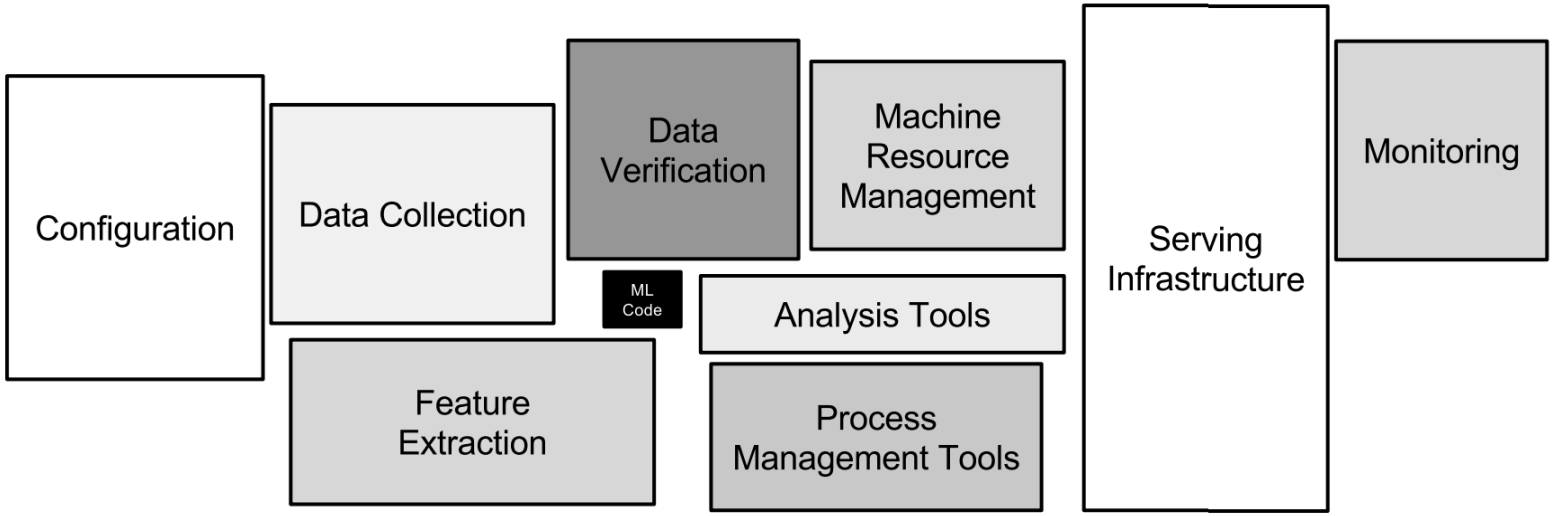


We need to validate



Comments like this on relevance for OpenGenesis will appear

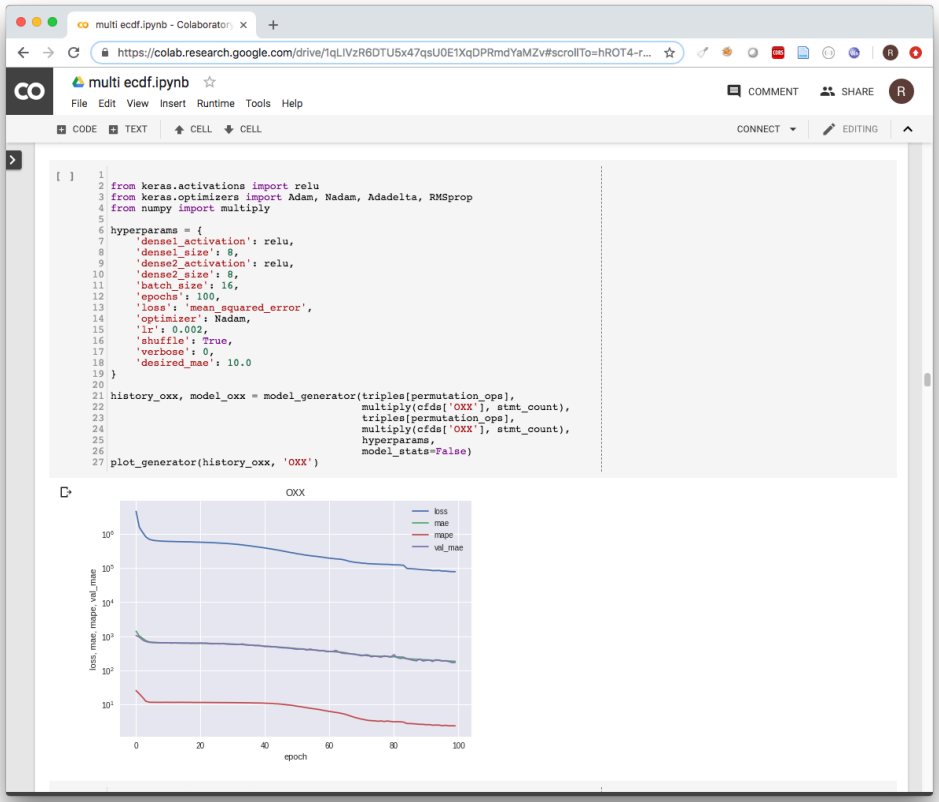
Improving the Process AI Platform

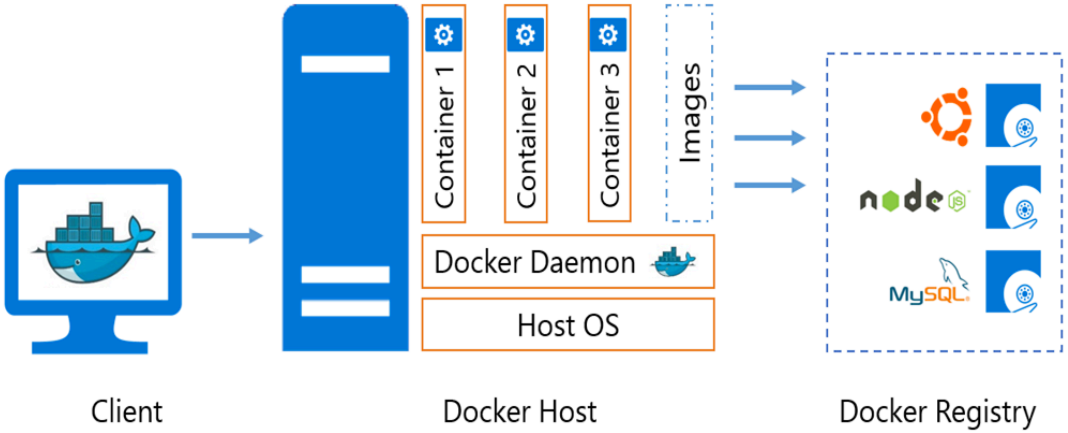


„Only a small fraction of real-world ML systems is composed of the ML code“

[Sculley et al, 2015]

Improving the Process AI Platform





Currently driven forward by German car industry and research institutes (VDA, VW, Bosch, DLR, Fraunhofer, DFKI) in multi-million EUR VDA lead initiative. (VDA = Association of the German automotive industry).

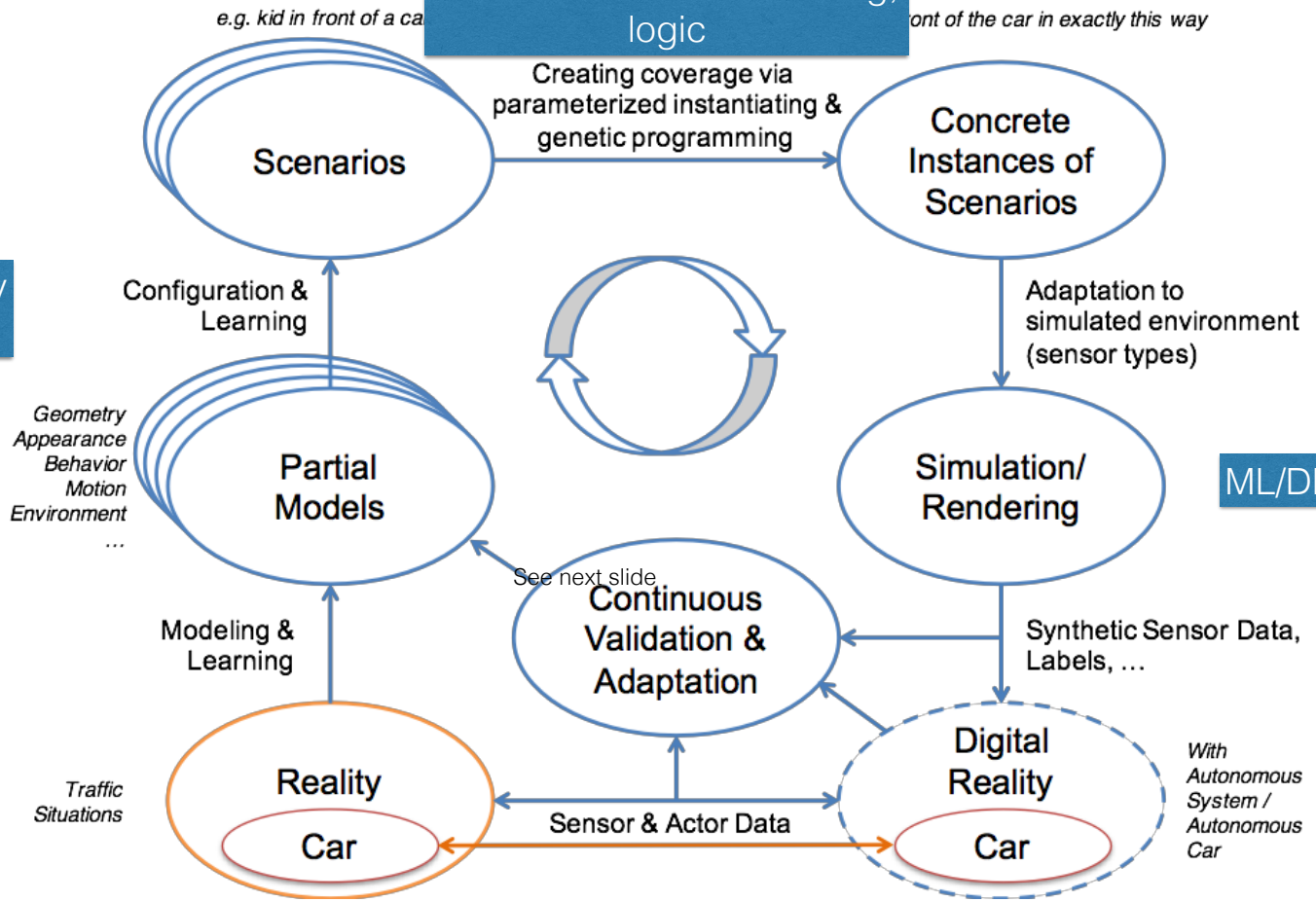
Area for OpenGenesis? I don't think so. Rather OpenGenesis uses an AI platform

Validation in a Digital Reality Scenario

Knowledge source

Constraint-based reasoning, logic

Multi-Agent technology tooling



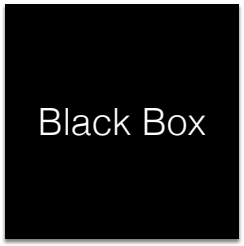
ML/DL based AI technologies

Dedicated novel ML/DL algorithms learned from Data

Area for OpenGenesis? Definitely, I believe.

Black Box, Grey Box, White Box

Learning Paradigm/Regime



We can test the AI module but we **cannot look inside** because a) we are not allowed b) we don't have competences c) the module does not reveal such insights — it is uninterpretable (like today's DNNs)

OEMs expect this situation

We can obtain **additional information about the behavior** of the model and its inner logic, i.e. by debugging it. We have **access** to the **inside** of the module and both the competencies and rights to derive such information

Important for reasonability

Goal

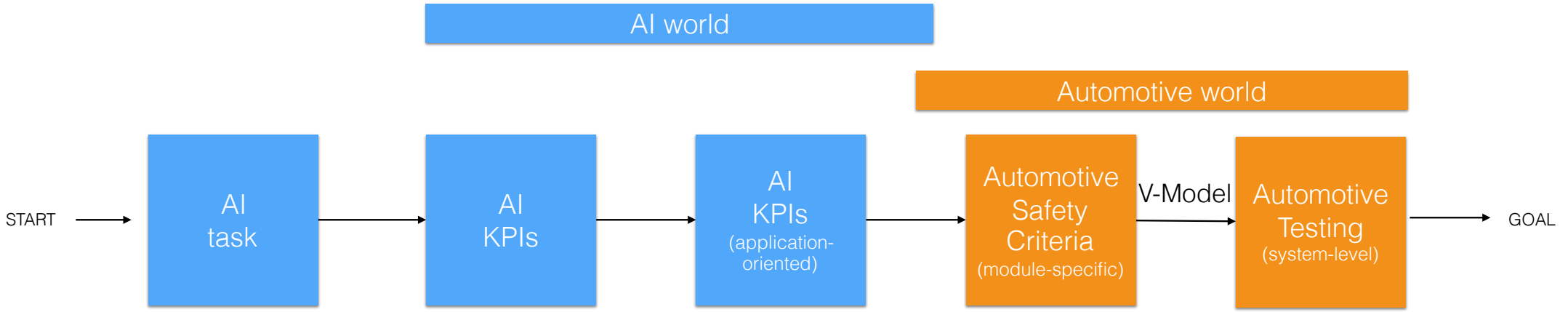
We **fully understand** the behavior of the module like with a human-engineered function. Not possible with todays DL regimes; to some extend with symbolic AI. Scientists works in this direction: hybrid learning seems promising. It is unlikely that in practice, boxes will ever be pure white (lack of competencies in testing facilities, residual inexplicable aspects of the model).

Scientific Utopia

A goal … also for OpenGenesis

AI and Automotive from a Validation Perspektive

KPI



e.g. frame-by-frame semantic segmentation of camera input

e.g. mean intersection over union

Mainly for scientific publications
Used in challenges, therefore popular

e.g. #of objects robustly recognized

Still for scientific publications
(+) more targeted
(-) less general
therefore rather rare

e.g. all vulnerable road users possibly colliding with ego vehicle recognized early enough

Functional safety engineering of individual or combined functions (e.g. fused environment perception)

e.g. for 99.99% of the cases collision with vulnerable road users can be ruled out.

Homologization, type and instance certification. Large-scale testing
e.g. EuroNCAP

ISO26262

SOTIF

Contract-Based Design

If this is not THE area for OpenGenesis, I wouldn't know what else would be.

Validation Methods — a Human Analogy

Neurological experiment

Debugging, AI tools

Systematic experiment as in cognitive psychology

Systematic test in digital (virtual) environments. See next slide(s)

Anomaly detection as in psychological expert reporting

Large-scale system level testing in real environments

All three „disciplines“ are necessary and deliver valuable insights

Area for OpenGenesis? Tools, systematic experiments: yes, large-scale testing: maybe supporting to some extend

Standards: ISO 262626

- Application of ISO262626 „Functional Safety“ used to support the necessary argumentation regarding the absence of an inappropriate risk **due to failure of the system**.

- It regards approaches for avoiding **systematic hardware and software errors** as well as random hardware failures.

- Applying only isolate26262 on the development processes for autonomous driving, in particular in application of AI-based functions, **cannot guarantee the desired level of safety**.

Standards: SOTIF

- A first attempt towards a new industrial consensus in the area of driver assistance systems (ISO/PAS 21448 „Safety of the Intended Functionality“) is supposed to minimize an inappropriate risk in relation to any danger, **i.e. through limitations of the system.**

- However, the analysis SOFTIF foresees, are only to a **limited extend applicable to AI-based functions**. Further measures are necessary.

- Furthermore, SOTIF is **applicable only to ADAS up to automation level SAE 2**. For higher automation levels, extensions are necessary.

Equivalence-Classes in the Input Space

- Promising approach: systematically identify system-critical influencing factors of the **input space**. (here: scenarios)
- Then, those factors could be analyzed and reduced with known methods and tools.
- On important step is to **summarize the input space in equivalence-classes**.

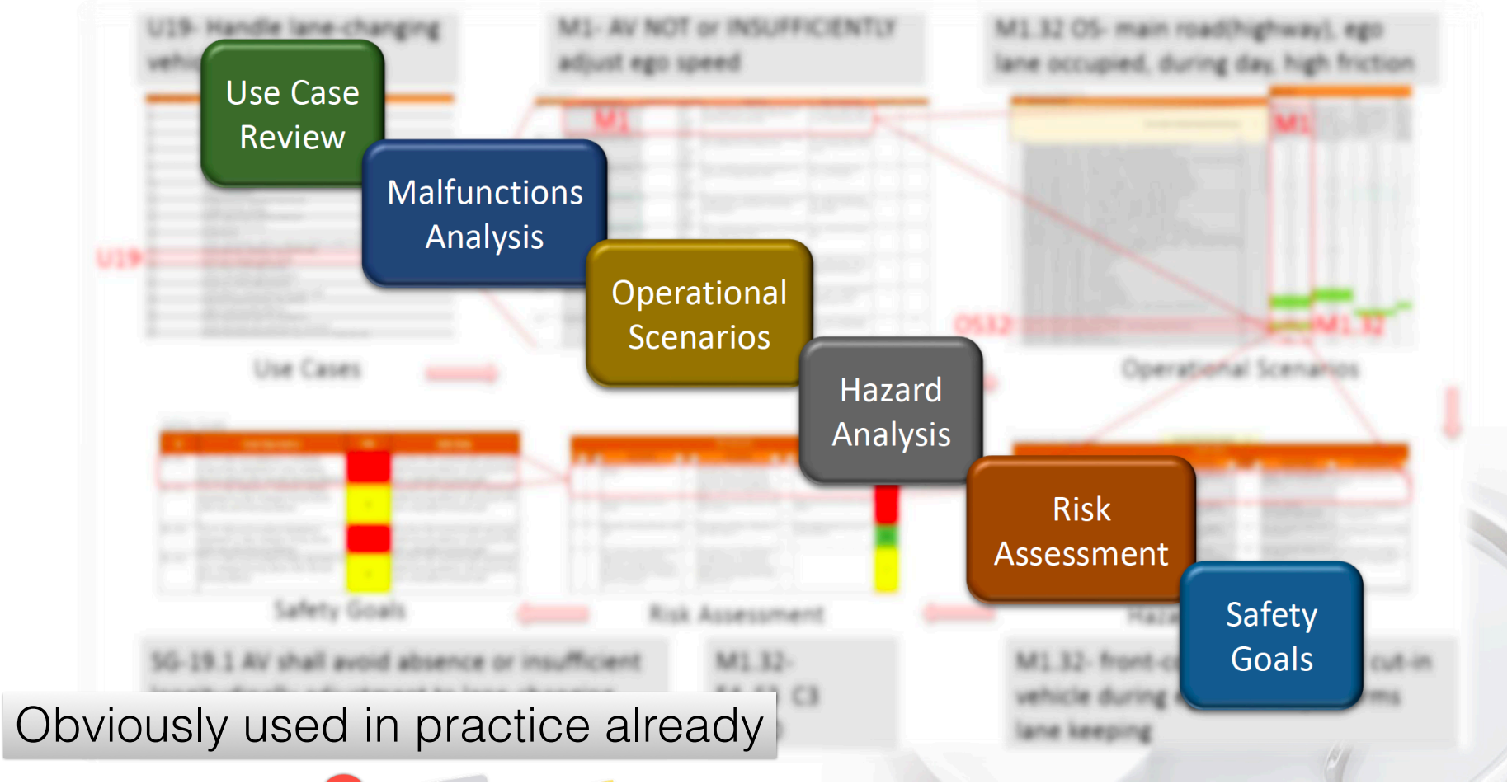
Contract-Based Design (CBD)

- CBD specification: guarantee for the result of a function under the assumption that the **input fulfills its specification**
- can also be used to specify that the use of an AI-based function in a **defined environment („assumption“) leads to a defined result („guarantee“)**
- Guarantees can be derived **according to safety criteria**
- Based on a **domain analysis** and characteristics of the system architecture, **assumptions can be derived.**
- For a given function, which fulfills this safety-„contract“, we can now validate **whether it satisfies the safety goals**.

Open Questions Regarding CBD

- How can we conduction a systematic domain analysis for a given function?
- How can we derive assumptions?
- How can we describe safety goals as guarantees?
- How can they be verified with a systematic test strategy?
  - In particular: are the KPIs valid with respect to assumptions and guarantees?
  - A KPU is valid if it allows implications about an AI-based function in the planned usage scope.
-

GWM – i-Pilot 1.0 – Safety Design



Obviously used in practice already

Conclusions

- For trustworthy AI, we need to consider the knowledge source, the learning paradigm, the KPIs (respectively safety criteria) and the process.
- A well-designed AI platform is important to improve the processes and thus the overall quality of the outcome
- Progress in AI towards whitening the box is essential. Together with semantic deep learning, hybrid learning is a very promising direction in research (also at iMotion Germany).
- We need to find ways to better translate AI KPIs into functional safety requirements
- Systematic experiments are also important, not only large-scale testing (analogy to Human)
- Contract-based validation has the potential to be the missing link at the end of the KPI „chain“.
- Announcement: Kick-off of OpenGenesis project with TÜV Süd, DFKI, others, April 30th