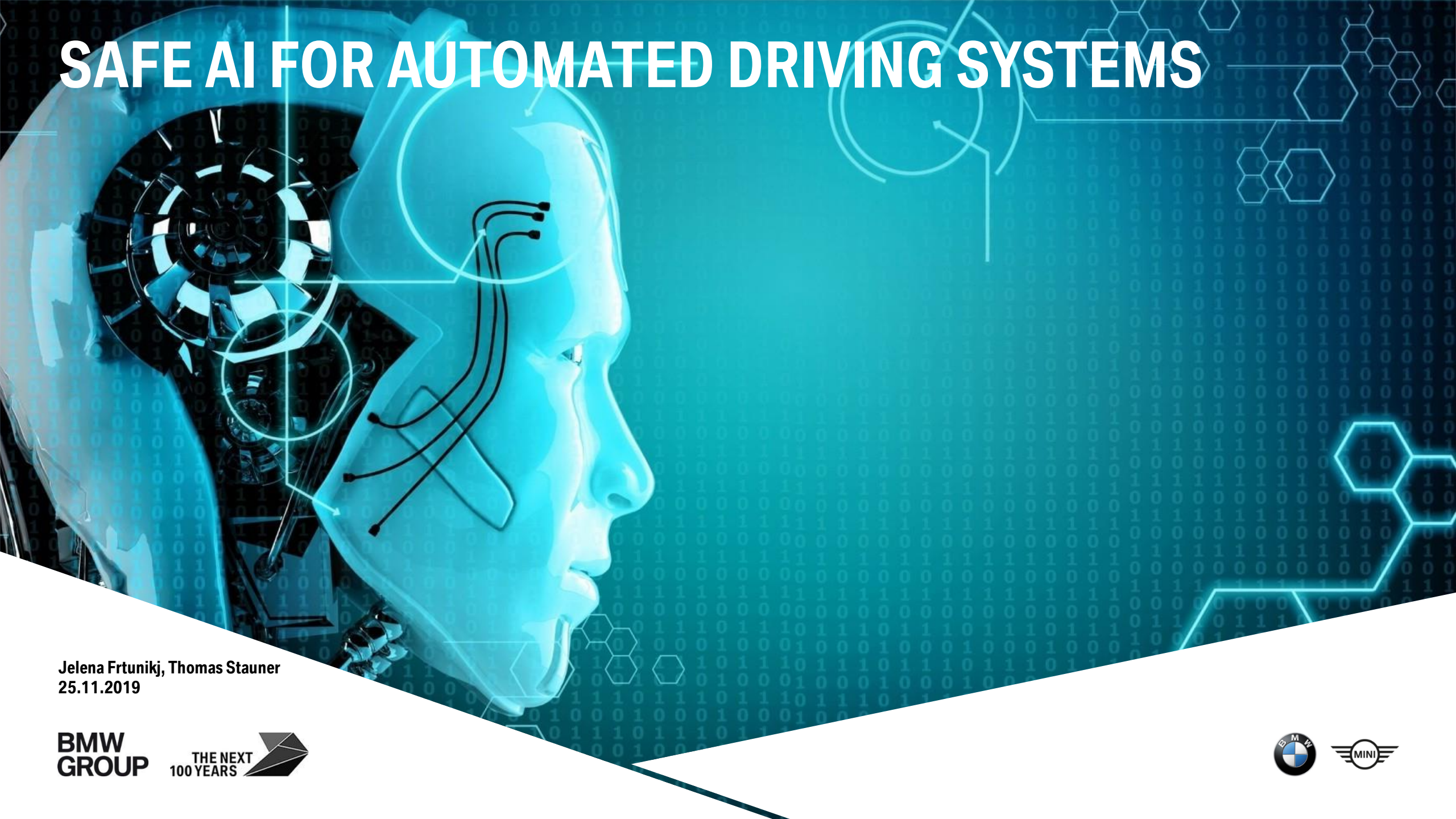
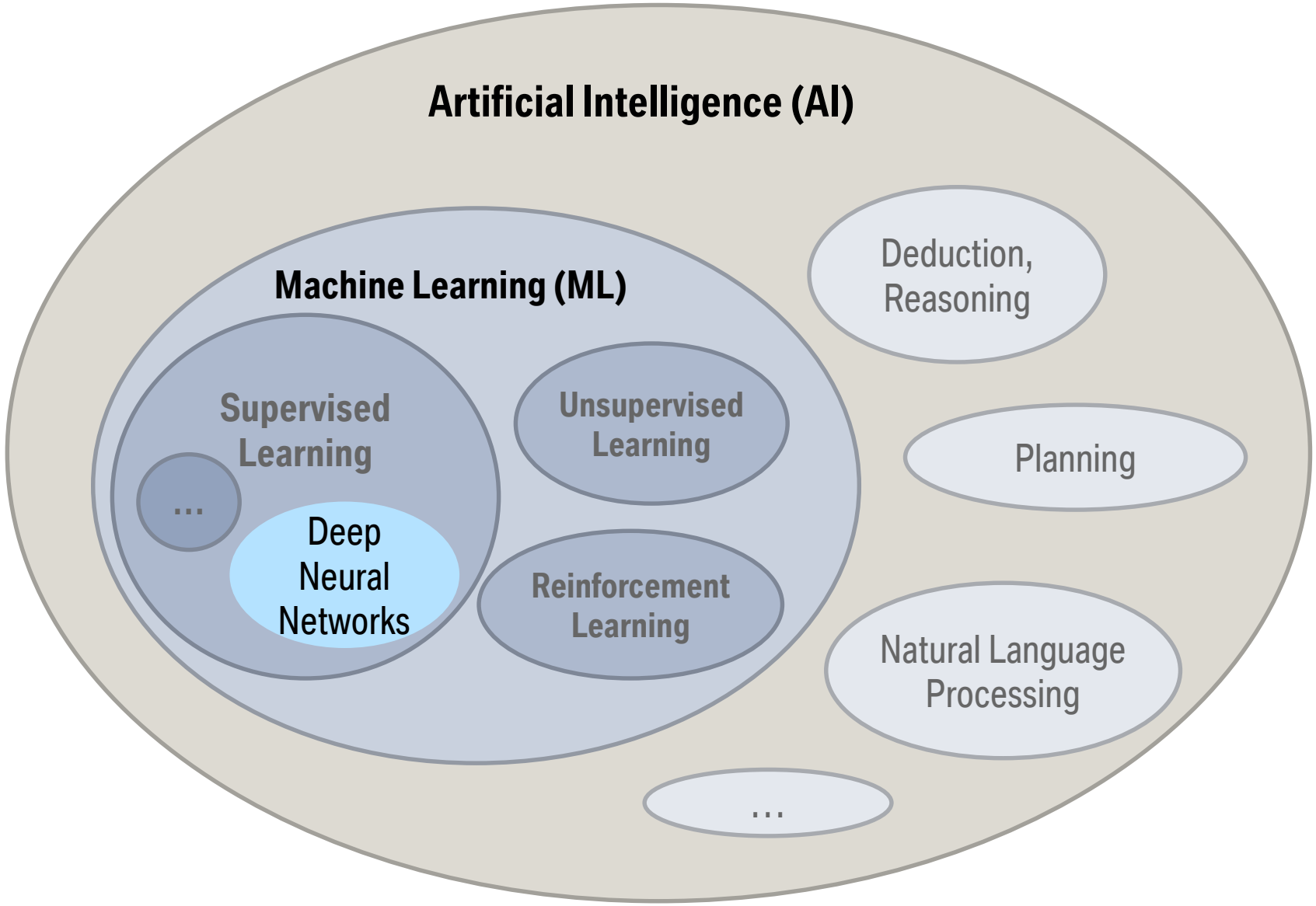
SAFE AI FOR AUTOMATED DRIVING SYSTEMS



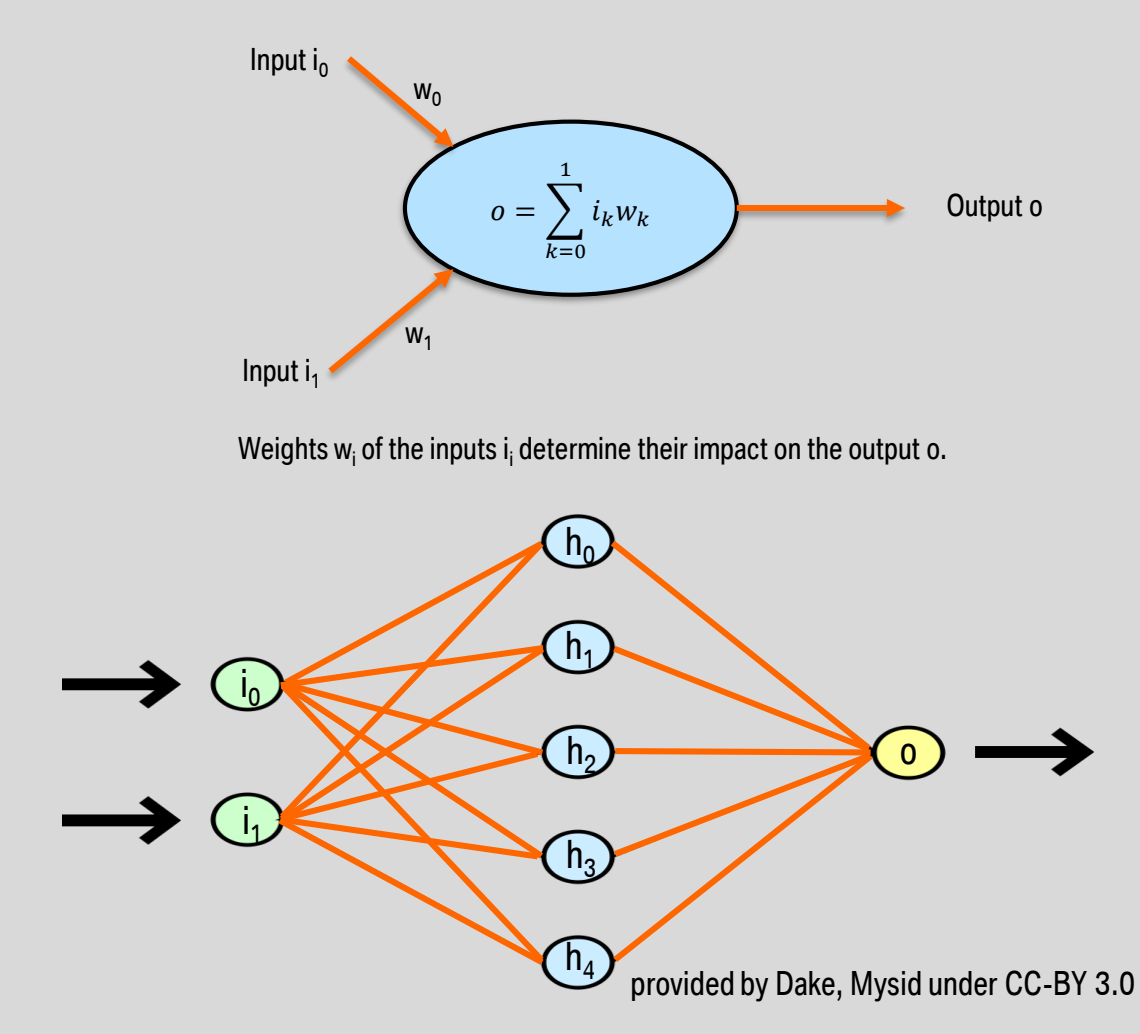
Jelena Frtunikj, Thomas Stauner
25.11.2019

BMW GROUP THE NEXT 100 YEARS

AI AND ML TERMS AND DEFINITIONS



ARTIFICIAL NEURAL NETWORKS

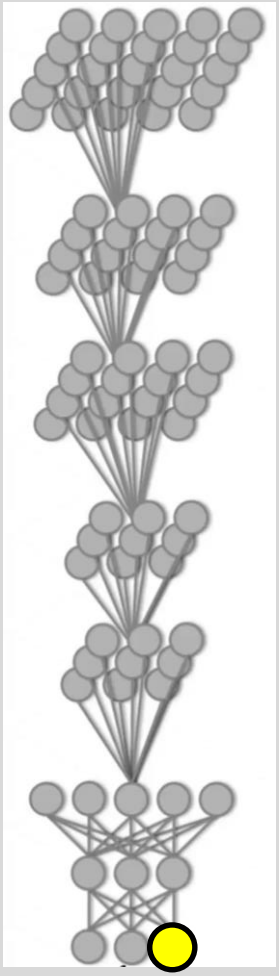


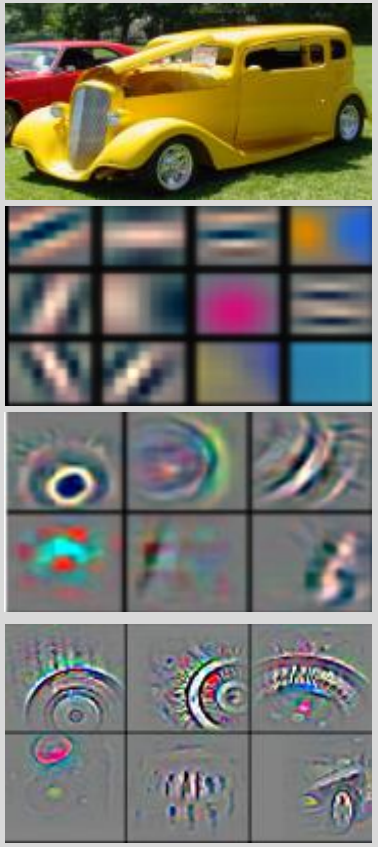
- Analysis is primarily by **matrix multiplications.**

- Weights have to be "learned", they are not pre-defined:
  - Default values for i_0, i_1, o – **training data**
  - Minimization of the error across all values – **back propagation**

- The first models of artificial neural networks were described as long ago as the 1940s, **major break-through** in the **last 5 years.**

DEEP LEARNING
NEURONAL NETWORKS WITH MANY HIERARCHICAL LAYERS





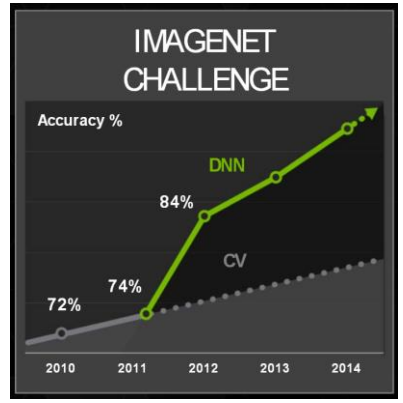
„What is this?“

„A car!“

Source: Geoff Hinton et al. (2015) Deep Learning NIPS‘2015 Tutorial

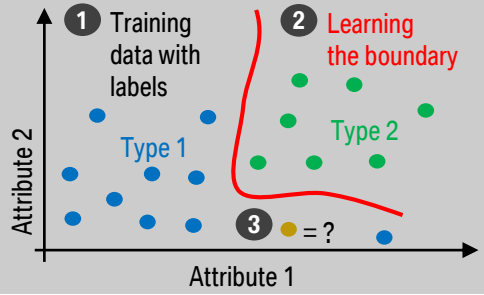
Idea of the 1980th, current success due to

- new **algorithms** for training of networks,
- availability of huge amount of data as **training sets** (e.g. image databases), and
- high performance, parallel **hardware accelerators** (GPUs, TPUs).

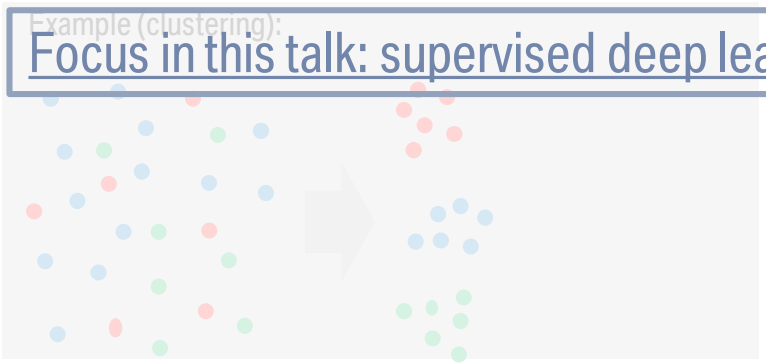


TYPICAL METHODS OF MACHINE LEARNING

Example (classification):



Example (clustering):

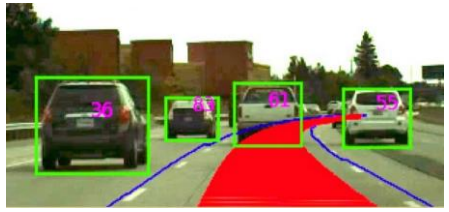


Focus in this talk: supervised deep learning

AlphaGo

Supervised learning

- Learning from associations of data
- **Classification**: Distinction of discrete classes/types
- **Regression**: Forecasting of continuous functions



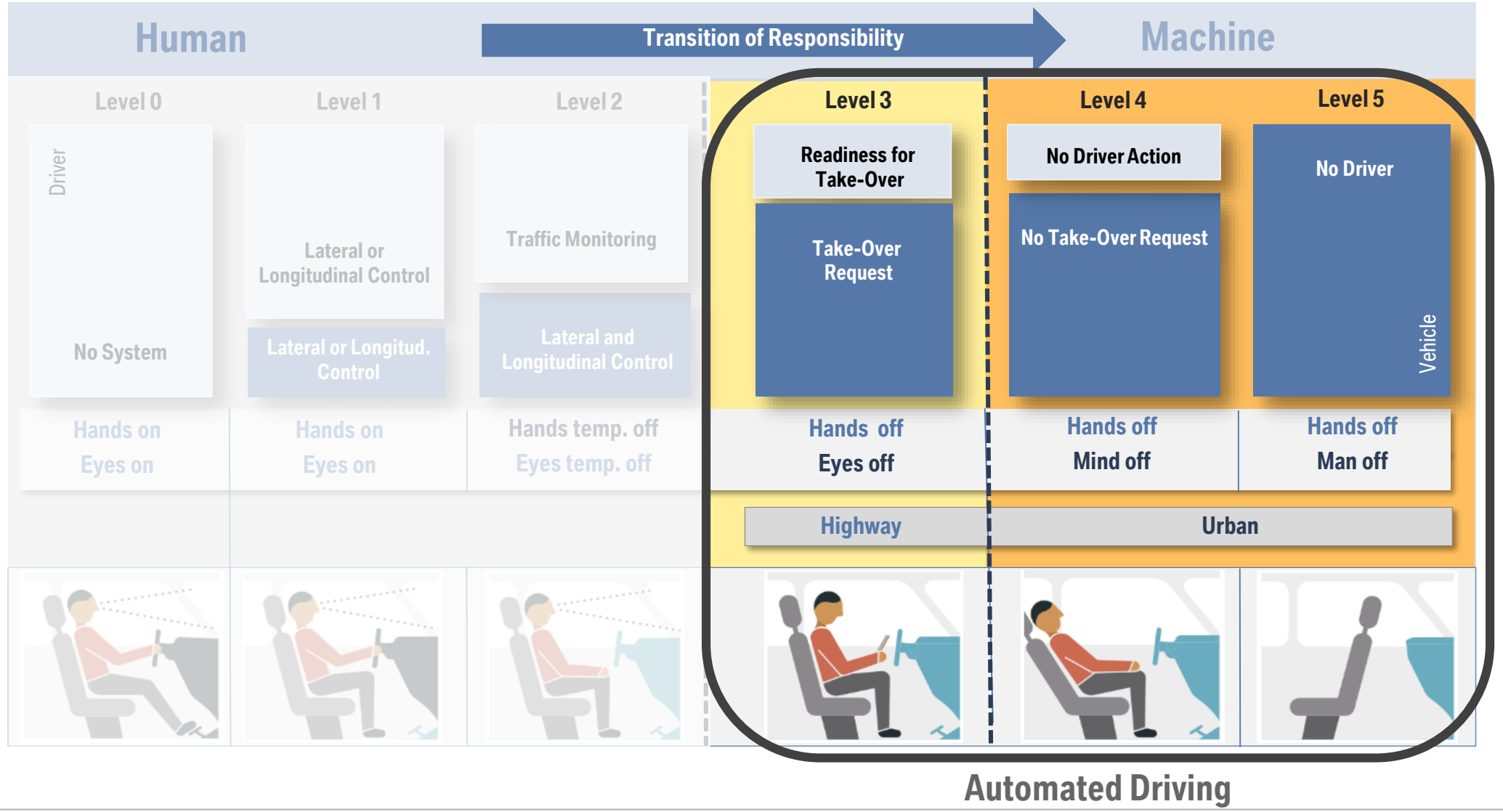
– Unsupervised learning

Example: Image segmentation

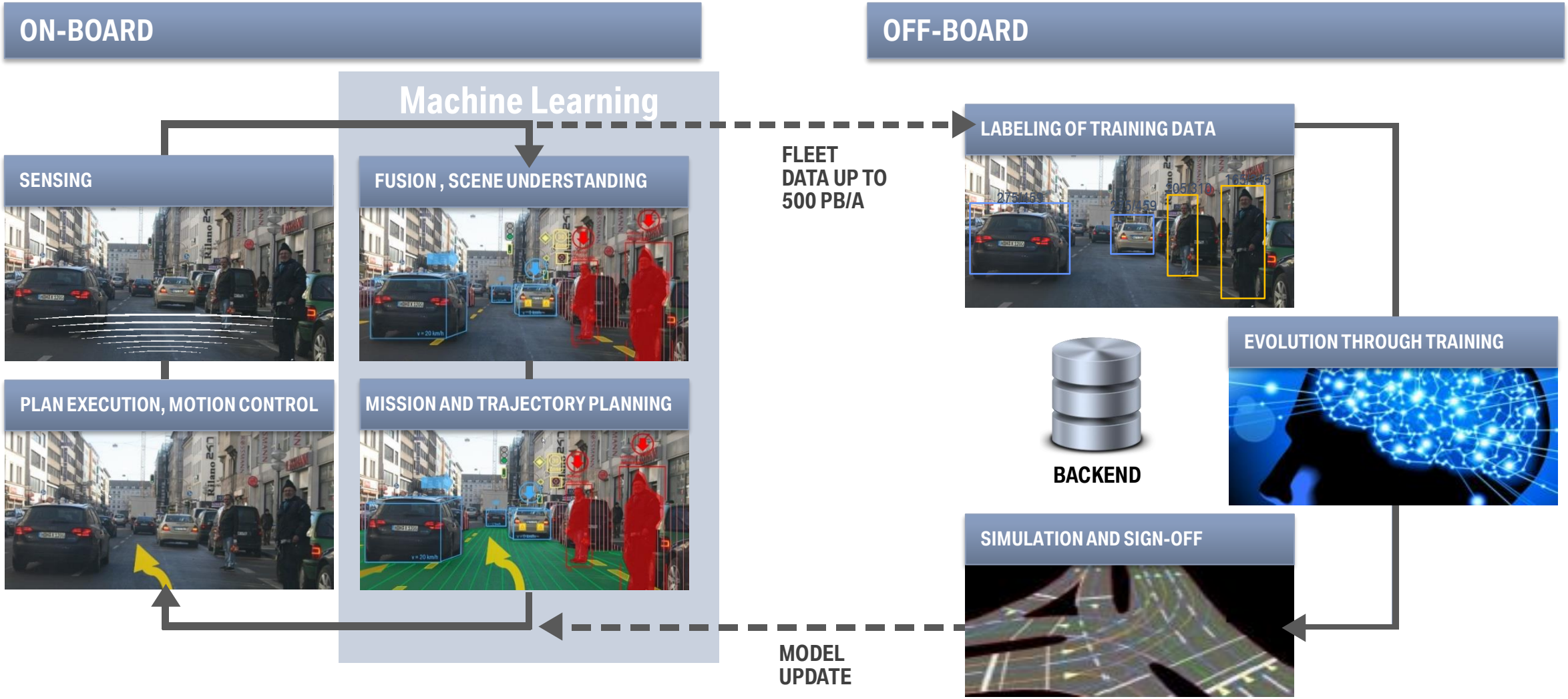
– Reinforcement learning

Example: Lane change maneuver

AUTOMATED DRIVING: FROM HANDS-ON TO MAN-OFF



ARTIFICIAL INTELLIGENCE IS THE KEY TO AUTOMATED DRIVING



SAFETY IS CRUCIAL FOR SUCCESS OF AUTONOMOUS DRIVING

– Shai Shalev-Shwartz, CTO MobilEye: "**Safety** and **Scalability** contain the risk of 'Winter of autonomous vehicles'." – similar to the "Winter of AI".[1]

– Classic safety standards are based of a „4 plus 1" principle:[2]

1. Define Safety requirements
2. Decompose safety requirements
3. Provide that the software satisfies the safety requirements
4. Identify and mitigate hazards from the software

– "The confidence established in addressing the software safety principles shall be commensurate to the contribution of the software to system risk."

– Additionally, process quality is required

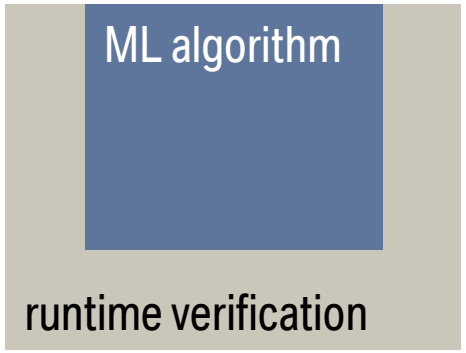
– Application to Machine Learning based algorithms:
  – Divide & conquer approach is not possible or not sensible
    – What would MCDC coverage mean for a DNN?

– Orthogonal problem: tool qualification of ML frameworks.

1 Shai Shalev-Schwartz et al.: On a Formal Model of Safe and Scalable Self-driving Cars, 2017.
2 The Safety of Autonomous Systems Working Group: Safety-Related Challenges for Autonomous Systems, 2018.

SOLUTION CLASSES: REFERENCE INFORMATION AVAILABLE

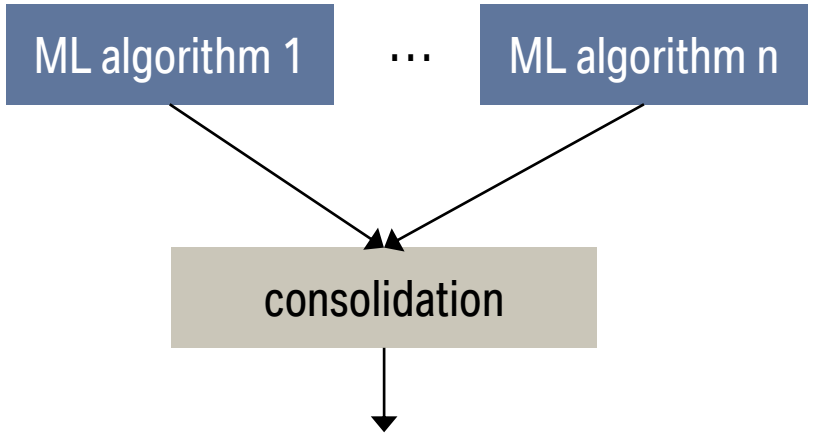
– Class 1: A formal reference exists against which correctness of the ML output can be established.



Example: Driving strategy
Formal reference: environment model data
Correctness criterion: e.g. keep distance

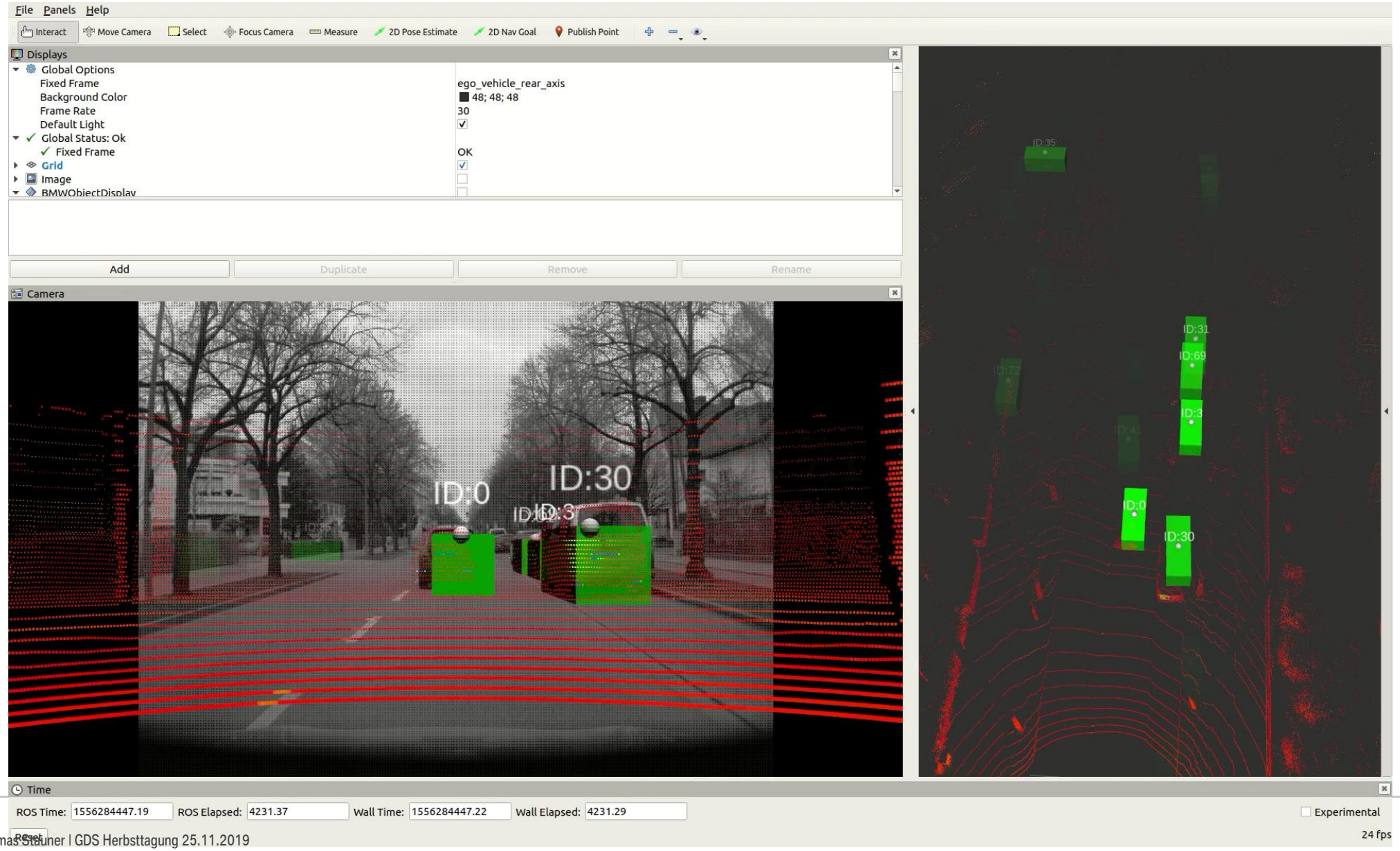
SOLUTION CLASSES: NO REFERENCE AVAILABLE

- Class 2: No formal reference available / input inherently unstructured
  - Means to guarantee reliability of ML solution itself are required
    - Redundancy of independent channels, to reduce required reliability of each individual channel
    - Exhaustive / statistically relevant testing
    - Process quality
    - Further measures see following example

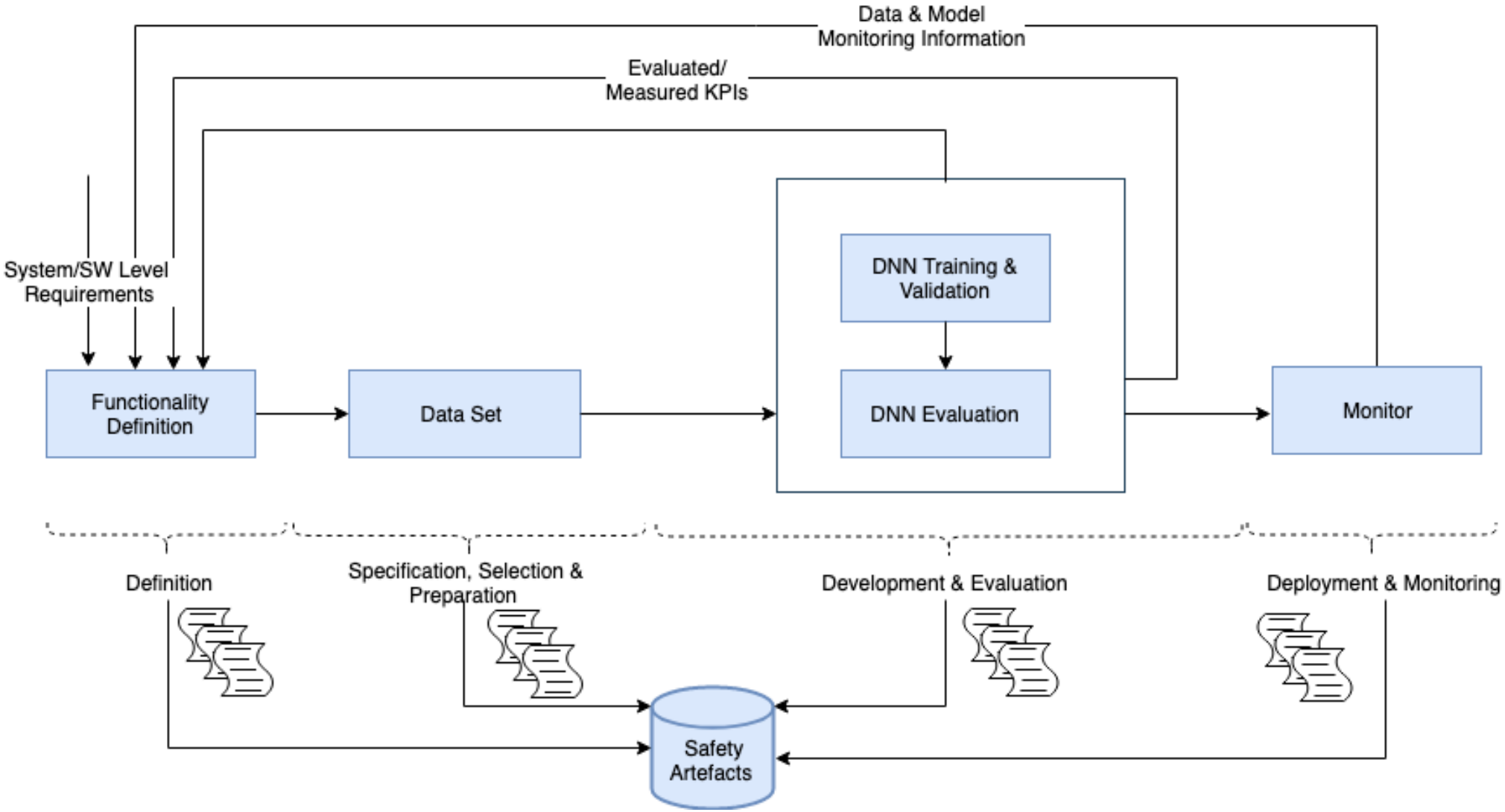


Example: sensor fusion, scene understanding.
Formal reference: -
Correctness criterion: correct detection

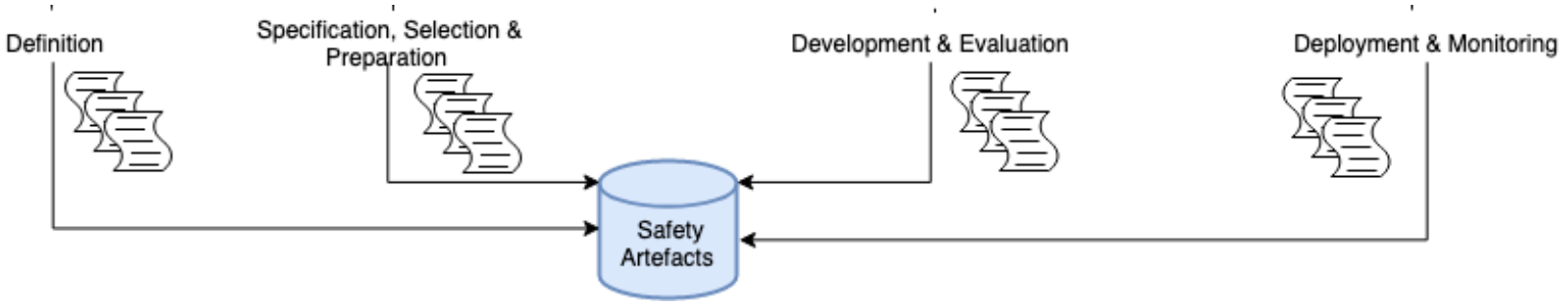
DEMO VIDEO: DEEP LEARNING BASED 3D OBJECT DETECTION



SAFETY ALONG THE ENTIRE DEEP LEARNING DEVELOPMENT CHAIN



EXEMPLARY SAFETY ARTIFACTS GENERATED DURING THE DNN DEVELOPMENT CHAIN



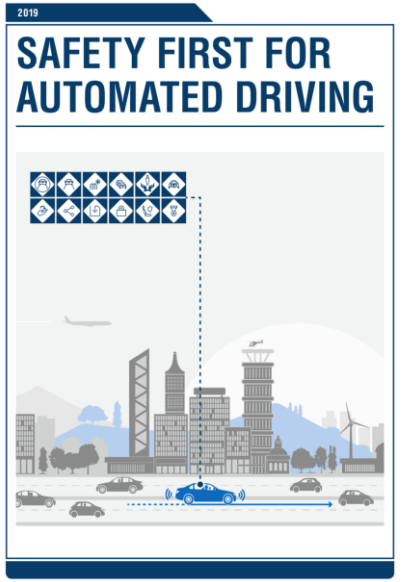
Definition	Data Set Specification, Selection, Preparation	Development & Evaluation	Deployment & Monitoring
Dataset Specification	Refined Dataset Specification	DNN Architecture	Detecting unseen situations
Labeling Specification	Refined Labeling Specification	Performance and Safety KPI Report	Performance and Safety KPI Monitoring
KPI Specification	Labeled Dataset	Data Set and Model Baseline	…
…	Dataset KPI Report	Software Code Baseline	
	Scripting tools	…	
	…		

SUMMARY & FUTURE WORK

- Deploying safe ML algorithm (incl. DNN) require mix of various safety measures along the whole development chain

- Further research is required to understand the safety insufficiencies of ML and identify counter-measures
  - ML algorithms should not be considered as black boxes, but their intrinsic properties should be used
  - Establish new methods and measures

- Extend safety standards to cover ML aspects



https://www.press.bmwgroup.com/global/article/attachment/T0298103EN/434404

THANK YOU FOR YOUR ATTENTION.
ANY QUESTIONS?

Jelena Frtunikj, Thomas Stauner

BMW GROUP THE NEXT 100 YEARS