Digital Reality: From Learned Models via Synthetic Data to Trusted-AI – and Back!

Philipp Slusallek

German Research Center for Artificial Intelligence (DFKI)
Saarland University, Computer Graphics Lab (UdS-CGL)
NHR South-West (Method Labs: Compiler & AI)

Saarland Informatics Campus
slusallek@dfki.de

German Research Center for Artificial Intelligence (DFKI)

- **Overview**
  - Largest independent AI research center worldwide (founded in 1988)
  - Germany’s leading research center for innovative software technologies
  - 6 main sites in Germany
    - Saarbrücken, Kaiserslautern, Bremen, Osnabrück/Oldenburg
    - Berlin, (Darmstadt) + sites in Lübeck & Trier
  - 24 research areas, 9 competence centers, 8 demonstration centers (living labs)
  - Almost 1400 research staff & support
  - Revenues of >73 M€ in 2020 (58 M€ in 2019, 50 M€ in 2018)
  - More than 100 spin-offs





SIC Saarland Informatics Campus

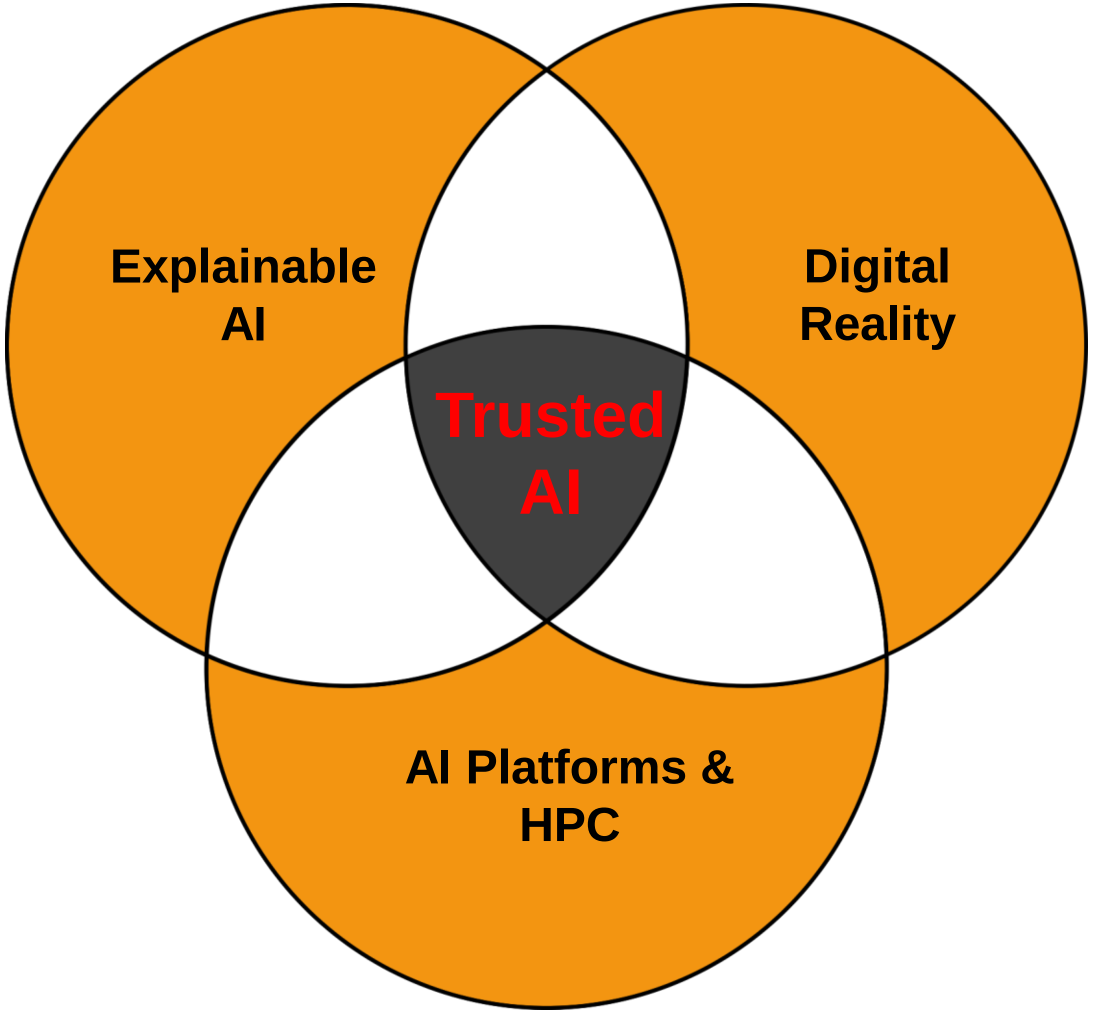
Germany Has a Head-Start

DFKI: The World´s Largest Center for Research & Application in AI

DFKI:
Covers the Complete Innovation Cycle

DFKI-ASR: Agents and Simulated Reality

How to design AI systems that can provide guarantees and that humans can understand and trust?



How can synthetic data from parametric models and simulations be used for *training, validating, and certifying AI systems?*

How can Trusted AI systems be realized technically in a *reliable and efficient way?*

Flexible Production Control Using Multiagent Systems at Saarstahl, Völklingen

DFKI multi-agent technology is running the steelworks, 24/7 for >12 years, 5 researchers transferred

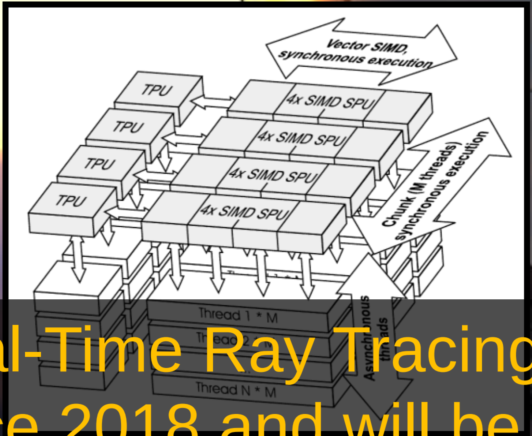
Physically-Based Image Synthesis with Real-Time Ray Tracing

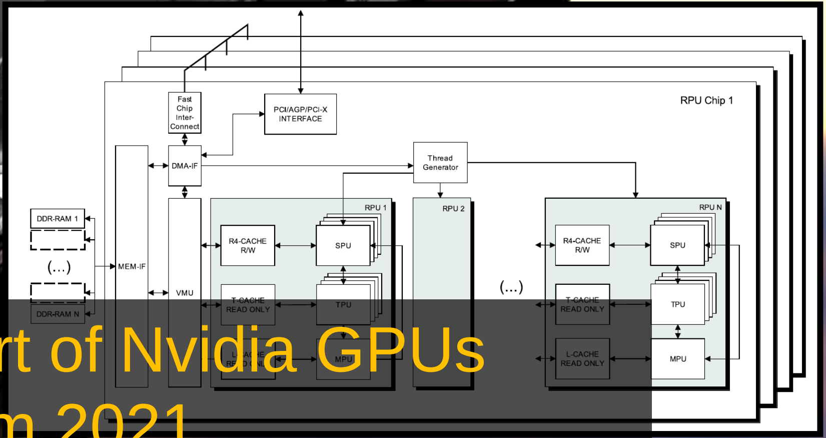
Technical Oscar, Feb 2021

Key product offered now by all major HW vendors:
e.g. Intel (Embree), Nvidia (OptiX), AMD (Radeon Rays) , …

Custom Ray Tracing Processor [Siggraph'05]

In EVERY GPU starting 2022



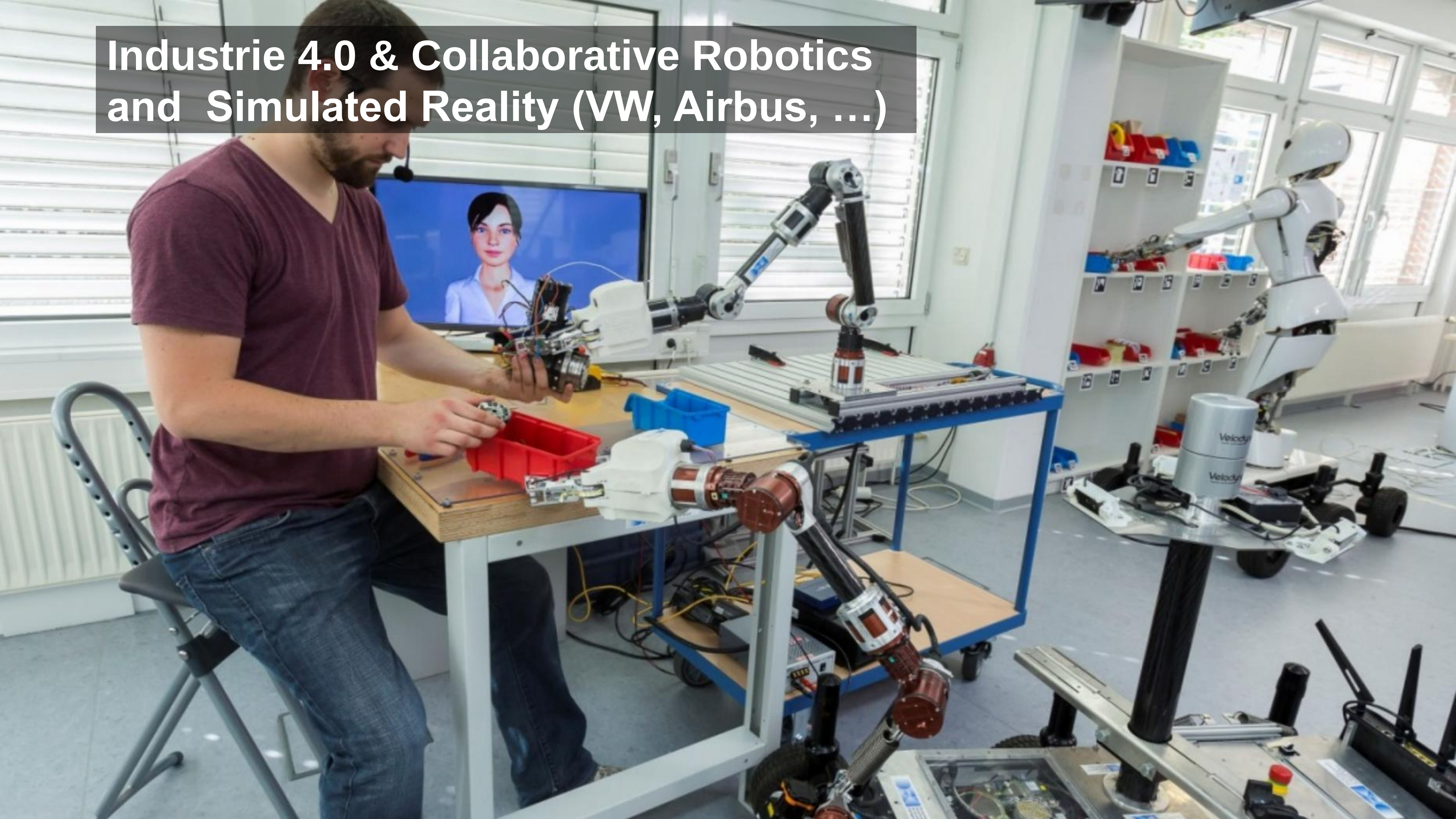


Real-Time Ray Tracing Hardware is part of Nvidia GPUs since 2018 and will be for all others from 2021

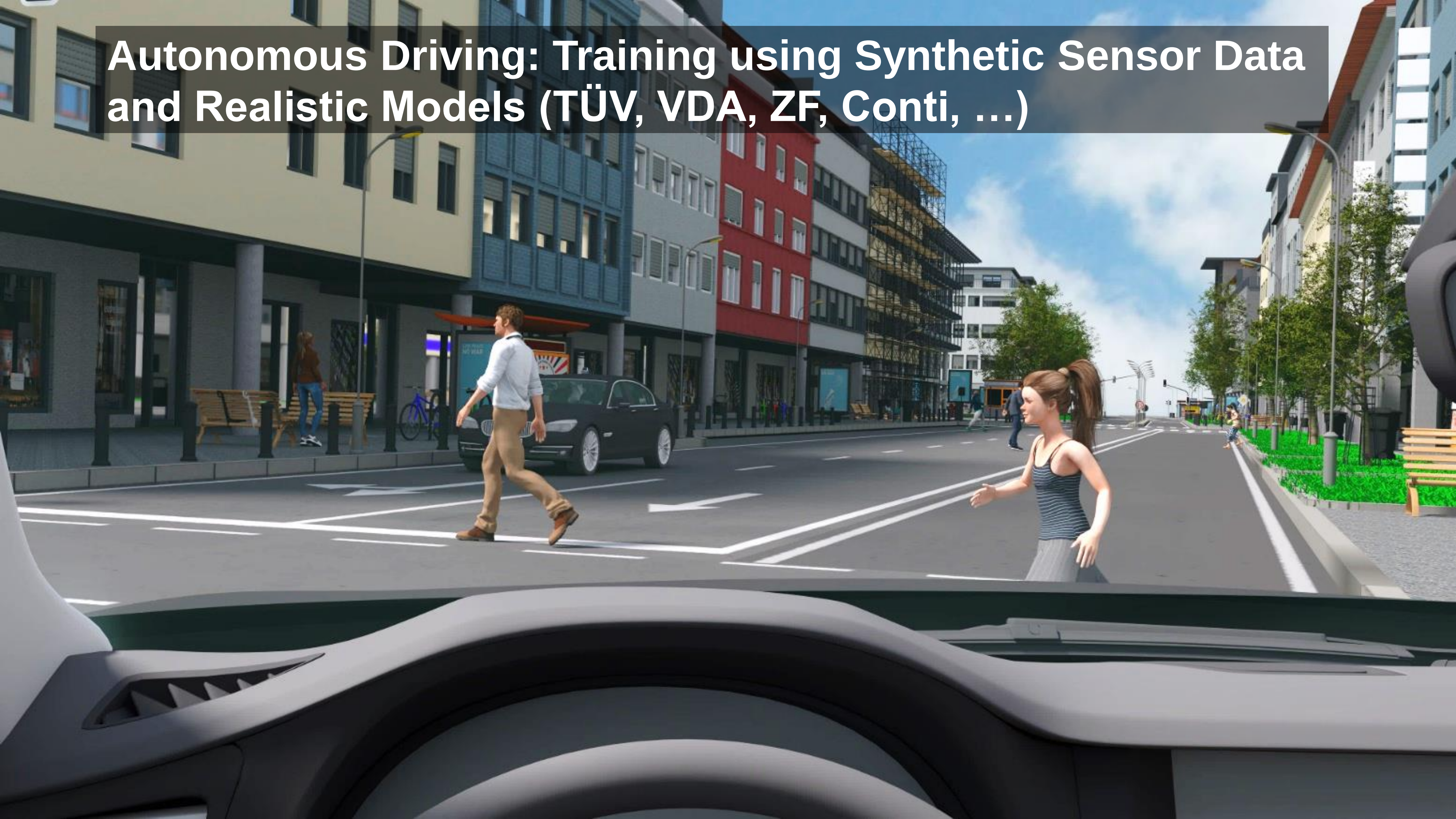
Efficient Simulation:
Radiation Propagation and Sensor Models

VCM now part of most commercial renders:
e.g. RenderMan, V-Ray, …

Industrie 4.0 & Collaborative Robotics and Simulated Reality (VW, Airbus, …)

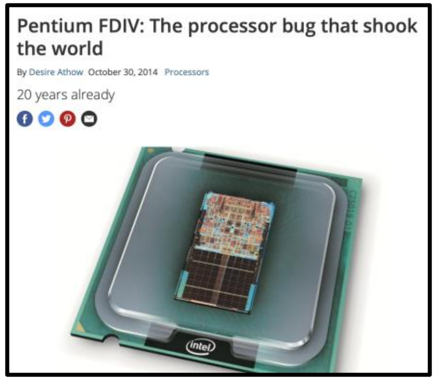


Autonomous Driving: Training using Synthetic Sensor Data and Realistic Models (TÜV, VDA, ZF, Conti, …)



State of AI

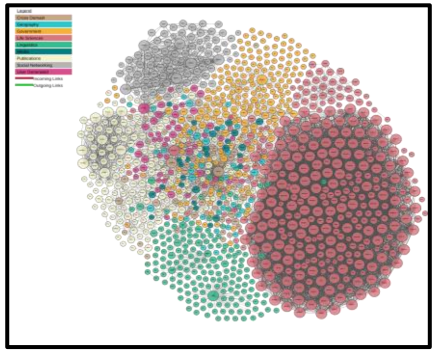
- **Many success stories**
  - HW Verification, Knowledge Graphs, Search & Optimization, …
  - Perception: Vision, Speech, …
  - Game playing: Chess, Go, video games, …
  - Some complex tasks: translation, autonomous driving, …
- **Amazing progress in recent years**
  - Most visible due to Deep Neural Networks (DNNs)
  - Focus shifting to hybrid/neuro-symbolic/neuro-explicit approaches
- **Still many fundamental challenges**
  - With severe consequences to the practical use of AI

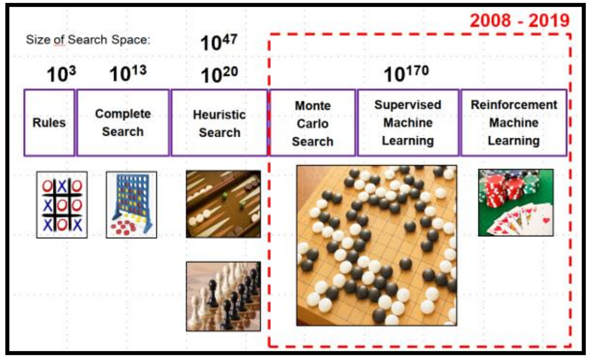


Pentium FDIV: The processor bug that shook the world

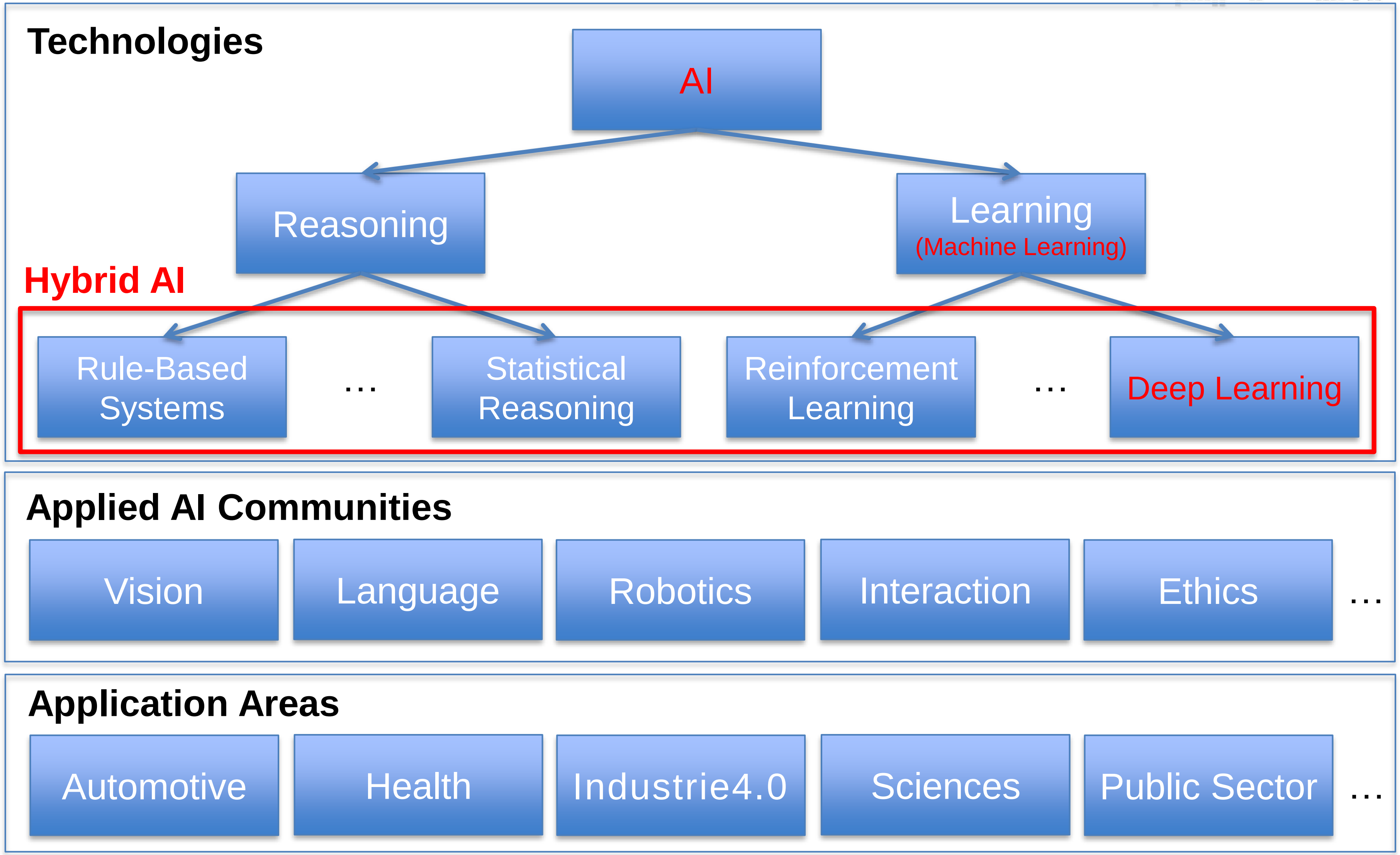
By Desire Athow October 30, 2014 Processors

20 years already

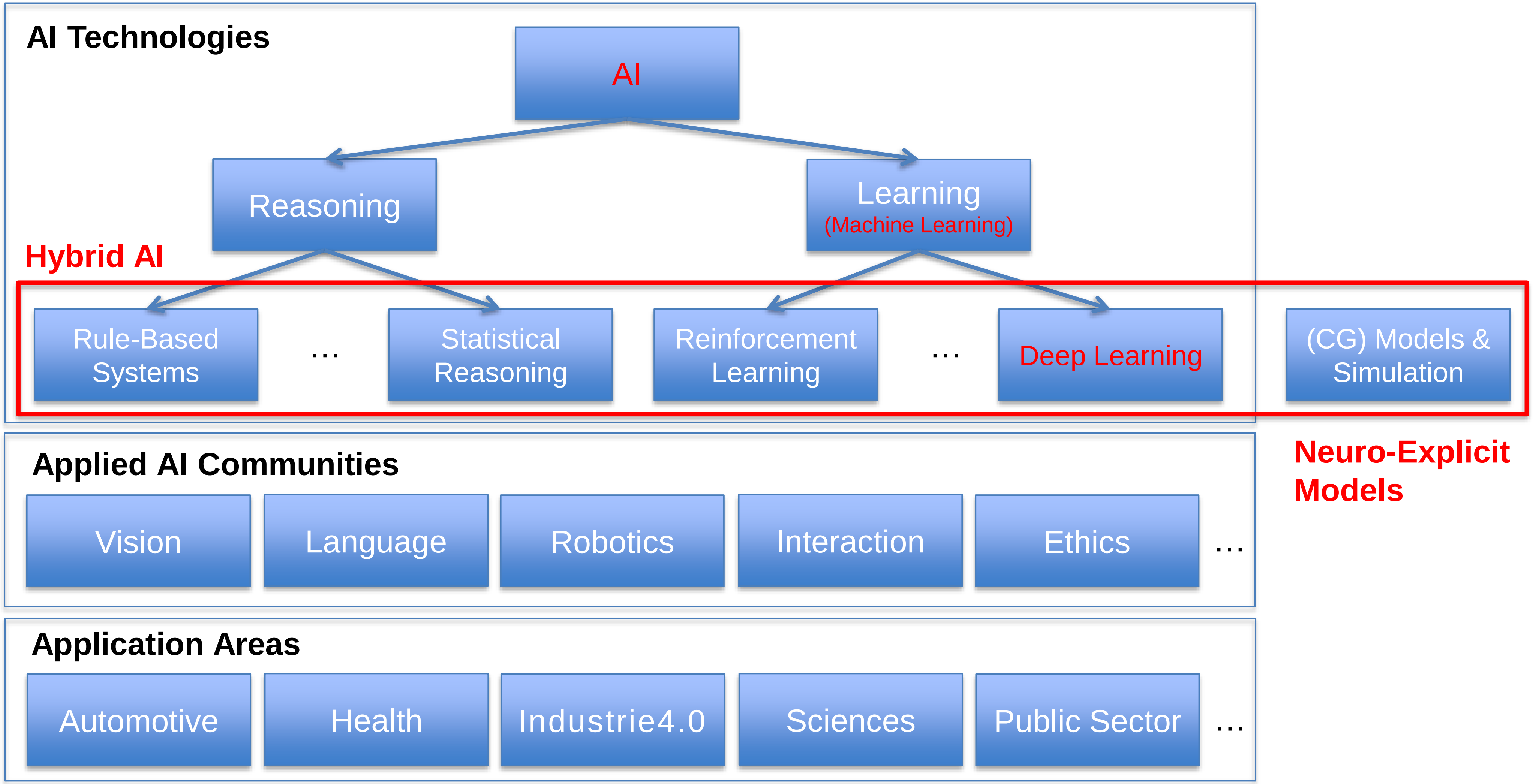




AI: Structure and Terminology

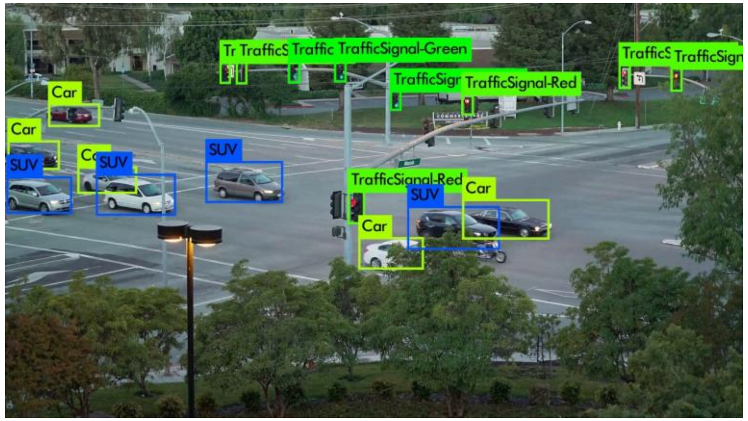


AI: Structure and Terminology

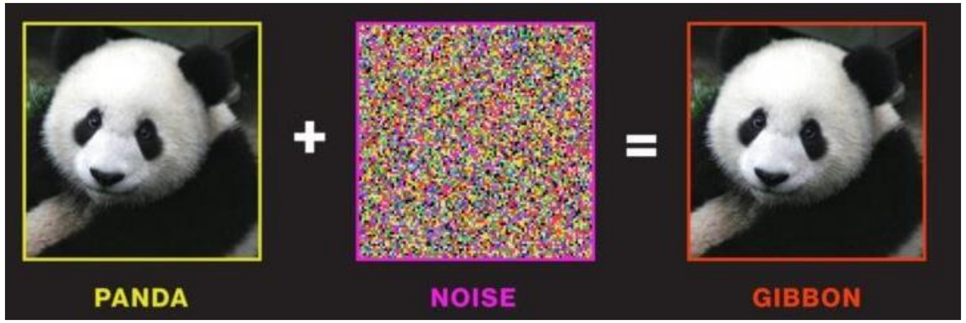


Challenges: Functionality vs. Robustness

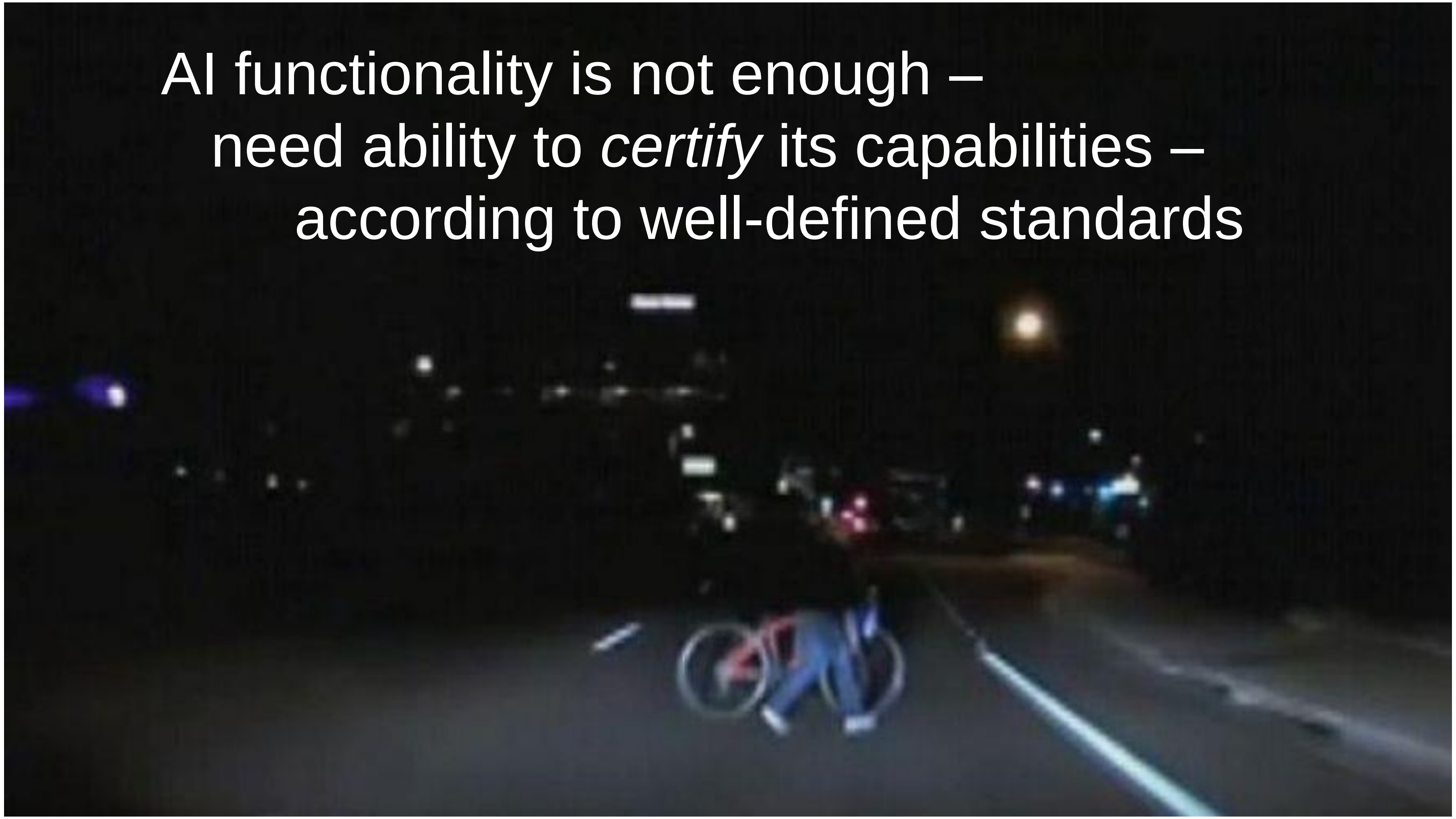
- AI/DL is highly capable already …



- … but we often cannot guarantee even basic functionality



Digital Reality:
Using AI to Optimize & Certify AI



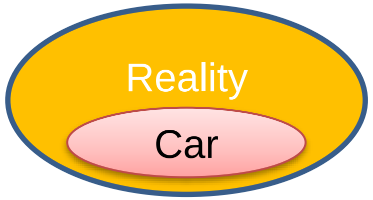
Autonomous Systems: The Problem

- **Our World is extremely complex**
  - Geometry/Shape, Appearance, Motion, Weather, Environment, …
- **Systems must make accurate and reliable decisions**
  - Especially in ***Critical Situations***
  - Increasingly making use of (deep) machine learning
- **Learning of critical situations is essentially impossible**
  - Often little (good) data even for "normal" situations
  - Critical situations rarely happen in reality – per definition!
  - Extremely high-dimensional models

➔ **Goal: Scalable Learning from *synthetic* input data**
  - Continuous benchmarking & validation ("Virtual Crash-Test")

Reality

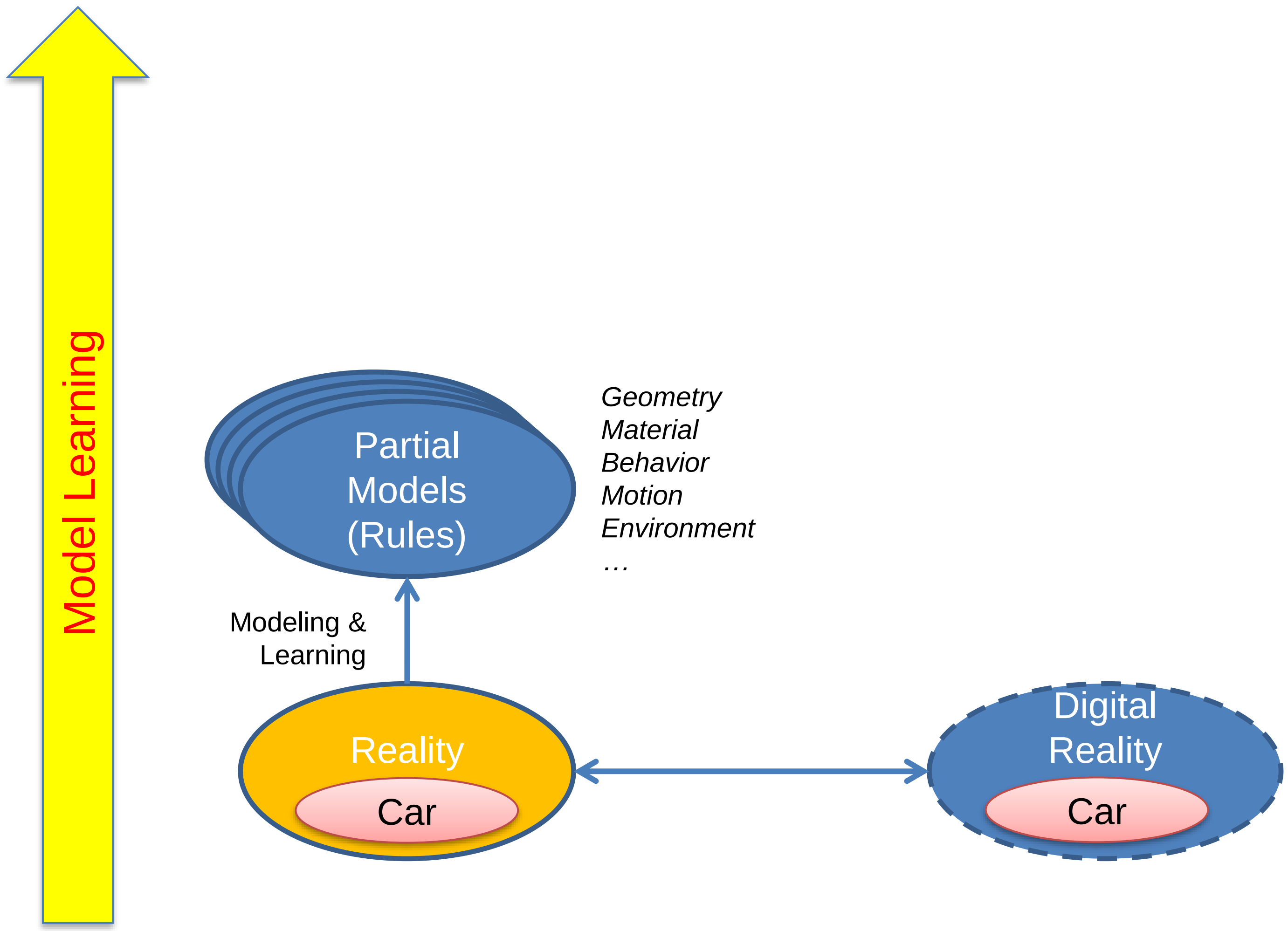
- **Training and Validation in Reality**
  - E.g. driving millions of miles to gather data
  - Difficult, costly, and non-scalable



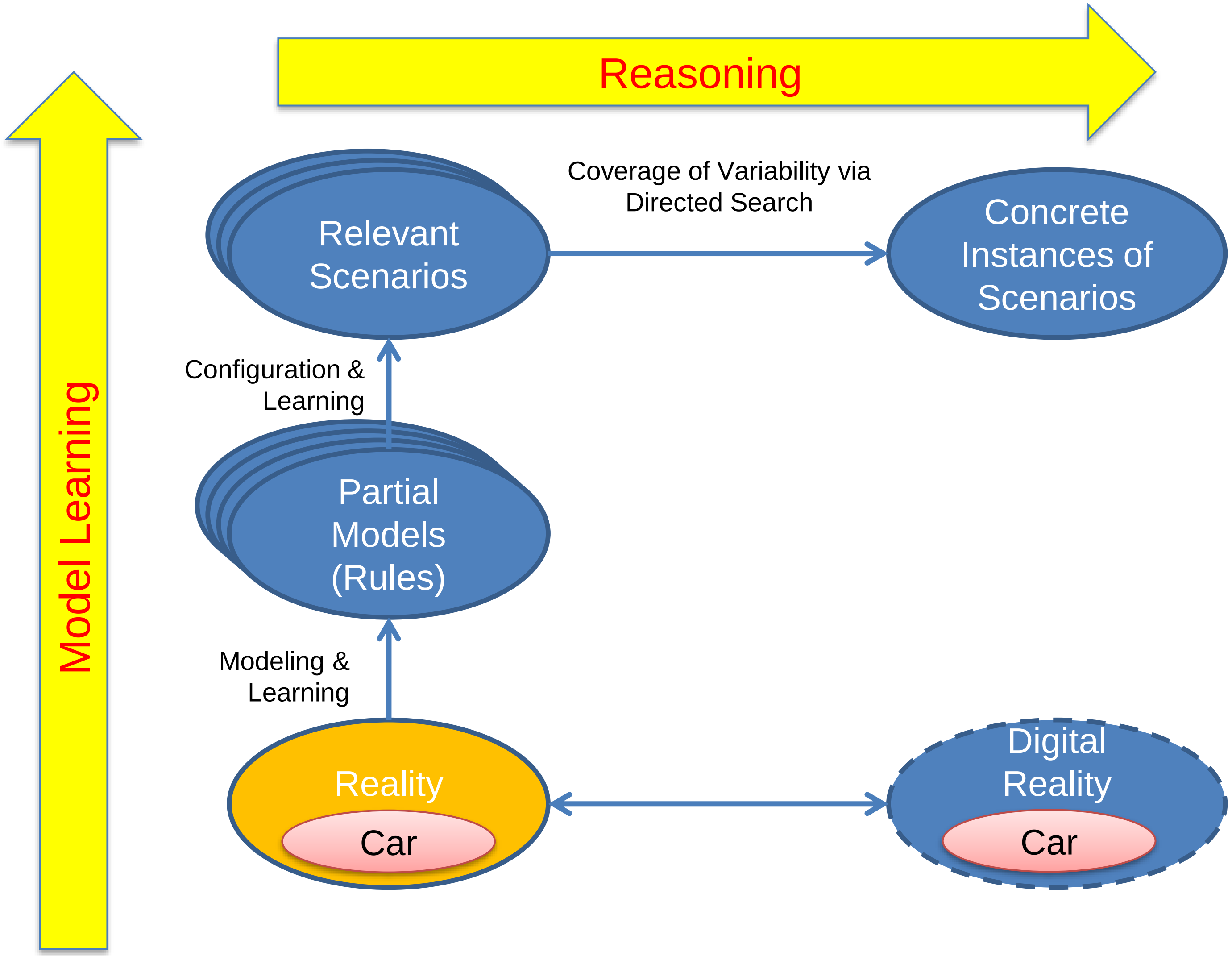
Digital Reality

- **Training and Validation in the *Digital Reality***
  - Arbitrarily scalable (given the right platform)
  - But: Where to get the models and the training data from?

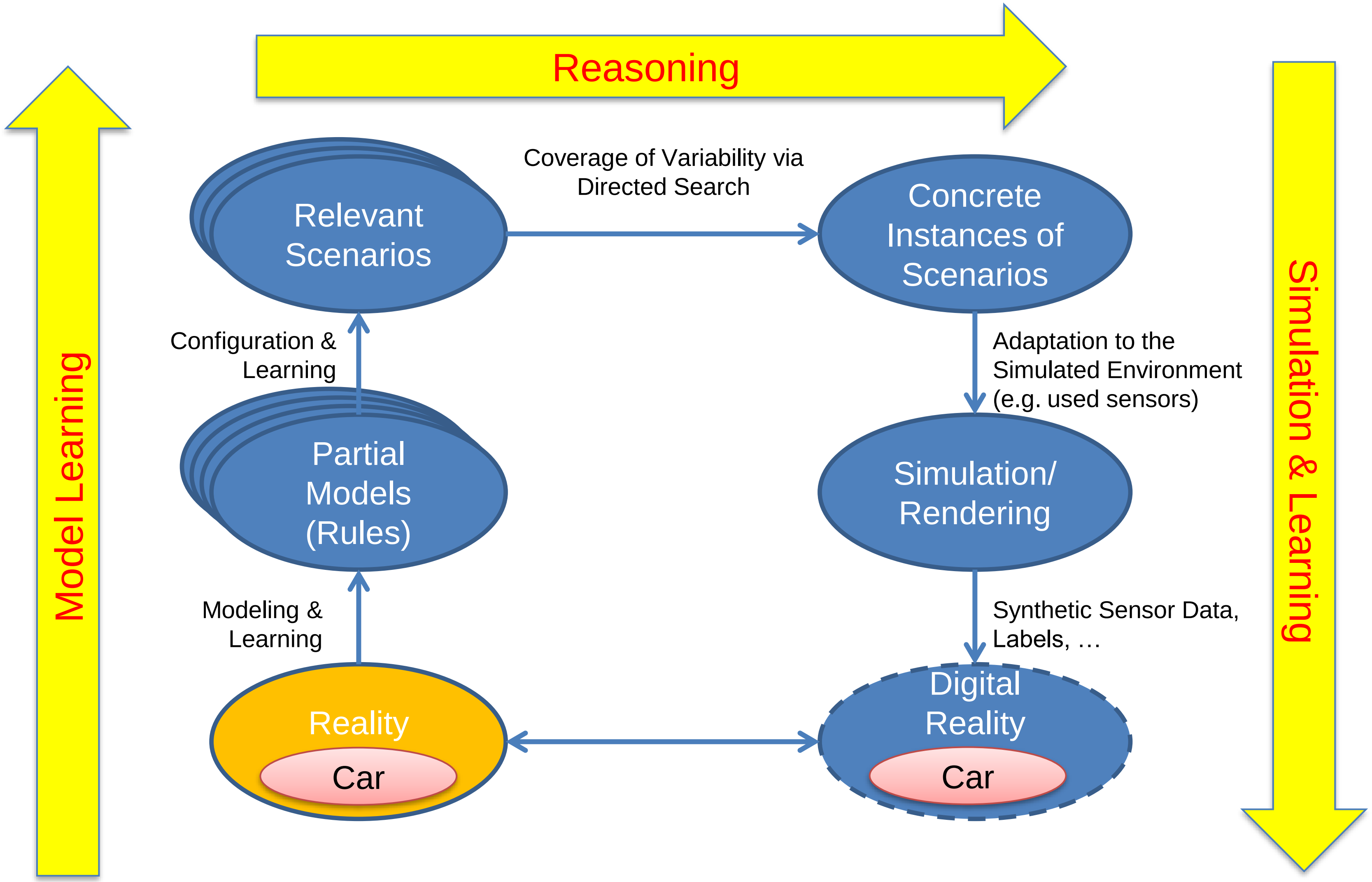
Digital Reality: AI to Optimize and Certify AI



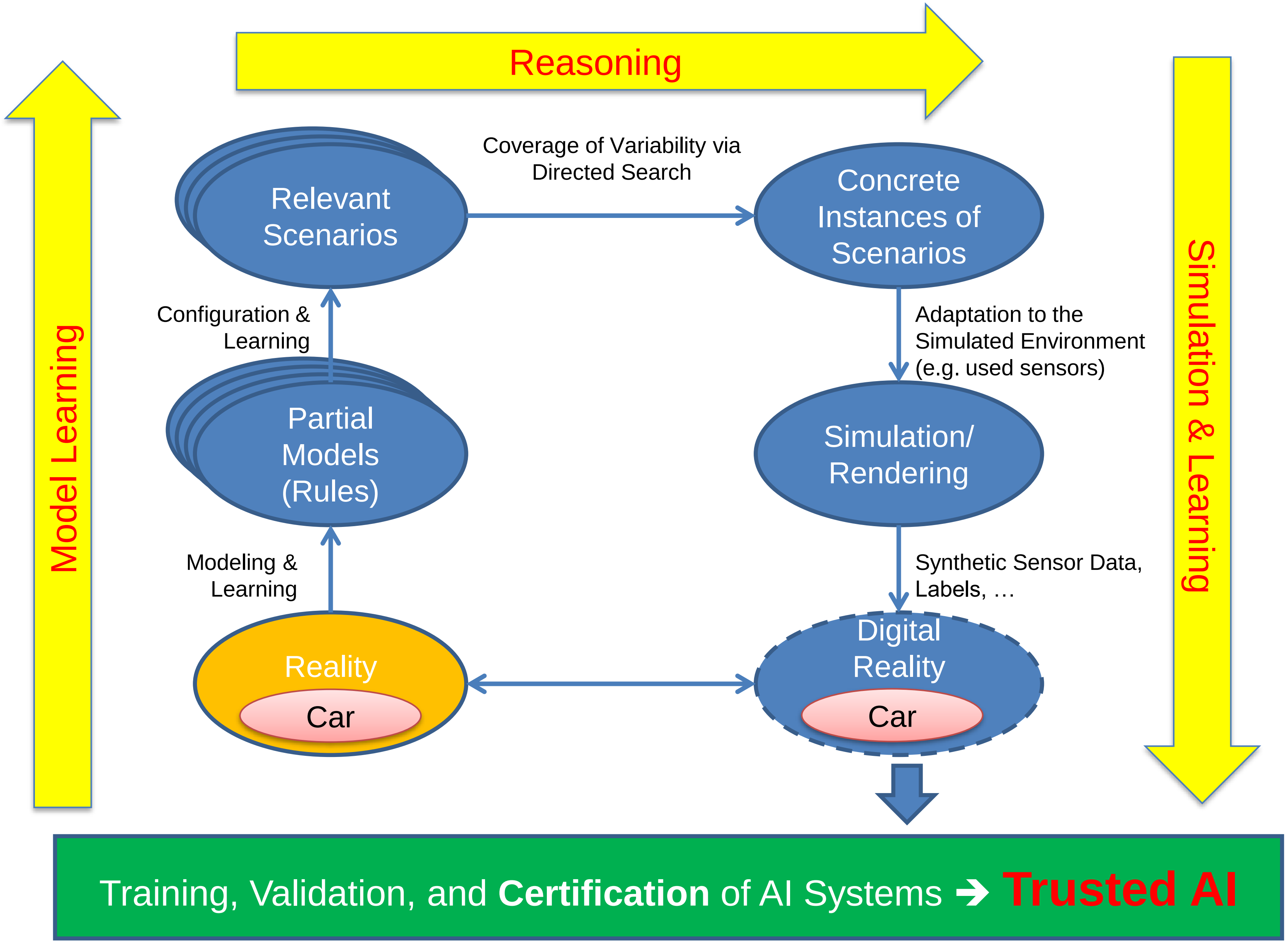
Digital Reality: AI to Optimize and Certify AI



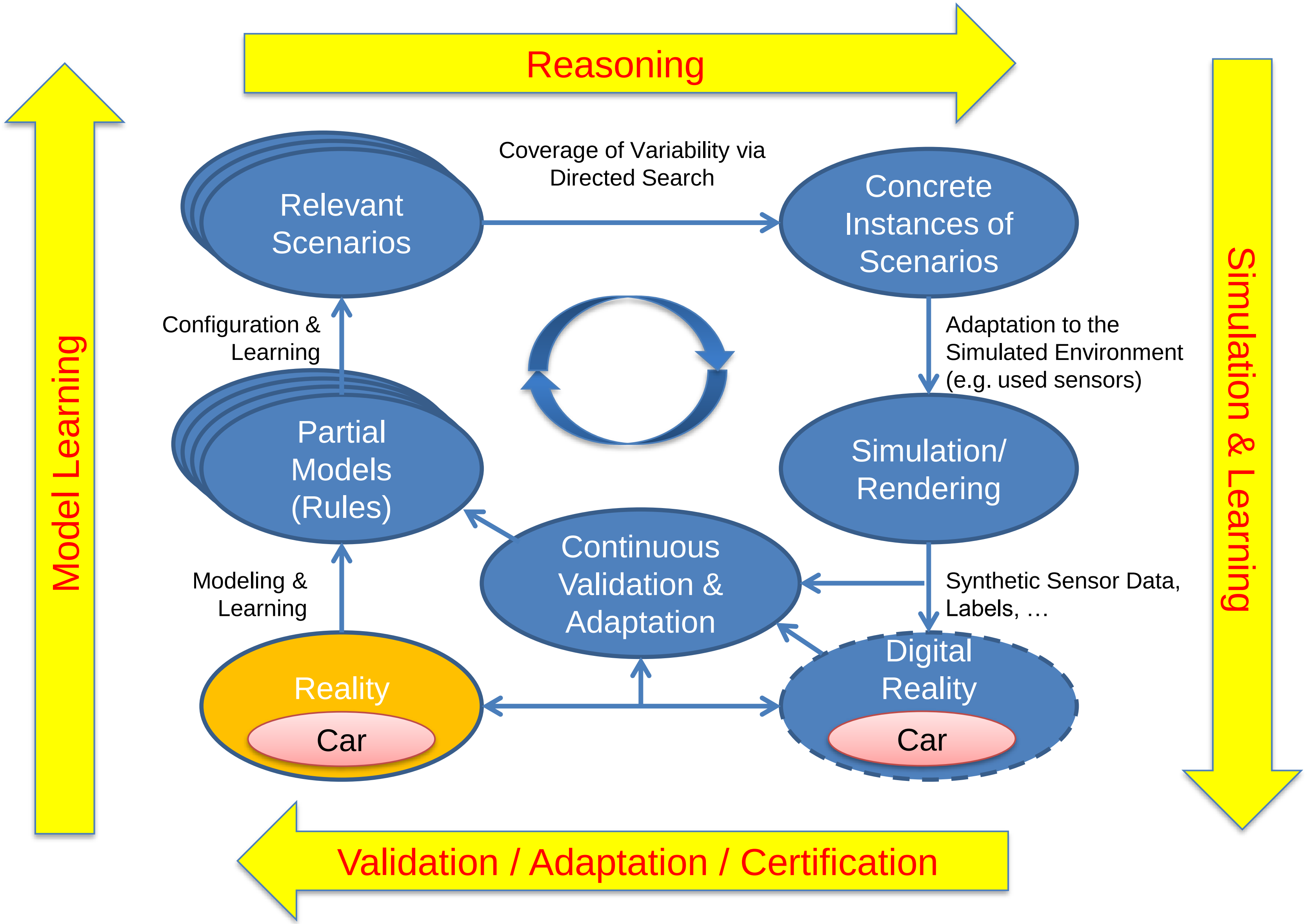
Digital Reality: AI to Optimize and Certify AI



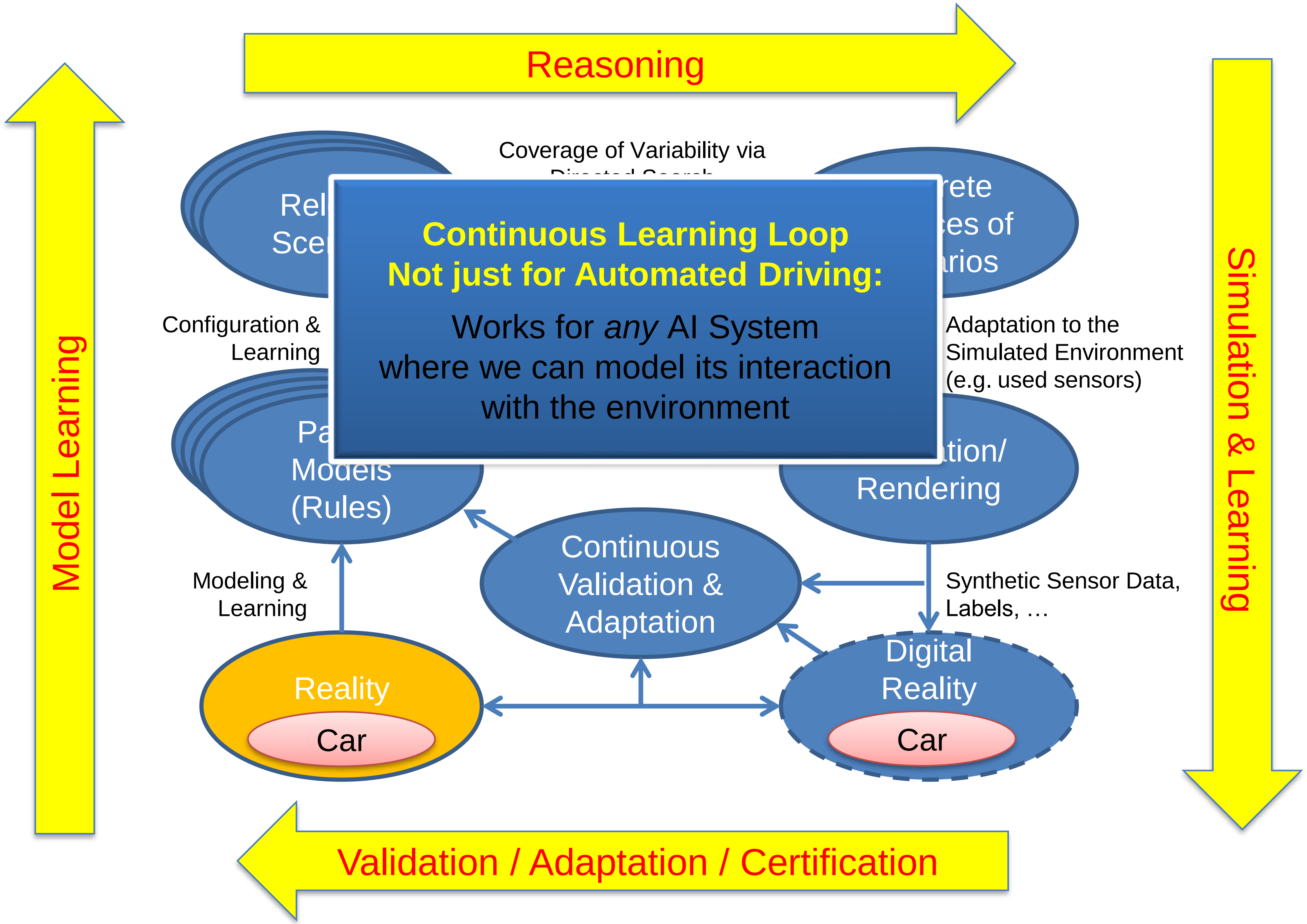
Digital Reality: AI to Optimize and Certify AI



Digital Reality: AI to Optimize and Certify AI

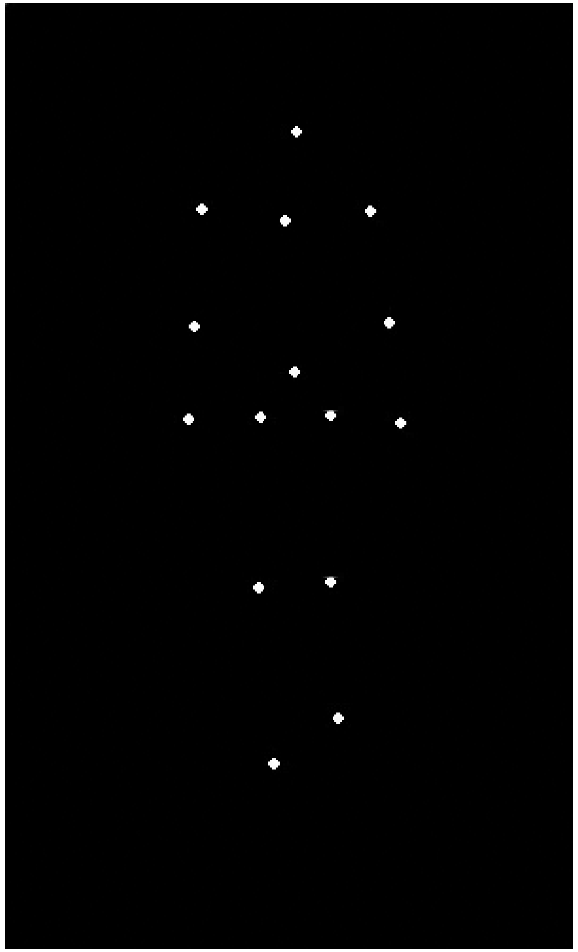


Digital Reality: AI to Optimize and Certify AI



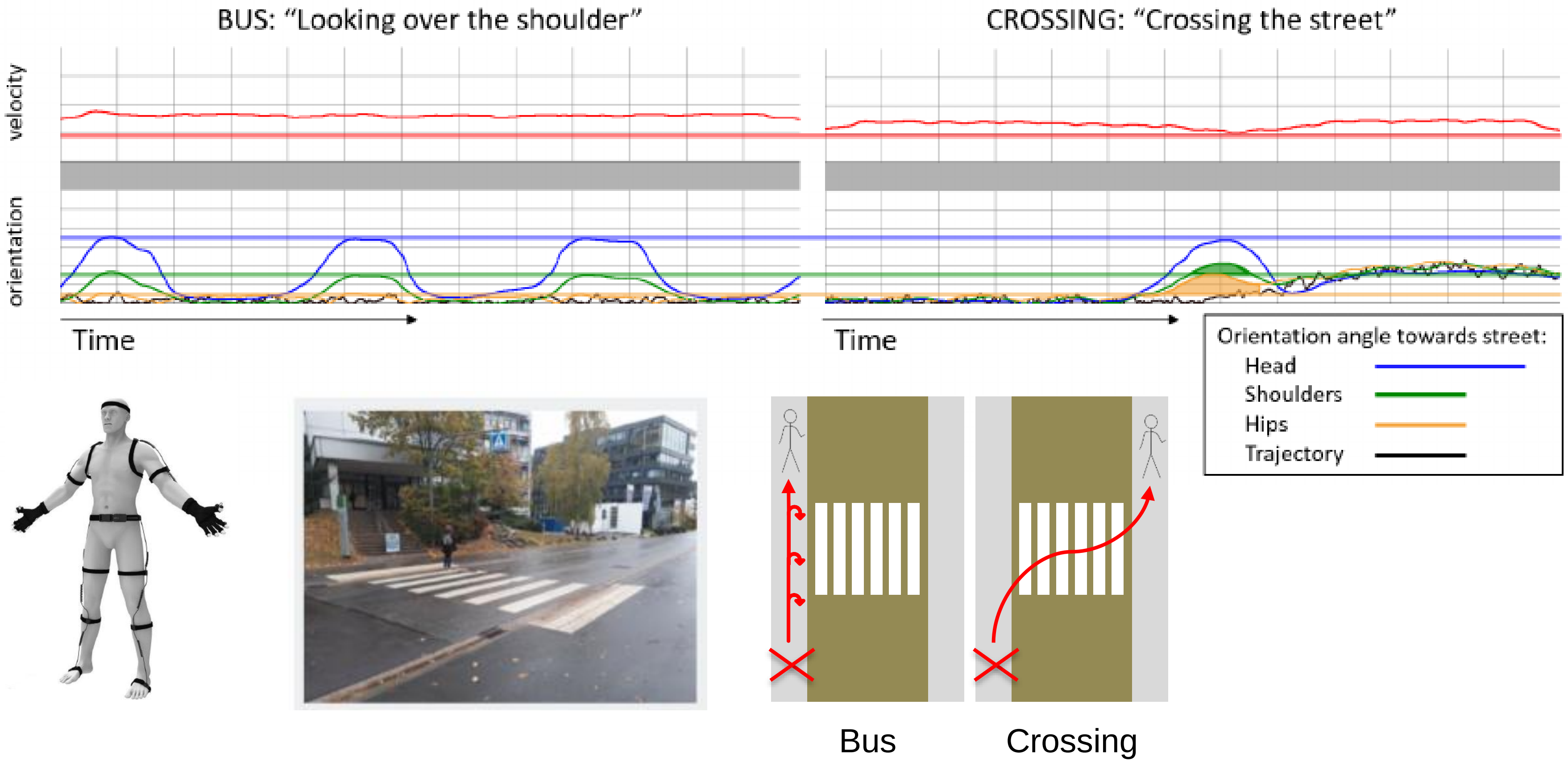
Challenge: Better Models of the World (e.g. Pedestrians)

- **Long history in motion research (>40 years)**
  - E.g. Gunnar Johansson's Point Light Walkers (1974)
  - Significant interdisciplinary research (e.g. psychology)
- **Humans can easily discriminate different styles**
  - E.g. gender, age, weight, mood, ...
  - Based on minimal information
- **Can we teach machines the same?**
  - Detect if pedestrian will cross the street
  - Parameterized motion model & style transfer
  - Predictive models & physical limits



Challenge: Pedestrian Motion

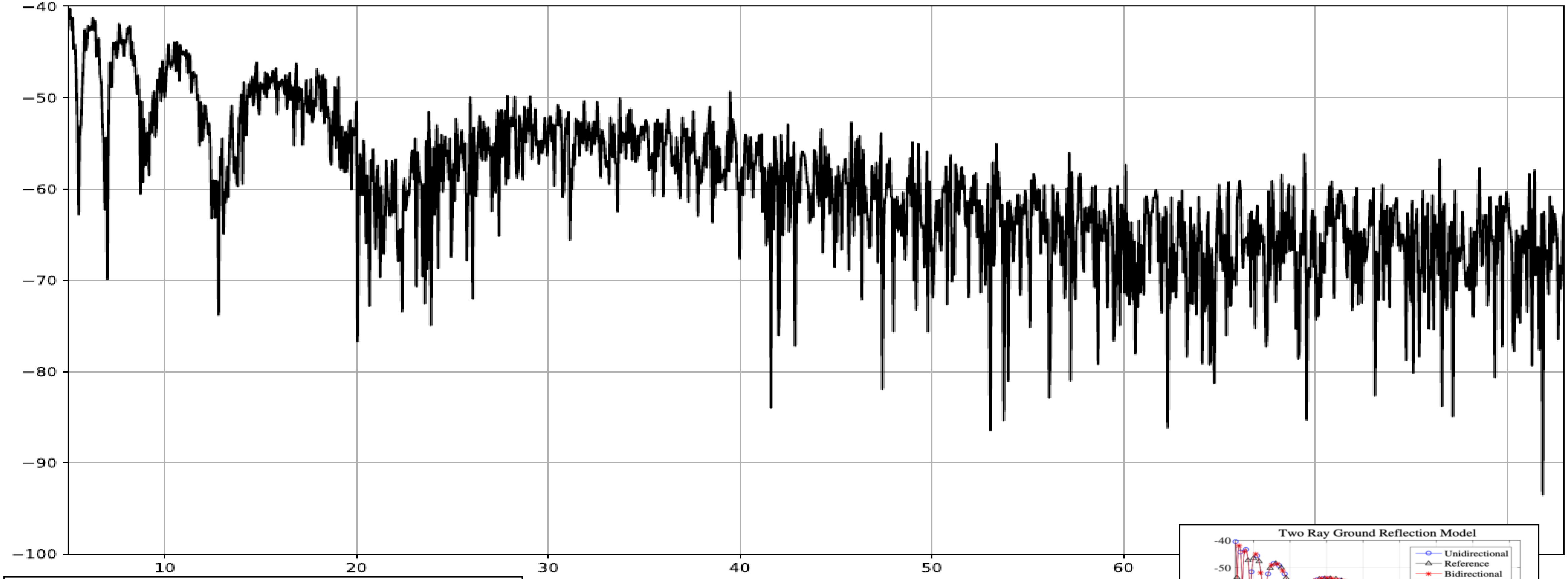
- **Characterizing Pedestrian Motion**
  - Clear motion differences when crossing the street

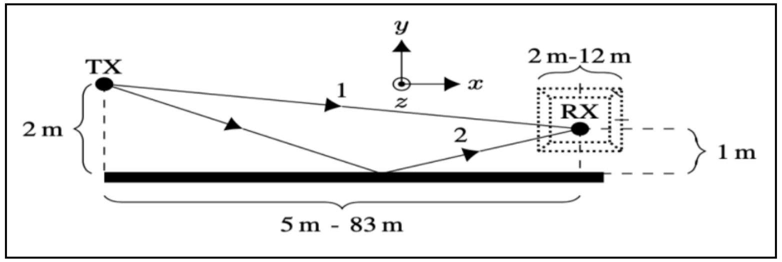


Challenge: Better Simulation (e.g. Radar Rendering)

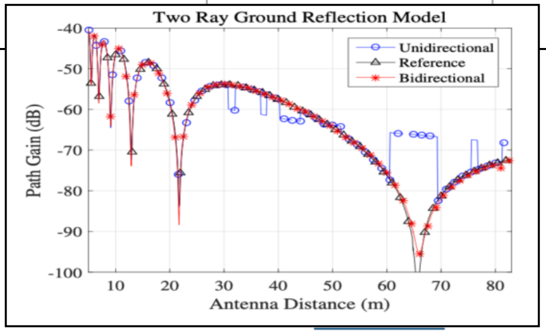
- **Key differences to simulation of optics**
  - Longer wavelength: Geometric optics (rays) not sufficient
  - Need for *some* wave optics
    - Interference of multi-path interactions (coherent radiation, GO/PO)
    - Need for polarization and phase information
    - Diffraction from rough surfaces and edges
  - Highly different goals
    - Optical: Focus on *diffuse* effects (+ some highlights, reflections, etc.)
    - Radar: Focus on *specular* transport only (i.e. caustic paths)
- **Completely novel approach (beyond ray tracing)**
  - Using latest Monte-Carlo techniques (BiDir, MIS, VCM, …)
  - Using recent work on Path Guiding [Grittmann, Rath et al., Siggraph`19, `20, `22]
- **Bringing together radar & latest research on Monte-Carlo simulation**

Radar (Two-Way Ground Reflection): Existing Algorithms



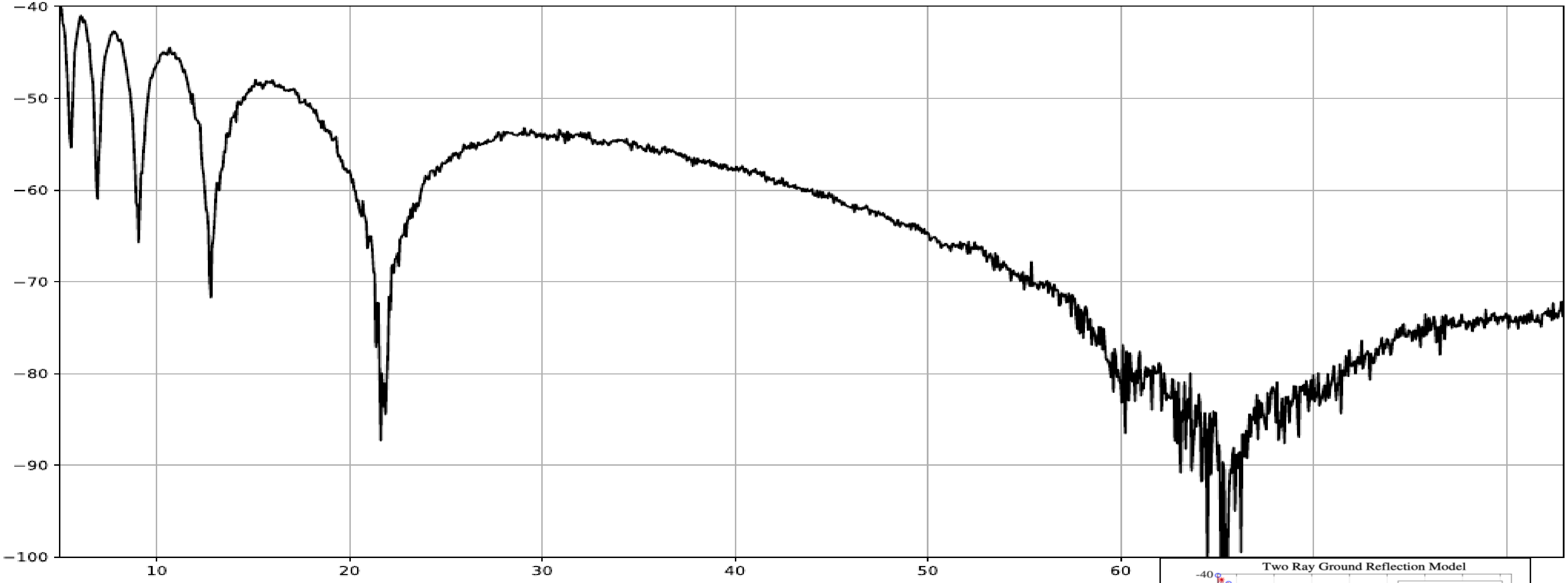


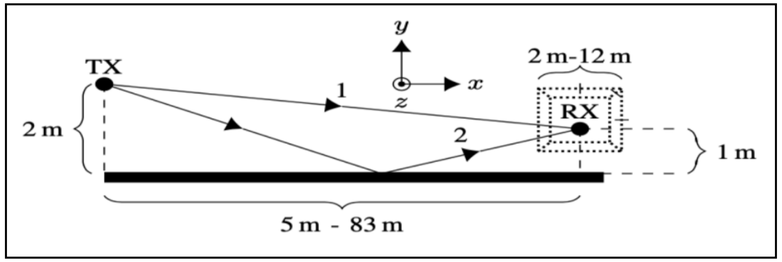
Path Tracing



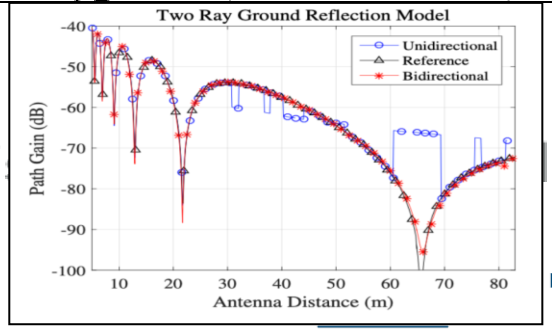
informatics

Radar (Two-Way Ground Reflection): Using Modern Monte-Carlo Algorithms!





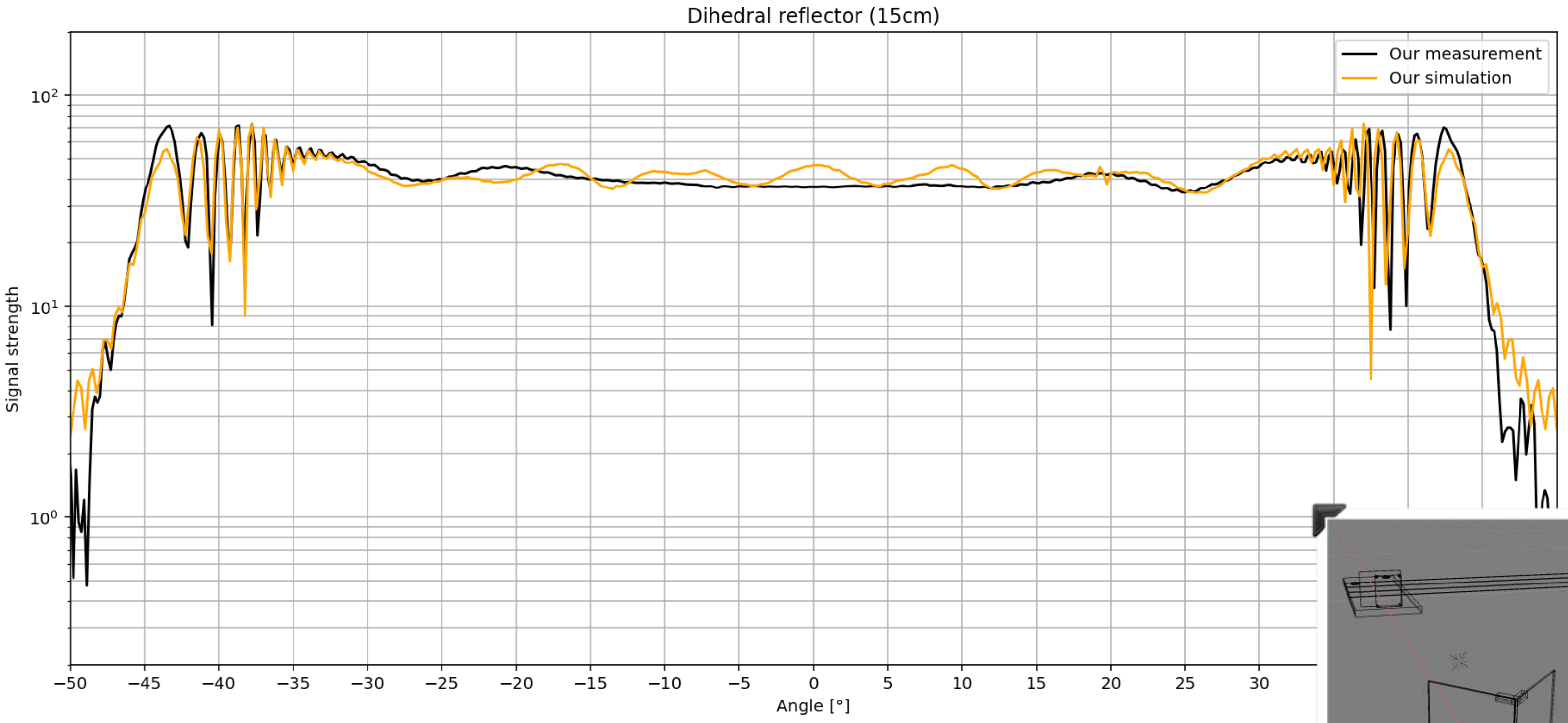
"Filtering" + Guiding + Low Di… …ling

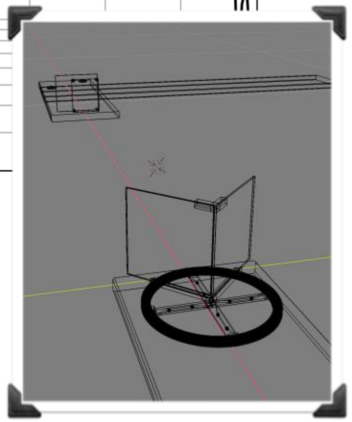


Informatics

Our Simulation

Ours (Physical Optics + Monte Carlo)

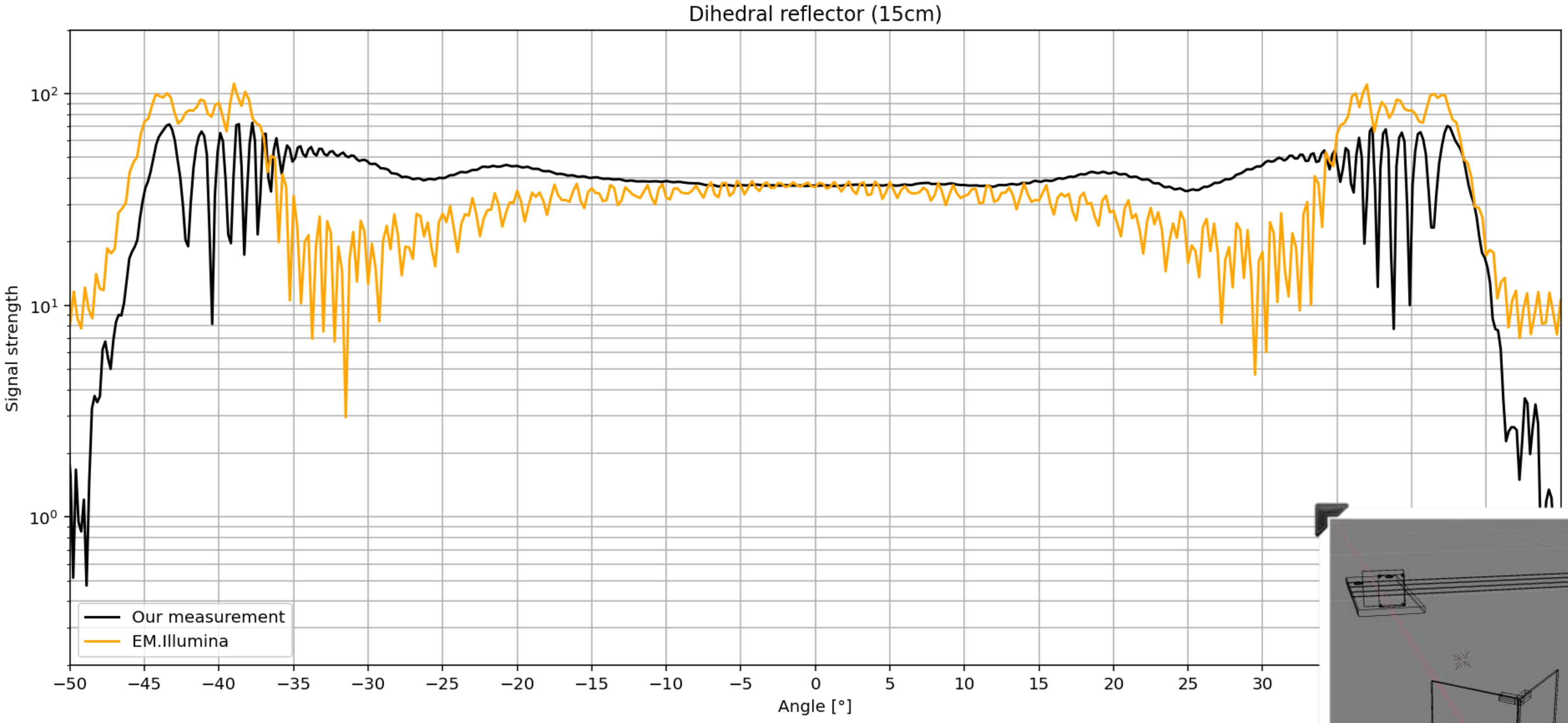


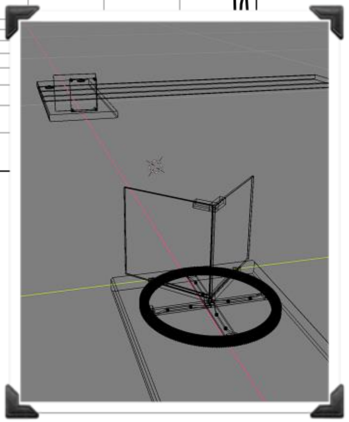


runtime: 33 seconds

Commercial Software

EM.Illumina (Physical Optics + Finite Elements)



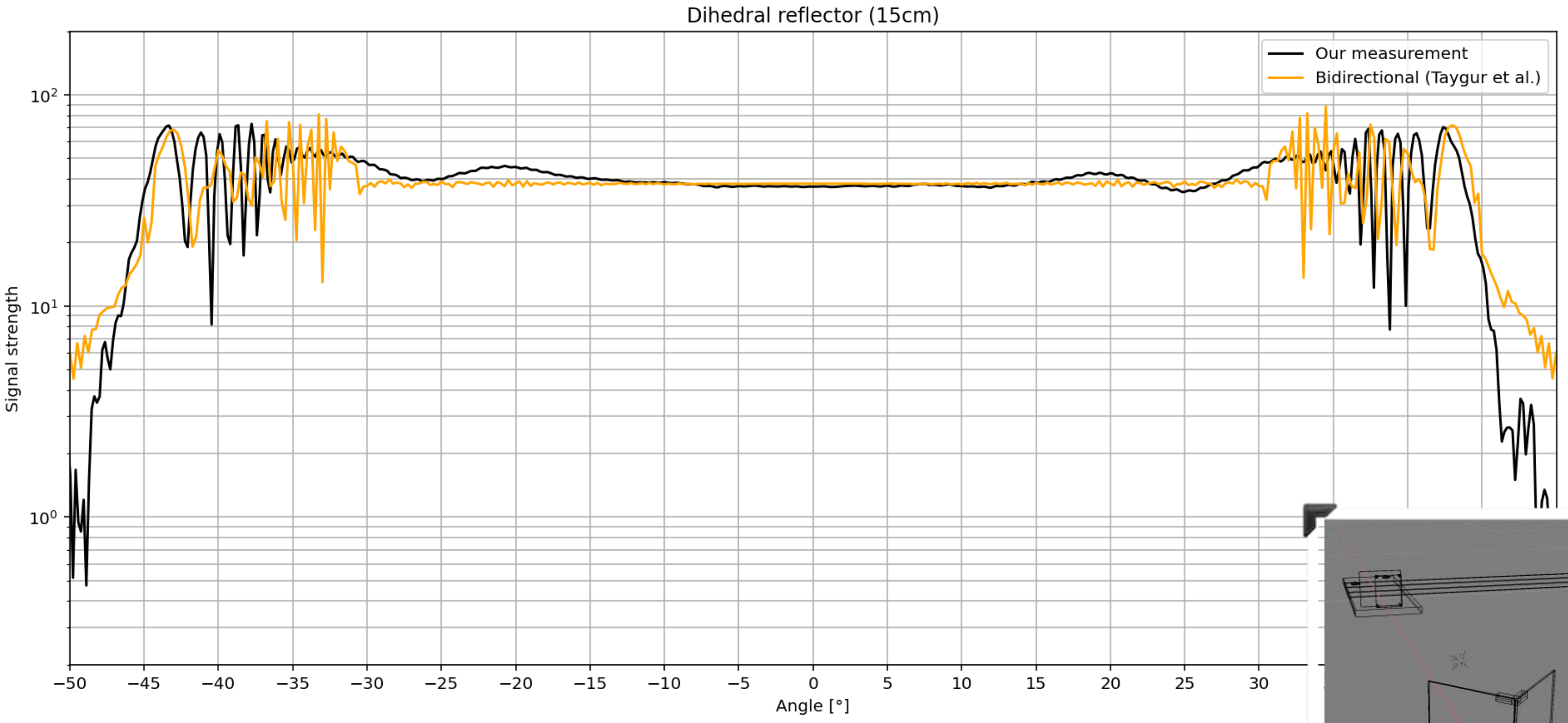


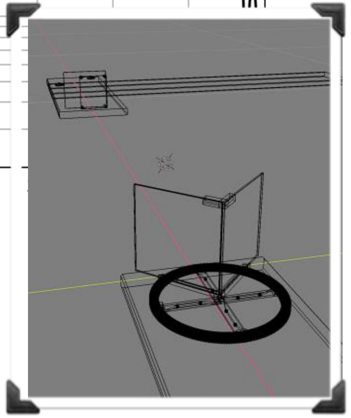
runtime: 13.4 hours

- *EM.Illumina is based on the same physical model (physical optics), but – like virtually all available simulators – uses Finite Elements instead of Monte Carlo*
- *This makes it a lot slower than our method (by a factor of 1,400) and produces results that are not as accurate*

Results from Academia

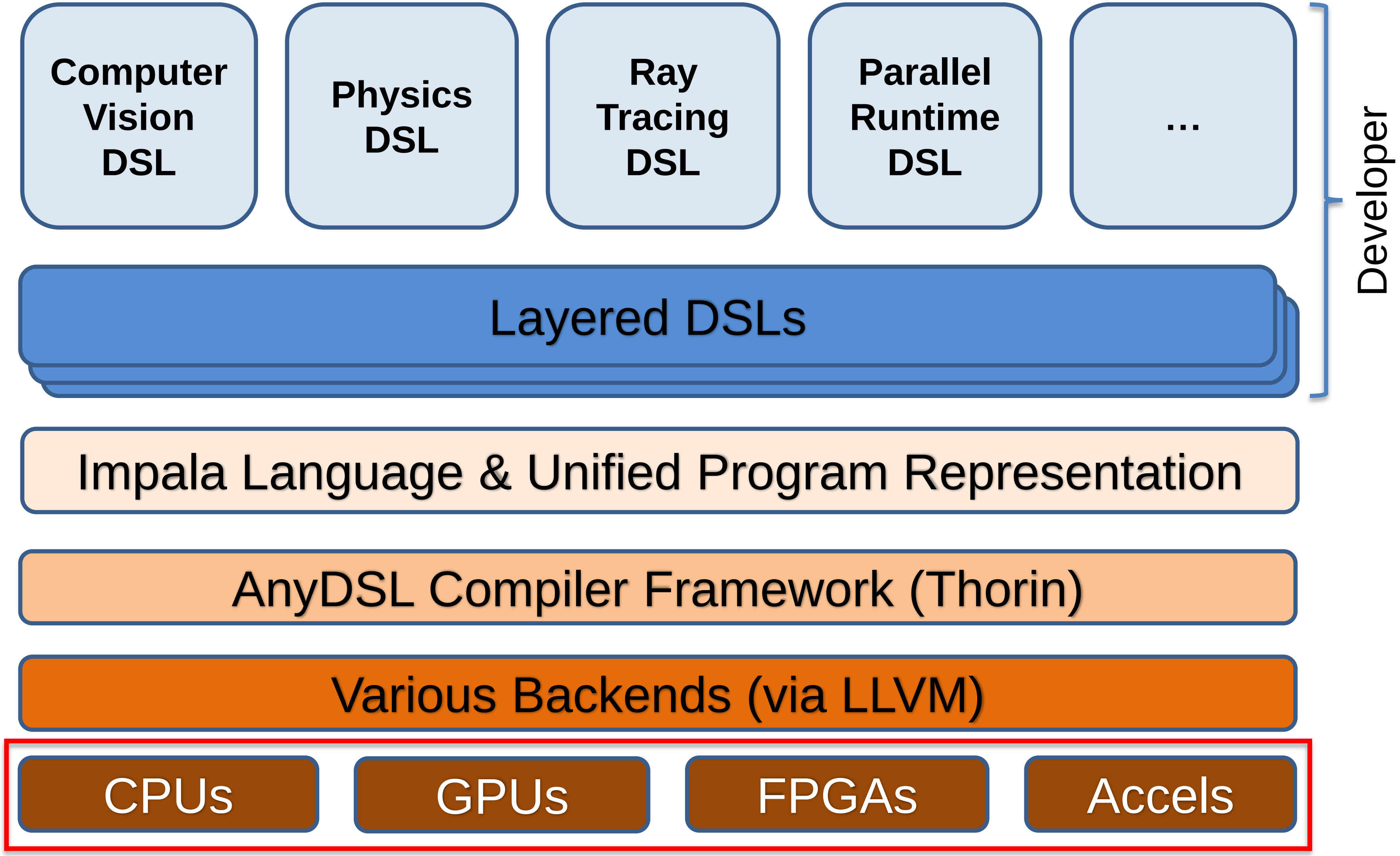
Bidirectional Antenna Coupling (Taygur et al.)





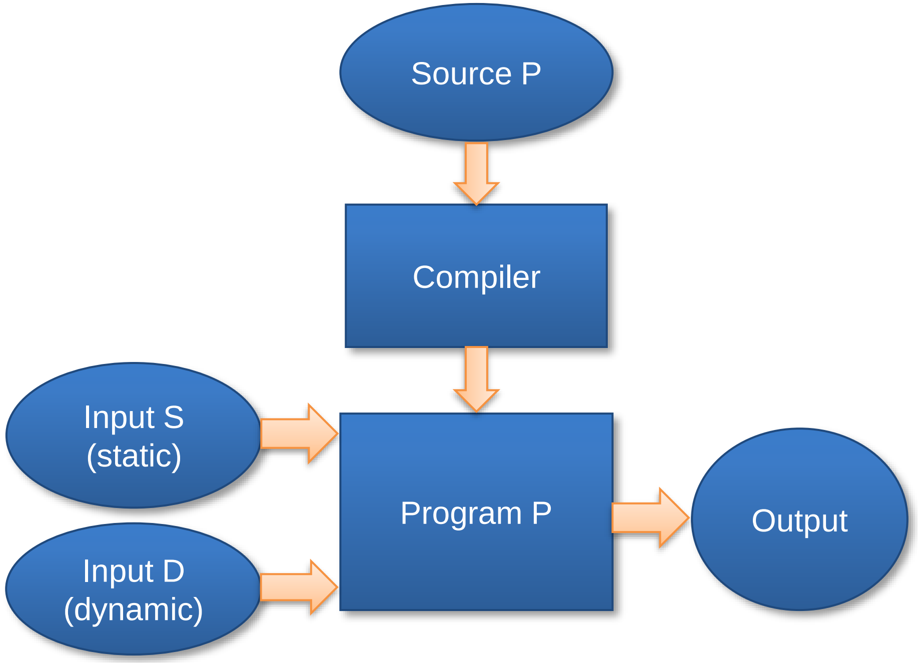
- *Bidirectional Antenna Coupling* (Taygur et al.) is a state-of-the-art algorithm that find connections between RX and TX antenna by starting paths from both sides and connecting them in the middle
- Unfortunately, it makes asymptotic assumptions and is therefore also less accurate for smaller features

AnyDSL Compiler Framework

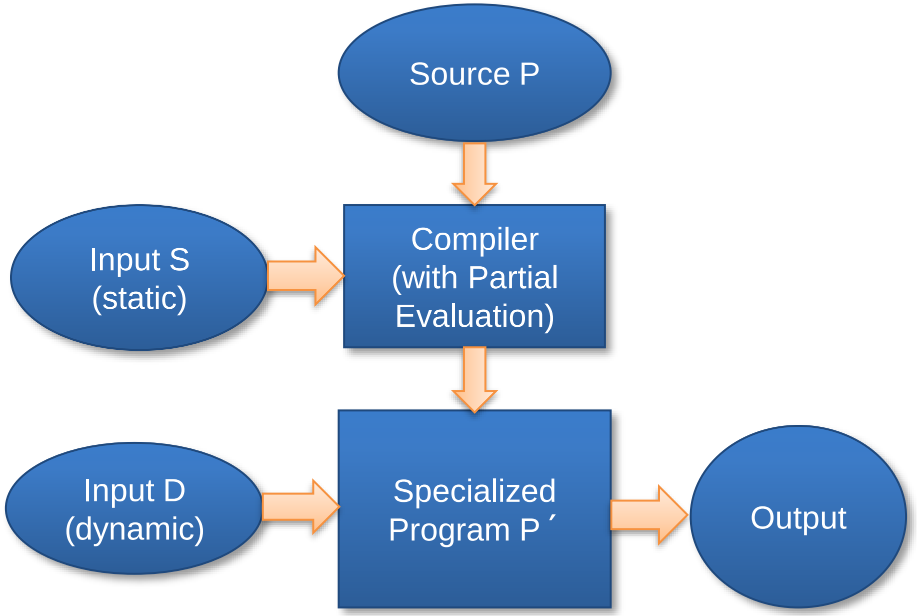


AnyDSL Key Feature: Partial Evaluation

- Left: Normal program execution
- Right: Execution with program specialization (PE)
  - Full control by developer via annotations & compiler evaluation



Traditional Compiler



AnyDSL Compiler

DFKI

SIC Saarland Informatics Campus

Rodent: Lighting Simulation Test Scenes

A. Perard-Gayot, R. Membarth, R. Leissa, S. Hack, P. Slusallek,
Rodent: Generating Renderers without Writing a Generator, **Siggraph 2019**

Performance Result: Lighting Simulation

- Cross-layer specialization (traversal + shading)
  - ~20% speedup vs. no specialization
- Optimal scheduling for each device
  - Megakernel vs. wavefront (of different sizes)

Scene	CPU (Intel™ i7 6700K)	
	Rodent^{WF}	Embree^{WF}
Living Room	9.77 (+23%)	7.94
Bathroom	6.65 (+13%)	5.90
Bedroom	7.55 (+ 4%)	7.24
Dining Room	7.08 (+ 1%)	7.01
Kitchen	6.64 (+12%)	5.92
Staircase	4.86 (+ 8%)	4.48

Msamples/s (higher is better). MK: Megakernel, WF: Wavefront.

Performance Result: Lighting Simulation

- Cross-layer specialization (traversal + shading)
  - ~20% speedup vs. no specialization
- Optimal scheduling for each device
  - Megakernel vs. wavefront (of different sizes)

	CPU (Intel™ i7 6700K)		GPU (NVIDIA™ Titan X)	
Scene	Rodent^{WF}	Embree^{WF}	Rodent^{MK}	OptiX^{MK}
Living Room	9.77 (+23%)	7.94	38.59 (+25%)	30.75
Bathroom	6.65 (+13%)	5.90	27.06 (+31%)	20.64
Bedroom	7.55 (+ 4%)	7.24	30.25 (+ 9%)	27.72
Dining Room	7.08 (+ 1%)	7.01	30.07 (+ 5%)	28.58
Kitchen	6.64 (+12%)	5.92	22.73 (+ 2%)	22.22
Staircase	4.86 (+ 8%)	4.48	20.00 (+18%)	16.89

Msamples/s (higher is better). MK: Megakernel, WF: Wavefront.

Performance Result: Lighting Simulation

- Cross-layer specialization (traversal + shading)
  - ~20% speedup vs. no specialization
- Optimal scheduling for each device
  - Megakernel vs. wavefront (of different sizes)

	CPU (Intel^{TM} i7 6700K)		GPU (NVIDIA^{TM} Titan X)		
Scene	Rodent^{WF}	Embree^{WF}	Rodent^{MK}	Rodent^{WF}	OptiX^{MK}
Living Room	9.77 (+23%)	7.94	38.59 (+25%)	43.52 (+42%)	30.75
Bathroom	6.65 (+13%)	5.90	27.06 (+31%)	35.32 (+42%)	20.64
Bedroom	7.55 (+ 4%)	7.24	30.25 (+ 9%)	38.88 (+29%)	27.72
Dining Room	7.08 (+ 1%)	7.01	30.07 (+ 5%)	40.37 (+29%)	28.58
Kitchen	6.64 (+12%)	5.92	22.73 (+ 2%)	32.09 (+31%)	22.22
Staircase	4.86 (+ 8%)	4.48	20.00 (+18%)	27.53 (+39%)	16.89

Msamples/s (higher is better). MK: Megakernel, WF: Wavefront.

Performance Result: Lighting Simulation

- Cross-layer specialization (traversal + shading)
  - ~20% speedup vs. no specialization
- Optimal scheduling for each device
  - Megakernel vs. wavefront (of different sizes)

	CPU (Intel^{TM} i7 6700K)		GPU (NVIDIA^{TM} Titan X)			GPU (AMD^{TM} R9 Nano)	
Scene	Rodent^{WF}	Embree^{WF}	Rodent^{MK}	Rodent^{WF}	OptiX^{MK}	Rodent^{MK}	Rodent^{WF}
Living Room	9.77 (+23%)	7.94	38.59 (+25%)	43.52 (+42%)	30.75	24.87	35.11
Bathroom	6.65 (+13%)	5.90	27.06 (+31%)	35.32 (+42%)	20.64	14.95	27.31
Bedroom	7.55 (+ 4%)	7.24	30.25 (+ 9%)	38.88 (+29%)	27.72	19.25	32.90
Dining Room	7.08 (+ 1%)	7.01	30.07 (+ 5%)	40.37 (+29%)	28.58	16.22	30.83
Kitchen	6.64 (+12%)	5.92	22.73 (+ 2%)	32.09 (+31%)	22.22	16.68	28.13
Staircase	4.86 (+ 8%)	4.48	20.00 (+18%)	27.53 (+39%)	16.89	11.74	22.21

Msamples/s (higher is better). MK: Megakernel, WF: Wavefront.

Take-Aways

- Hybrid / Neuro-Explicit AI
  - **Bringing together simulation/HPC with AI**
  - New challenges and trade-offs for HW/SW architectures
- Digital Reality as a Fundamental Tool in AI
  - Modeling, simulation, and learning even in complex environments
  - **Learning and reasoning via continuous feedback loop (e.g. RL)**
- Trusted AI: Providing Guarantees About AI Systems
  - Certification & Validation required to establish trust in AI systems
  - **Needs significant HPC resources for simulations and AI**
- Big Challenges Ahead
  - Many promising partial results already (e.g. car industry) – but still many gaps
  - Tools for new HW/SW architectures (incl. compiler) are still largely missing
  - **Requires closer collaboration of HPC & AI as well as industry & academia**

➔ **Great future for HPC and AI working together: HW, SW, tools, and applications**

Human Centric AI.

Thank you very much for your attention !

Human Centric AI.

SIC Saarland Informatics Campus