



KI Data Tooling Final Event | 05/06 December 2023

Project Overview, Challenges and Stories to tell

Armin Koehler | Robert Bosch GmbH

Our mission within VDA-LI KI Family

EXPERT centric

KI WISSEN Development of methods for the integration of knowledge into machine learning

MODEL centric

KI DELTA LEARNING Development of methods and tools for the efficient expansion and transformation of existing AI modules in autonomous vehicles to meet the challenges of new domains or more complex scenarios



SAFETY centric

KI ABSICHERUNG Methods and measures to safeguard AI-based perception functions for automated driving

DATA centric

KI DATA TOOLING Methods and tools for the generation and refinement of training, validation and safeguarding data for AI functions in autonomous vehicles

Our mission within VDA-LI KI Family

DATA centric

KI DATA TOOLING Methods and tools for the generation and refinement of training, validation and safeguarding data for AI Functions in autonomous vehicles

Challenge: Model vs Data



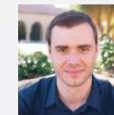
A Chat with Andrew on MLOps: From Model-centric to Data-centric AI

Improving the code vs the data

	Steel defect detection	Solar panel	Surface inspection
Baseline	76.2%	75.68%	85.05%
Model-centric	+0% (76.2%)	+0.04% (75.72%)	+0.00% (85.05%)
Data-centric	+16.9% (93.1%)	+3.06% (78.74%)	+0.4% (85.45%)

Source: [YouTube: A Chat with Andrew on MLOps: From Model-centric to Data-centric AI](#)

Data centric AI – an experience



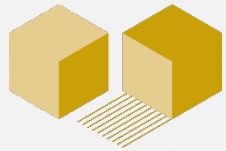
Andrej Karpathy
Formerly PhD Student at Stanford.
Now at Tesla

Amount of lost sleep over ...



Source: [What is Data-Centric AI? – Alectio](#)

Our mission within VDA-LI KI Family



**GOAL-ORIENTED
DATA BASIS**

**METHODS
AND TOOLS**

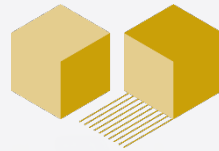
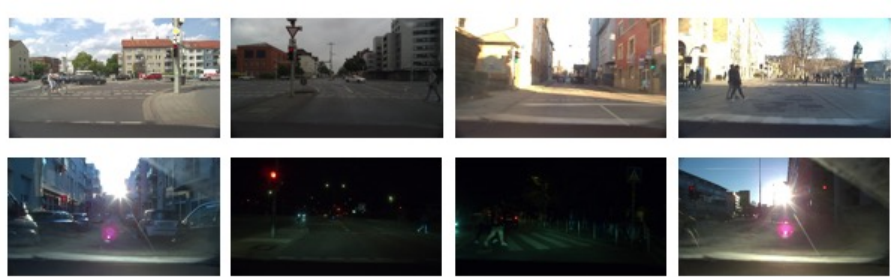
**OPTIMISED AI TRAINING STRATEGY
FOR AUTONOMOUS VEHICLES**

KI Data Tooling develops the complete data solution for the training and validation of AI-based functions for highly automated driving.

Project Goals

Real Data

Goal: Validation data



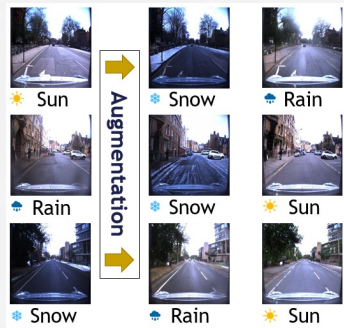
Goal: Data quality assessment

Synth. Replay Data



Augmented Data

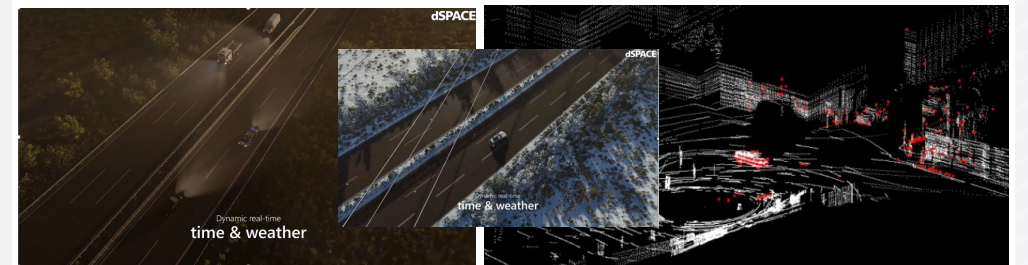
Goal: Increase data diversity



GOAL-ORIENTED DATA BASIS

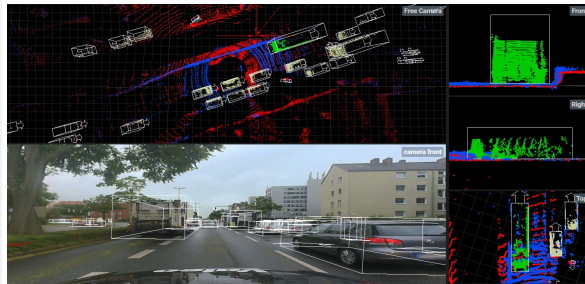
Goal: Training strategies

Diverse Synth. Data

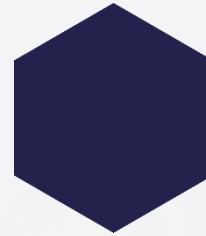


Project Goals

Label QC

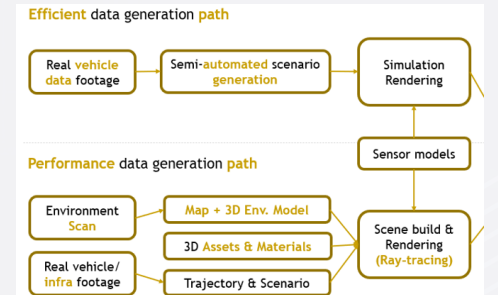


Asset checker

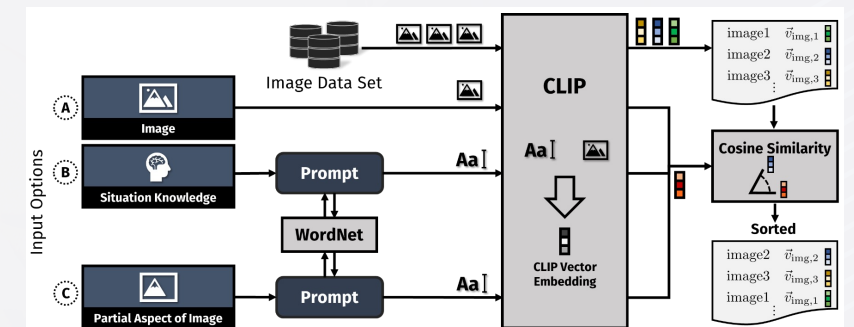


METHODS AND TOOLS

Synth. Data Toolchain

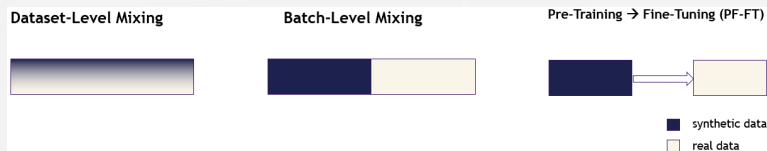
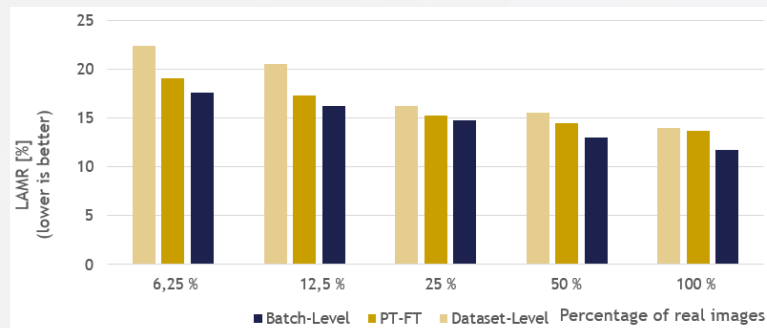


Foundation Model Search



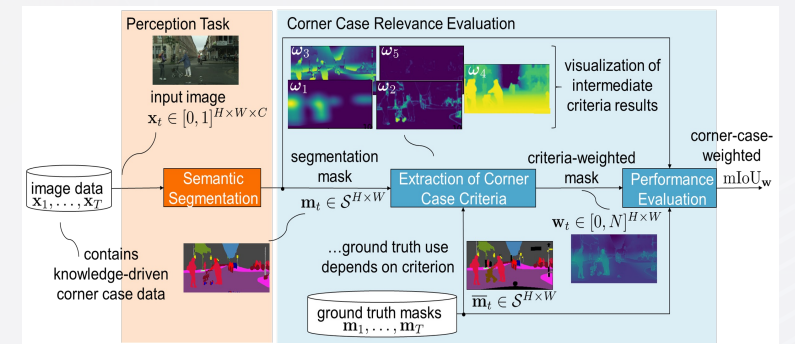
Project Goals

Mixed Training



OPTIMISED AI TRAINING STRATEGY FOR AUTONOMOUS VEHICLES

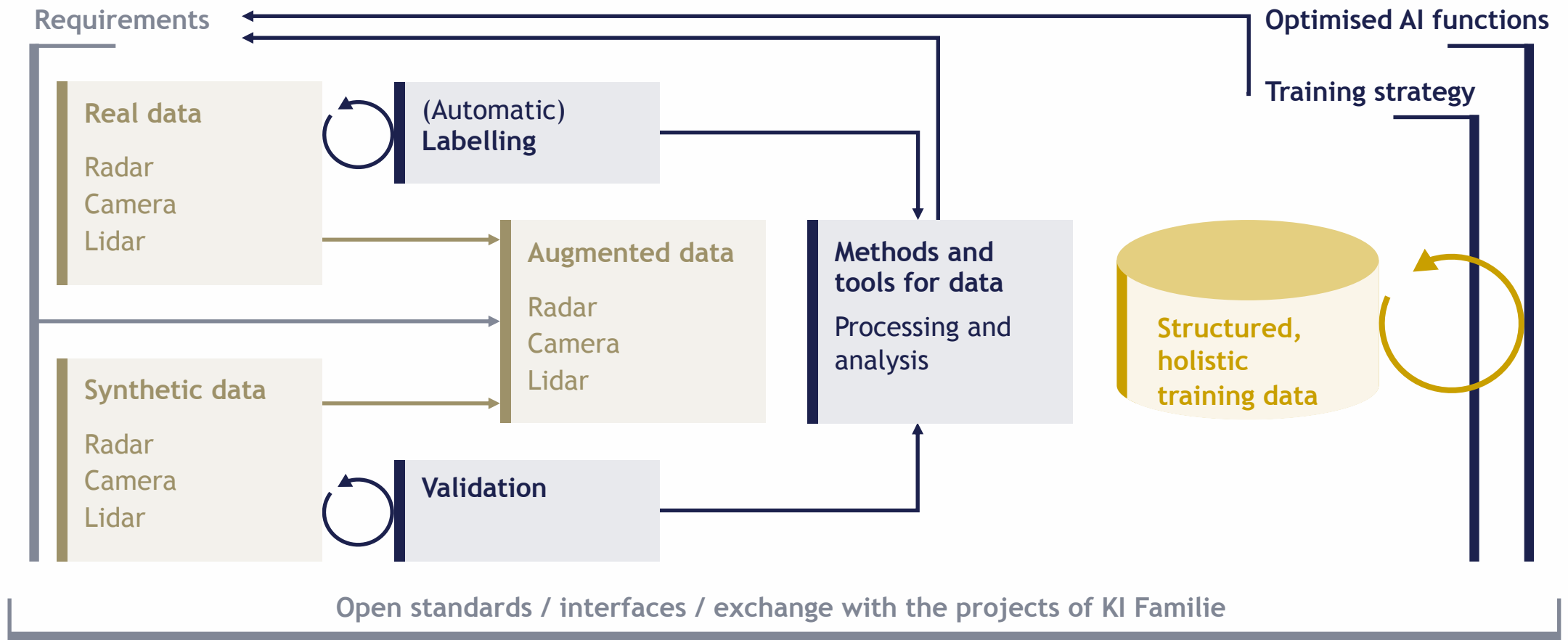
Optimize on Safety & Corner Case



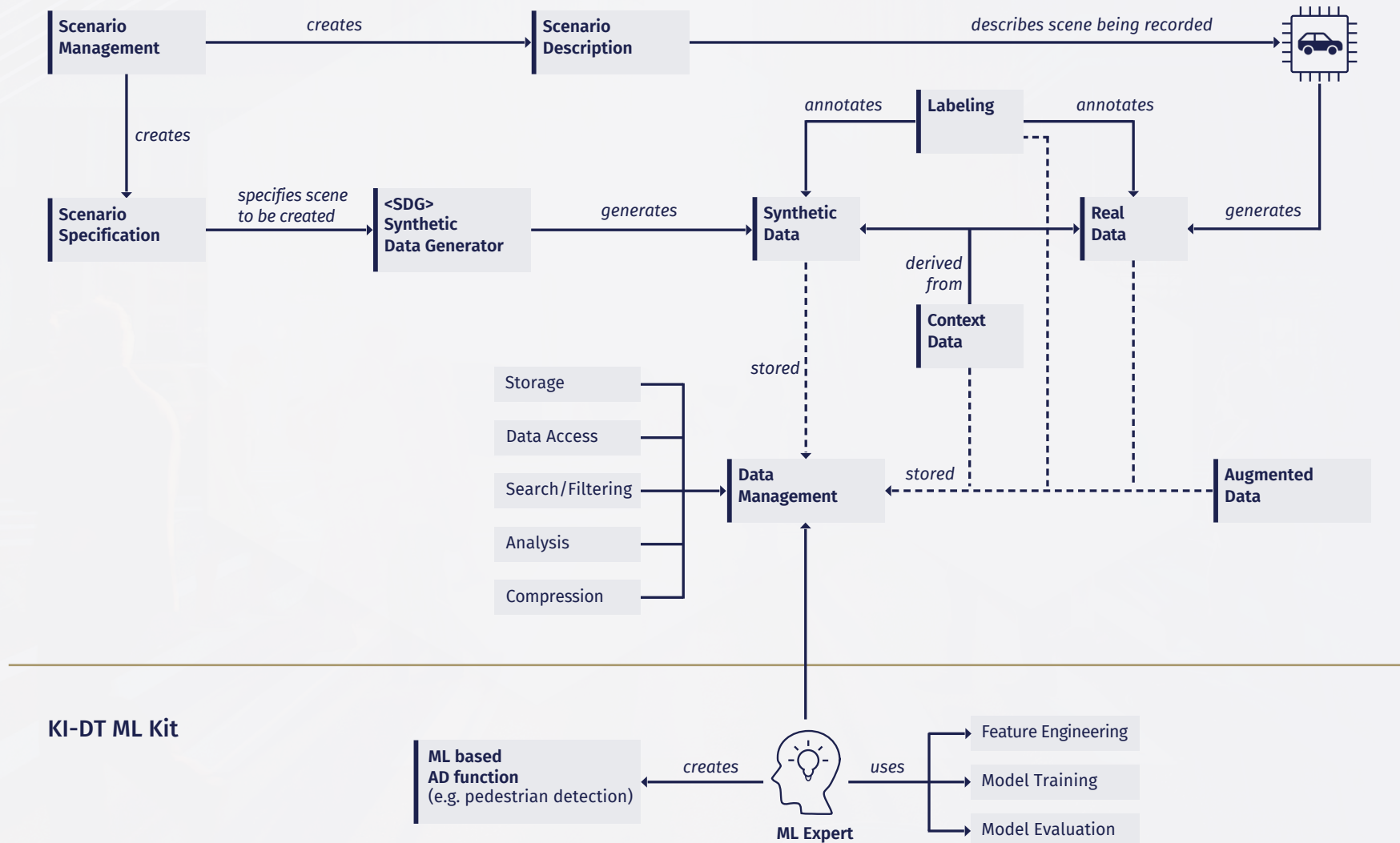
Amodal Training



Conceptual approach



KI-DT Data Kit Framework Architecture



Project Figures, Partners & Data

Consortium lead: **BMW AG**

Budget: **25.7 M €**

Funding: **16.2 M €**

Duration: **45 months**
04/2020 - 12/2023

17 partners



Ki-datatooling.json

```
{
  "project_name": "ki-datatooling",
  "duration": "45 months",
  "partners": 17+3 associated,
  "active members": 219,
  "GUA": 9,
},
"paperwork": [
  {
    "regulatory documents/contracts": 11,
    "confluence pages": 3558,
    "common report pages": 465,
    "publications": 63,
  },
],
"collaboration": [
  "steering committee meetings": 11,
  "project meeting": 27,
  "PMT,GUA,ACR meetings": 292,
  "TP/AP meetings": 814,
],
}
```

KIDT project-data repo
<https://www.ki-datatooling.de/>

A journey through time - 2020

Covid19 Lockdown

Tesla Data Engine

2020 Q2 Q3 Q4 2021 Q2 Q3 Q4 2022 Q2 Q3 Q4 2023 Q2 Q3 Q4

Kick-off

KI Data Tooling
VDA Vorstand der Automobilindustrie

VORHABENBESCHREIBUNG

KI Data Tooling
Methoden und Werkzeuge für das Generieren und Veredeln von Trainings-, Validierungs- und Absicherungsdaten für KI-Funktionen autonomer Fahrzeuge

Ein Projektvorhaben der VDA Leitinitiative autonomes und vernetztes Fahren aus der Projektfamilie Künstliche Intelligenz und maschinelles Lernen im automatisierten Umfeld.

Version 3.0

Vorstellungen der TP-Leitungen

TP1 Validierung synthetischer Sensordaten
Lead: Bosch

Armin Koehler (Bosch), Minokano Li (Bosch)

TP2 Qualitätsanforderungen und Effizienzpotenziale der Datengenerierung und -bereitstellung
Lead: BMW, Co-Lead: Uni Kassel

Maarten Bleshaar (Uni Kassel), Hans-Jörg Vogel (BMW), Cornelia Denk (BMW)

TP3 Methoden und Tools für die Erfassung und Veredelung von Realdaten
Lead: Valeo

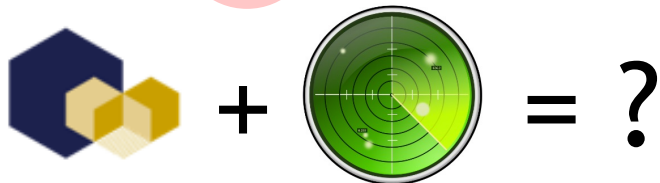
Marcel Matz (Valeo)

TP4 Leitlinien und Übertragbarkeit
Lead: Continental, Co-Lead: DLR

Evren Firmis (Continental), Sven Hallerbach (DLR)

12 | 23.04.2020 | KI Data Tooling Kickoff

~~ASTY-eXit~~



Wir leben in herausfordernden Zeiten – wir brauchen im Projekt viel Kommunikation und die Initiative jedes Einzelnen.

Offenheit, Wertschätzung und Vertrauen sind die Grundlage, um uns laufend zu verbessern. Sprechen Sie an, was nicht gut läuft. Loben Sie, was gut läuft.

#GernePerDu

A journey through time - 2020/2021

Waymo: world's 1st driverless robotaxi



Consortia agreement contract



Starting TPx cross project working groups

- Scenario catalogue
- Data formats
- Model zoo
- Toolchains
- Labeling

1st measurement campaign



KIDT @ BMWi/VDA
“Durch Kooperation
an die Spitze”



ca. 3.000 Besucher während der Veranstaltung

Davon

1.440 vorab registriert
und

Rd. 90 Pressevertreter

Präsentation des Projektes **KI Data Tooling** innerhalb der Fachsession KI Familie mit anschließender **Diskussion**

A journey through time - 2021/2022

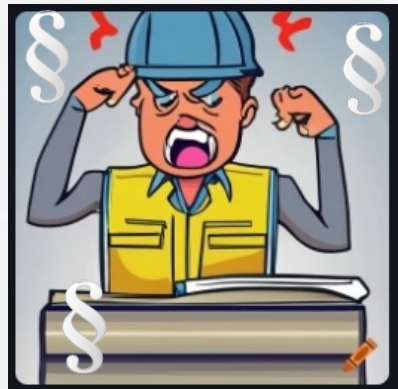
“Ampel” Coalition

Cruise: world’s 2nd robotaxi service
Russia invades Ukraine



1st GUAs signed & started (6,8)

- Synthetic data
- Data storage & compute platform



Crayon V3

A frustrated engineer has to deal with law book



Data transfer agreement

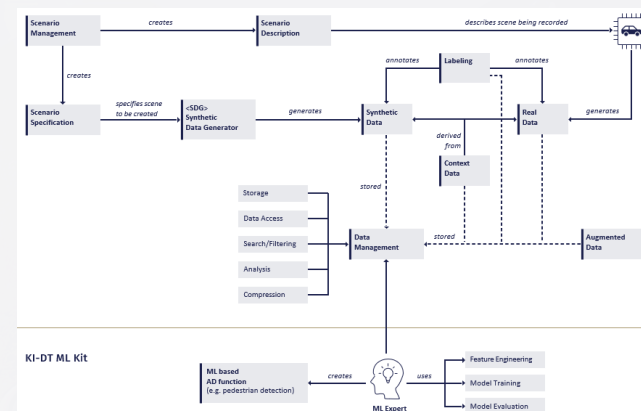
GUA3 update: call for bidding & awarding
(Scan, Mapping, 3D Reconstruction)

GUA - 2nd wave (GUA4,5,7)

- Static & Dynamic Assets for synthetic data
- Labeling



KI-DT Framework Architecture was born



A journey through time - 2022

DeepMind's GATO transformer



2020	Q2	Q3	Q4	2021	Q2	Q3	Q4	2022	Q2	Q3	Q4	2023	Q2	Q3	Q4
------	----	----	----	------	----	----	----	------	----	----	----	------	----	----	----

- Collaborative event with KI Absicherung
- Hybrid event approx. 110 participants, YouTube livestream
- 2 keynotes (Nvidia, Bosch), 2 panel discussions, specialist presentations on the KI-DT workflow, 3 Deep Dives, poster session, driving simulator



Mid-term event

Jonas Lödefink passed away

Von: Prof. T. Fingscheidt via Ki-dt-all <ki-dt-all@eict.de>
Gesendet: Donnerstag, 30. Juni 2022 10:18
An: ki-dt-all@eict.de
Betreff: Jonas Löhdefink

Dear colleagues,

I have to convey the sad message that our dear colleague Jonas Löhdefink has passed away in a hospital in New Orleans yesterday at 5 pm CEST. After a wonderful week at the CVPR conference, after presentation of his recent achievements of machine-learned image compression, a terrible accident happened last Sunday.

We are lacking words and are greatly saddened.

Tim Fingscheidt

A journey through time - 2022

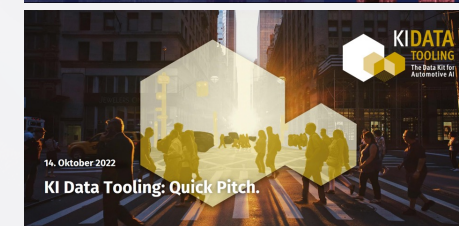
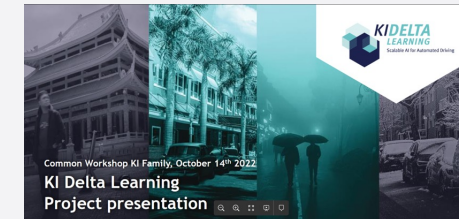
Mercedes: world's 1st certified L3 system — Free Chat GPT 3.0



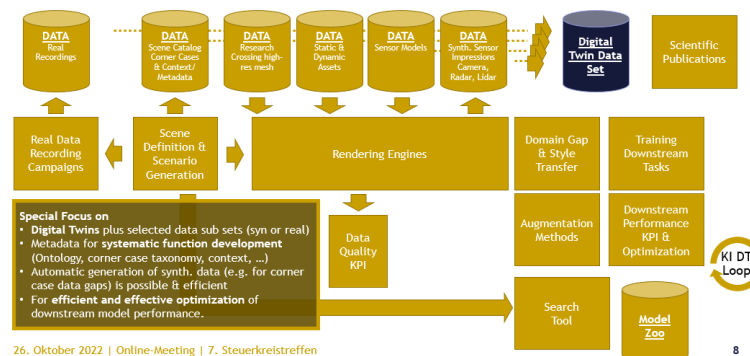
Project Meeting Munich:

- Focus update: common results
- Idea: publication of dataset with challenges was born

KI Family Tech exchange WS



Publishing a KI DT Data Set - A central KI DT product.



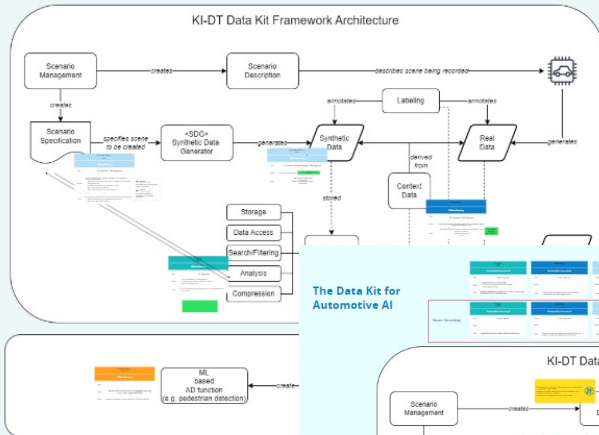
A journey through time - 2023

GTP 4.0 (Text/Image)

EU agrees on AI act

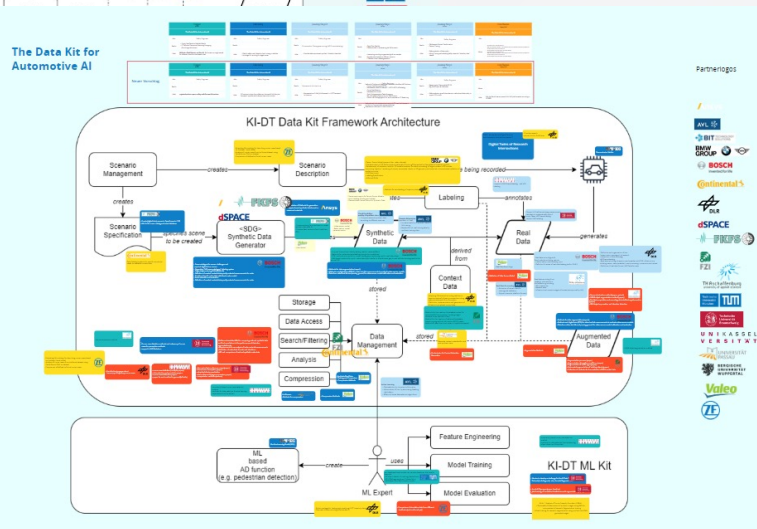


KI-Entwicklerstory
Challenge & Datensatz
Veröffentlichung



Project Meeting Stuttgart:

- Big picture workshops
- The storylines for final event presentation were born



Two additional GUA & re-budgiting:

- GUA9: synthetic data with high automation
- GUA6E: diverse synthetic training data
- Savings by labelling and compute platform



A journey through time - 2023

Hamas terror in Israel
Cruise cancels robotaxi service

2020	Q2	Q3	Q4	2021	Q2	Q3	Q4	2022	Q2	Q3	Q4	2023	Q2	Q3	Q4
------	----	----	----	------	----	----	----	------	----	----	----	------	----	----	----



“Praying for data”

• Data set publication on risk

KIDT Final Event in Munich



Request for participation

Please scan QR code to join our  **Mentimeter** channel.



- At the end of each day, we will ask for feedback.
- Questions for the panel discussion can be posted during the whole event.



Armin Koehler, armin.koehler@de.bosch.com

KI Data Tooling is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.



Supported by:



on the basis of a decision by the German Bundestag

www.ki-datatooling.de  @KI_Familie  KI Familie