

# IMPROVING SYNTHETIC DATA QUALITY FOR SAFE AI ENGINEERING APPLYING GENAI

Dr. Elena Hoemann, Dr.-Ing. Sven Hallerbach  
Institute for AI Safety & Security – AI Engineering

Conference Quantifying Simulation Quality

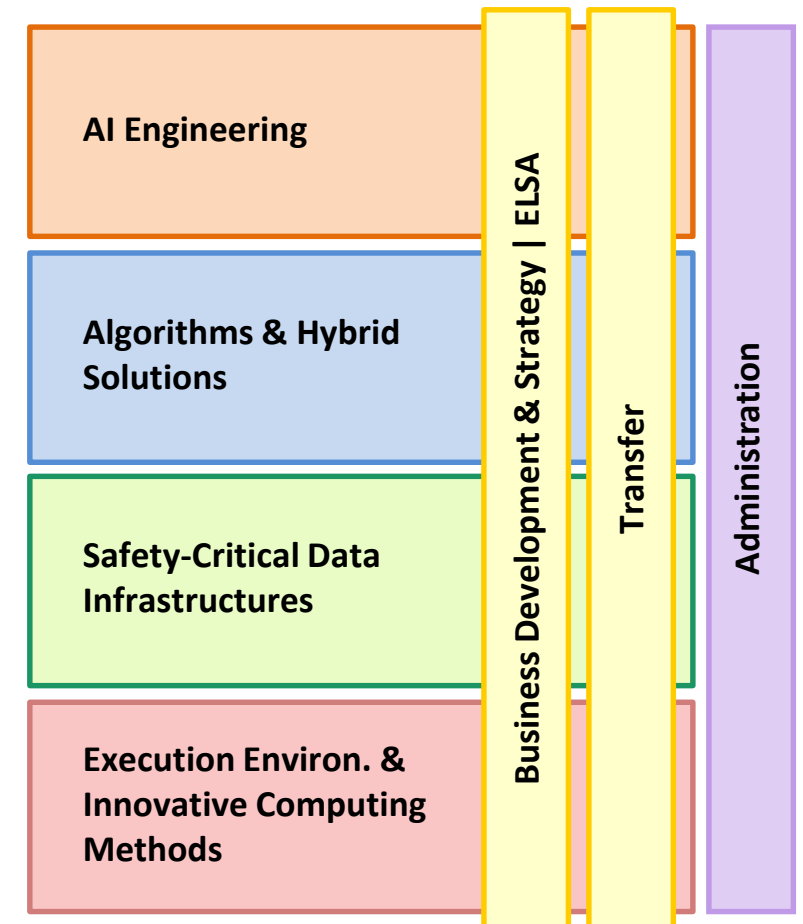
Session Verification of Tool Chains





## Integrated Management of **Safety & Security** during the **Design | Development & Implementation | Operation | Improvement** of **AI-based Solutions/Systems** for ambitious/advanced Applications.

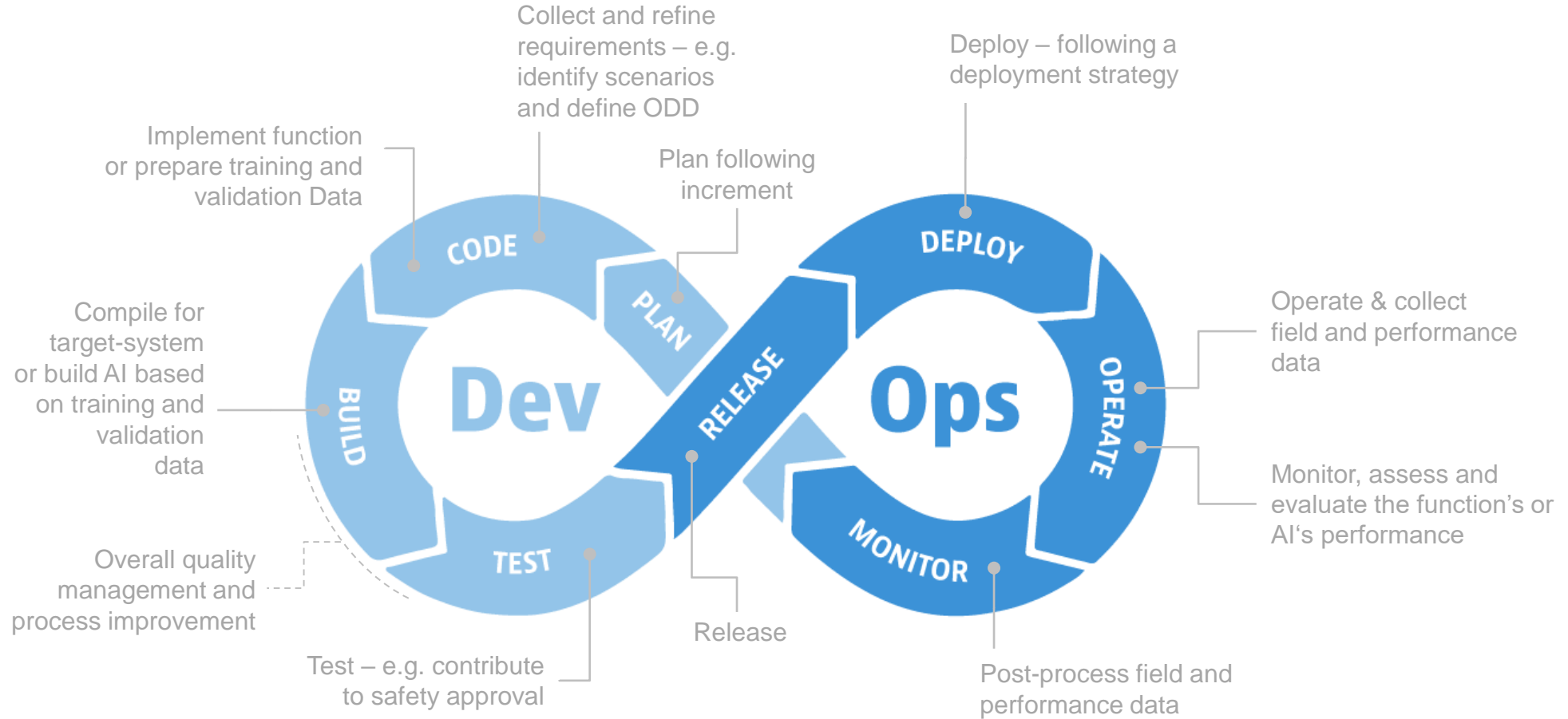
- **Safe and Standard-Compliant AI**
  - Guaranteeing safety and security as well as ensuring standard-compliant design of AI algorithms throughout their entire lifecycle.
  - Approval and certification of AI-based systems that are continuously updated or learn.
- **Cyber-Security and -Resilience for AI and Data/Service Ecosystems**
  - Secure management and trustful utilization of sensitive data in open data – e.g. in automotive and rail applications.
  - Protection of data, AI-components and -algorithms, as well as complex AI-based systems against cyberattacks.
- **Autonomous Mobility and Logistics Facilitation**
  - AI-based assistance and automated services to support the introduction and maintenance of advanced mobility and logistics solutions in the market as well as their scalability – e.g. for road, rail, and the maritime domain.





# Institute for AI Safety & Security

AI Engineering – Using the DevOps cycle to provide AI safety

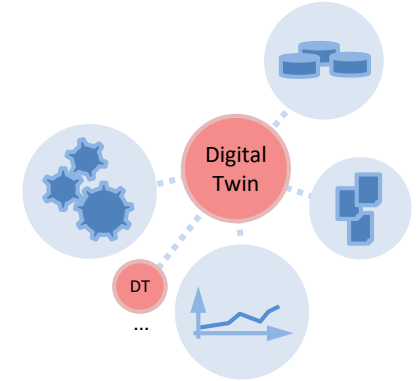
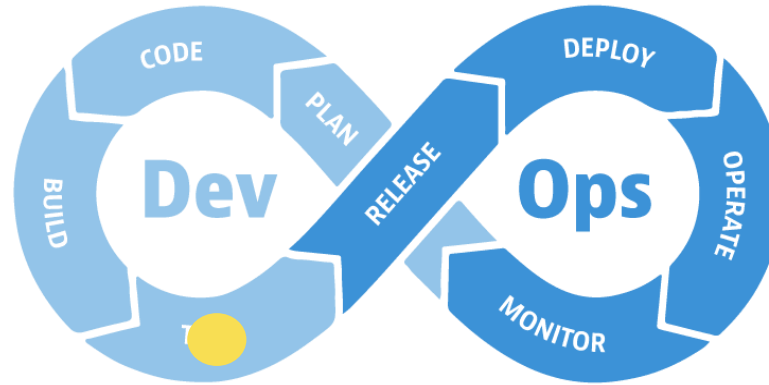
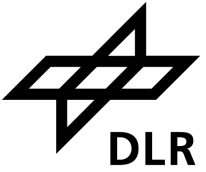


→ to provide AI safety



# Institute for AI Safety & Security

AI Engineering – Simulation-based Engineering | Processes & Itemization Core Dimensions



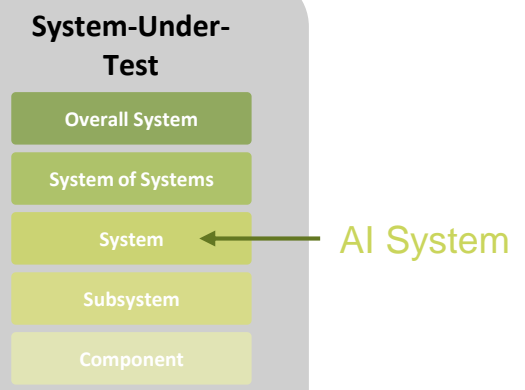
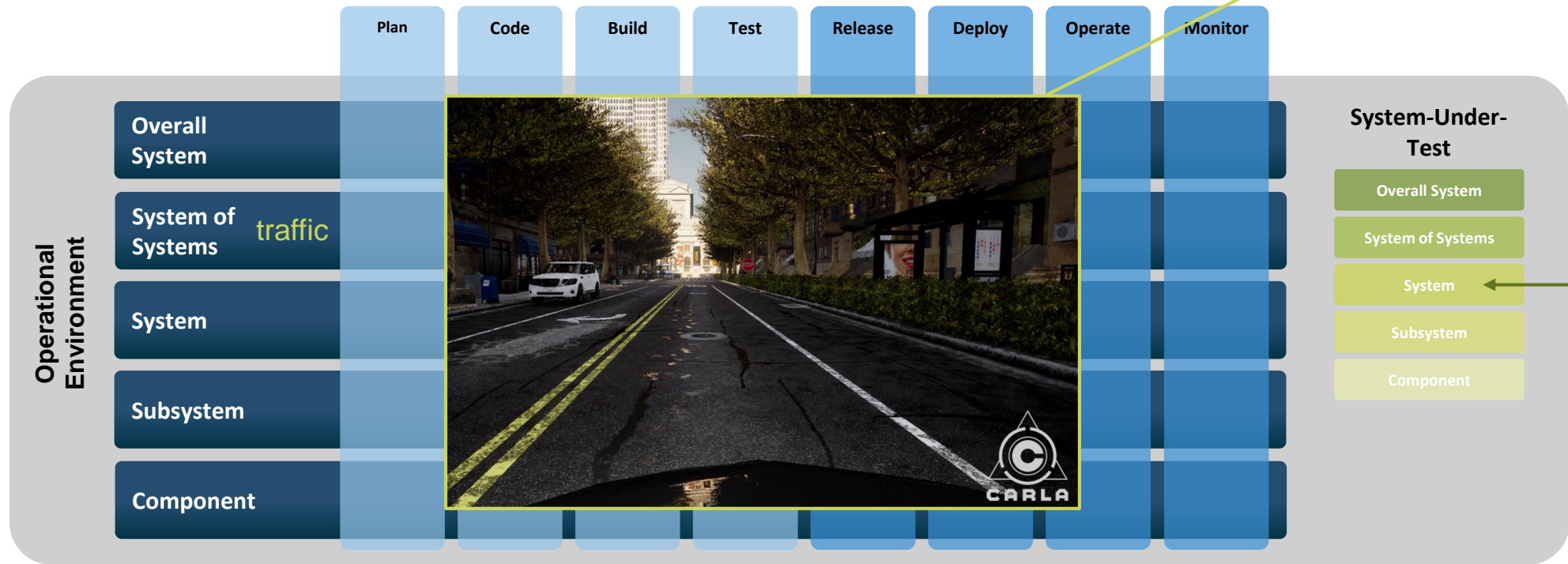
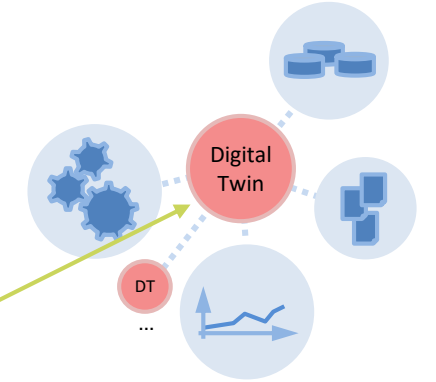
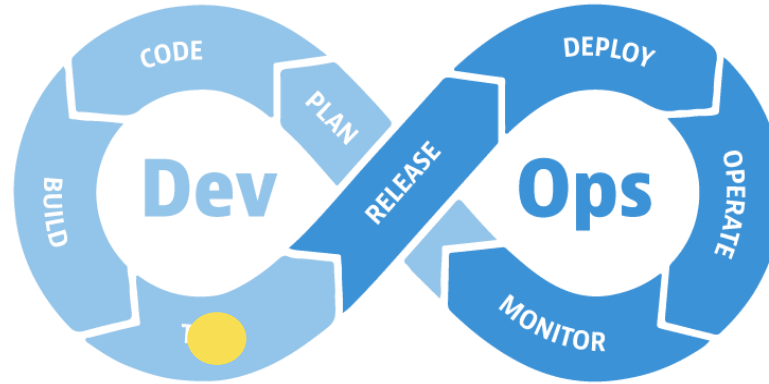
	Plan	Code	Build	Test	Release	Deploy	Operate	Monitor		
Operational Environment	Overall System									System-Under-Test
	System of Systems									
	System									
	Subsystem									
	Component									

- Overall System
- System of Systems
- System
- Subsystem
- Component



# Institute for AI Safety & Security

AI Engineering – Simulation-based Engineering | Processes & Itemization Core Dimensions





# Institute for AI Safety & Security

AI Engineering – How to generate data in the required quality?



Example: Automated driving

Motivation

Providing safe AI systems

Problem

Difficult to get sufficient training data

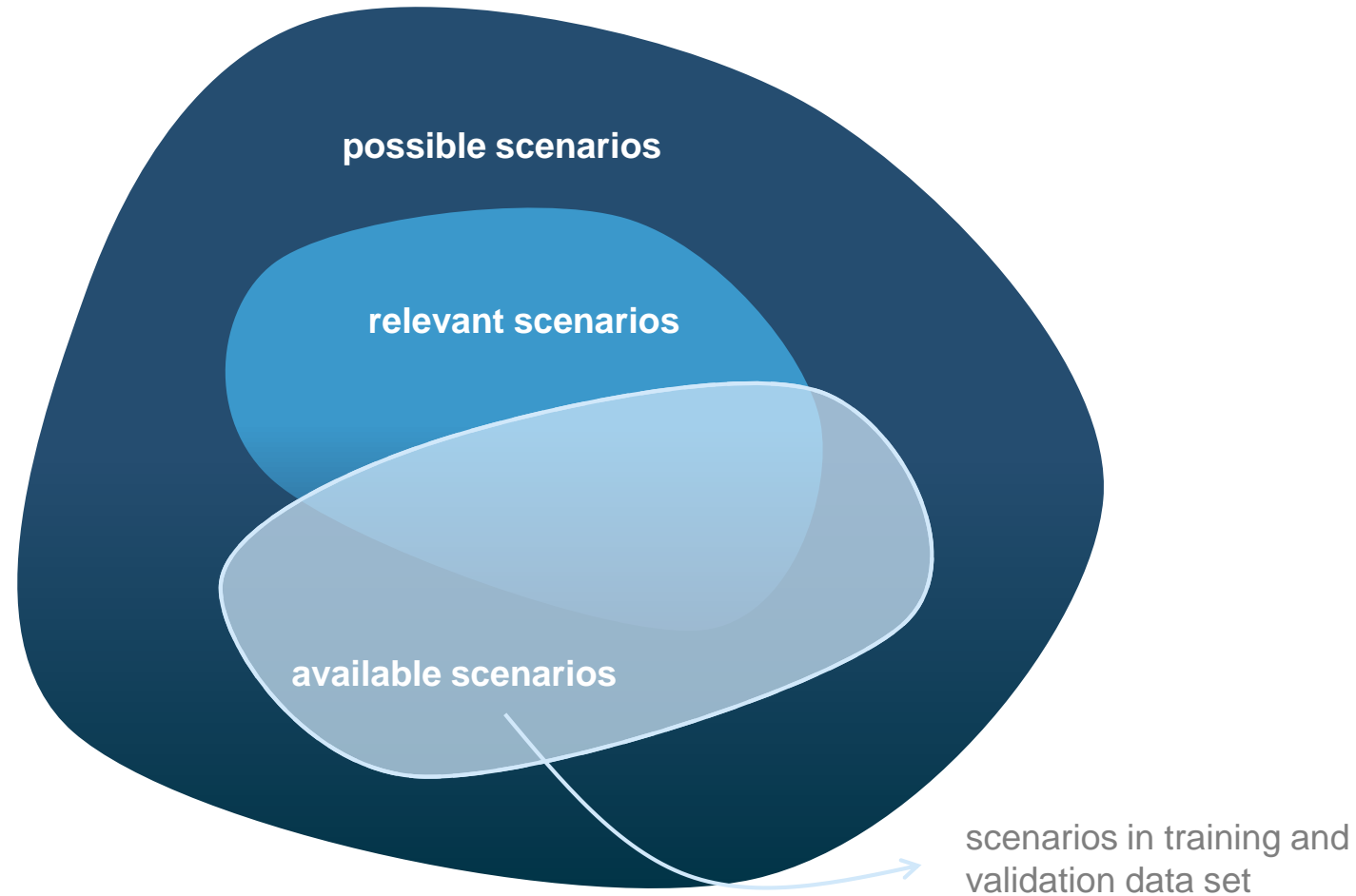
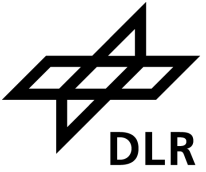
Question

How to generate data in the required quality?



# Institute for AI Safety & Security

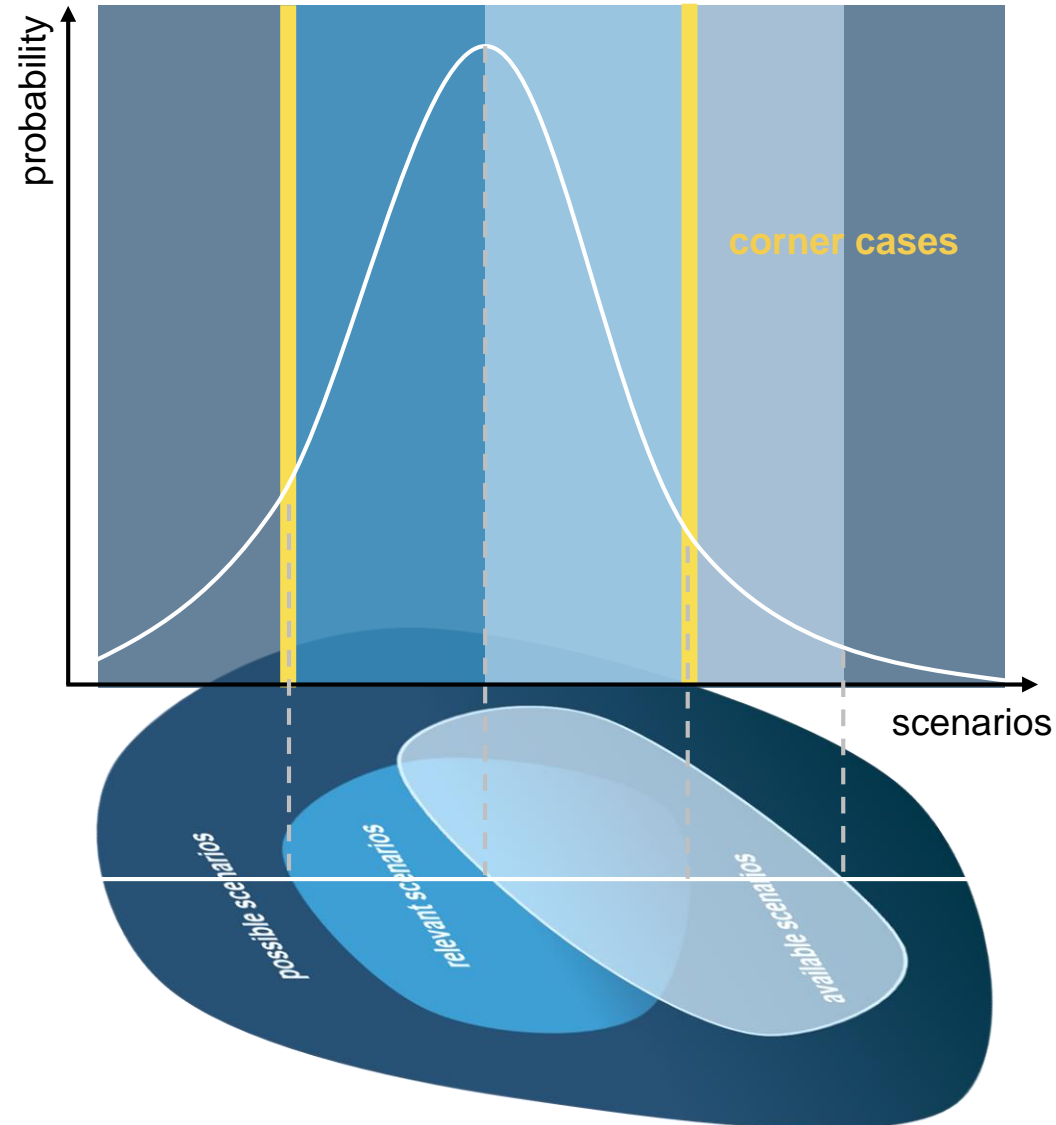
AI Engineering – Problem: discrepancy between relevant and available scenarios



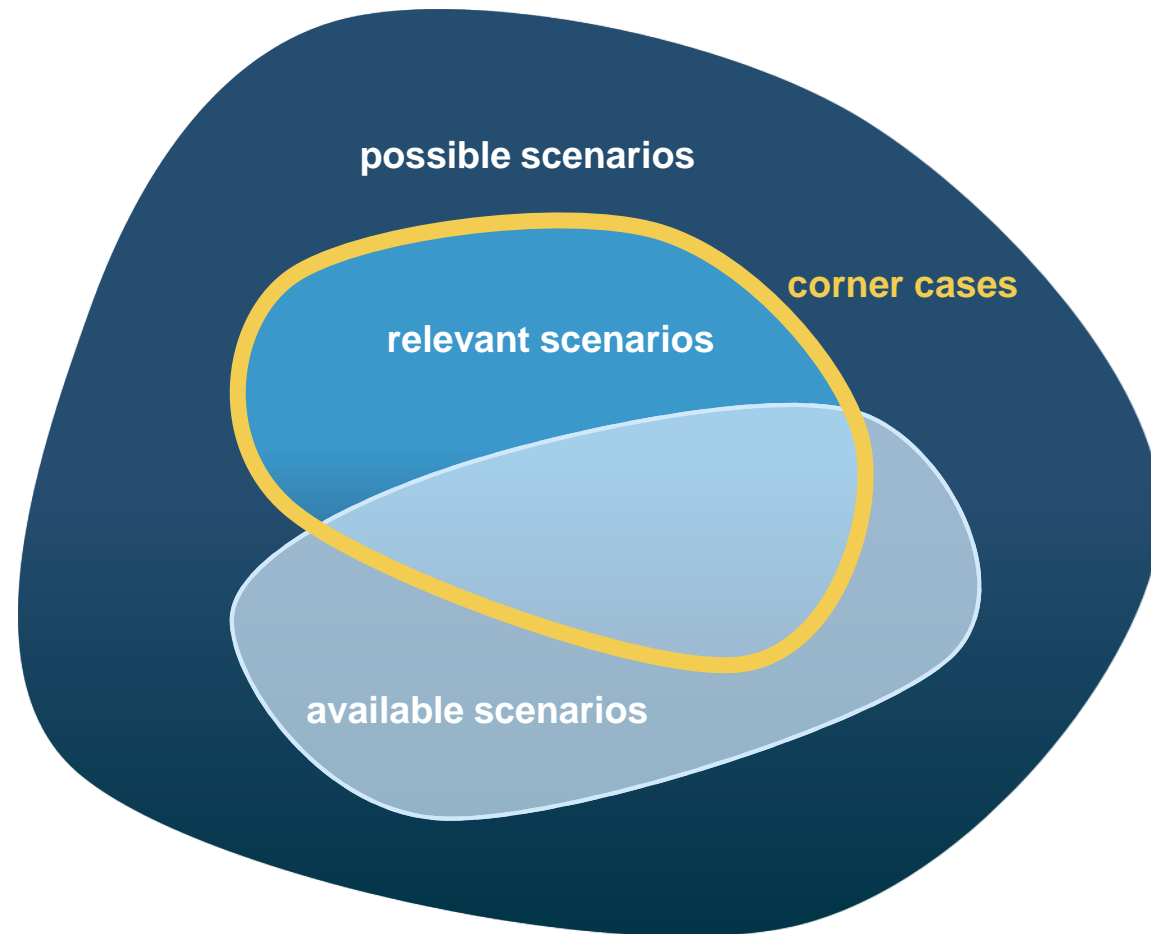
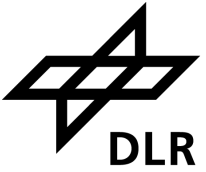


# Institute for AI Safety & Security

AI Engineering – Probability of scenario occurrence defines relevancy



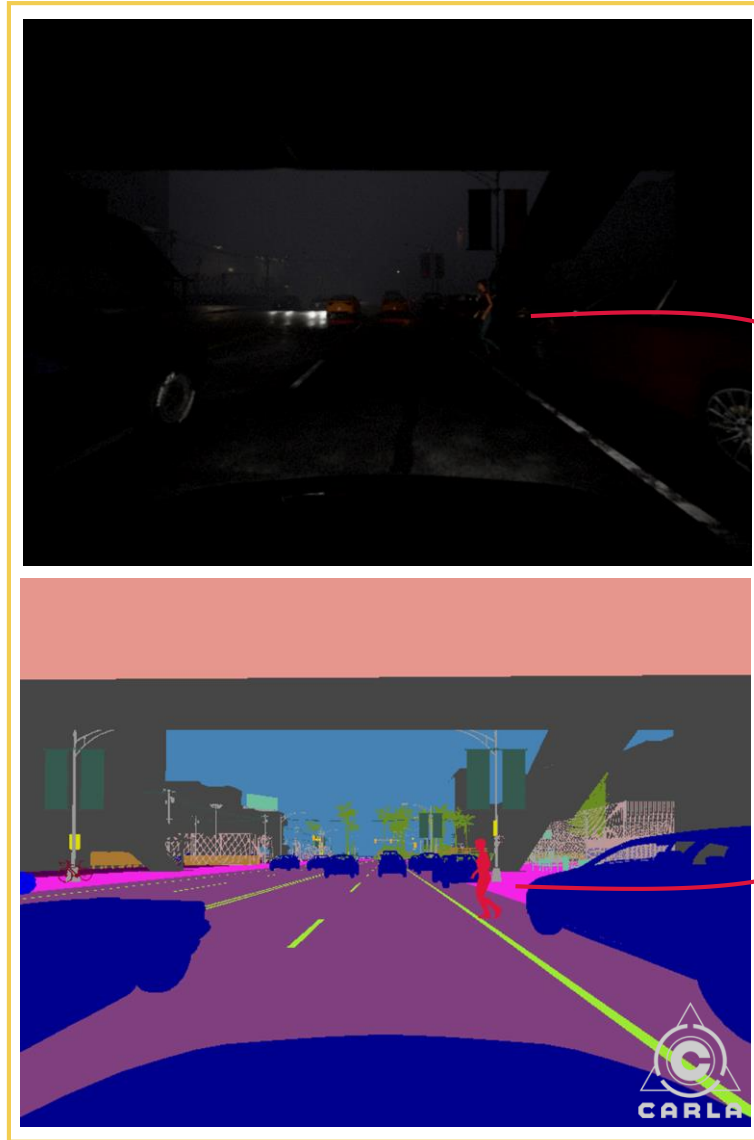
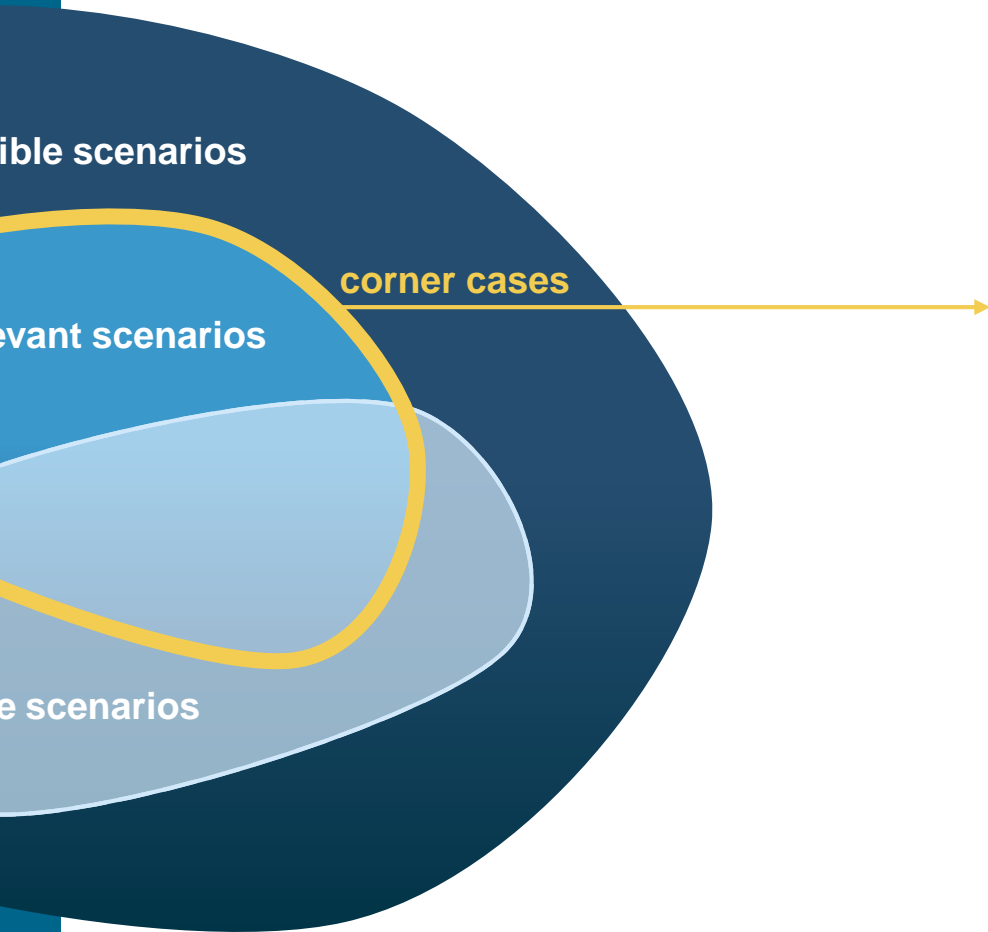
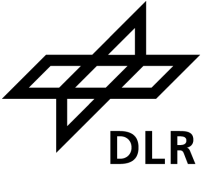






# Institute for AI Safety & Security

AI Engineering – Corner case example

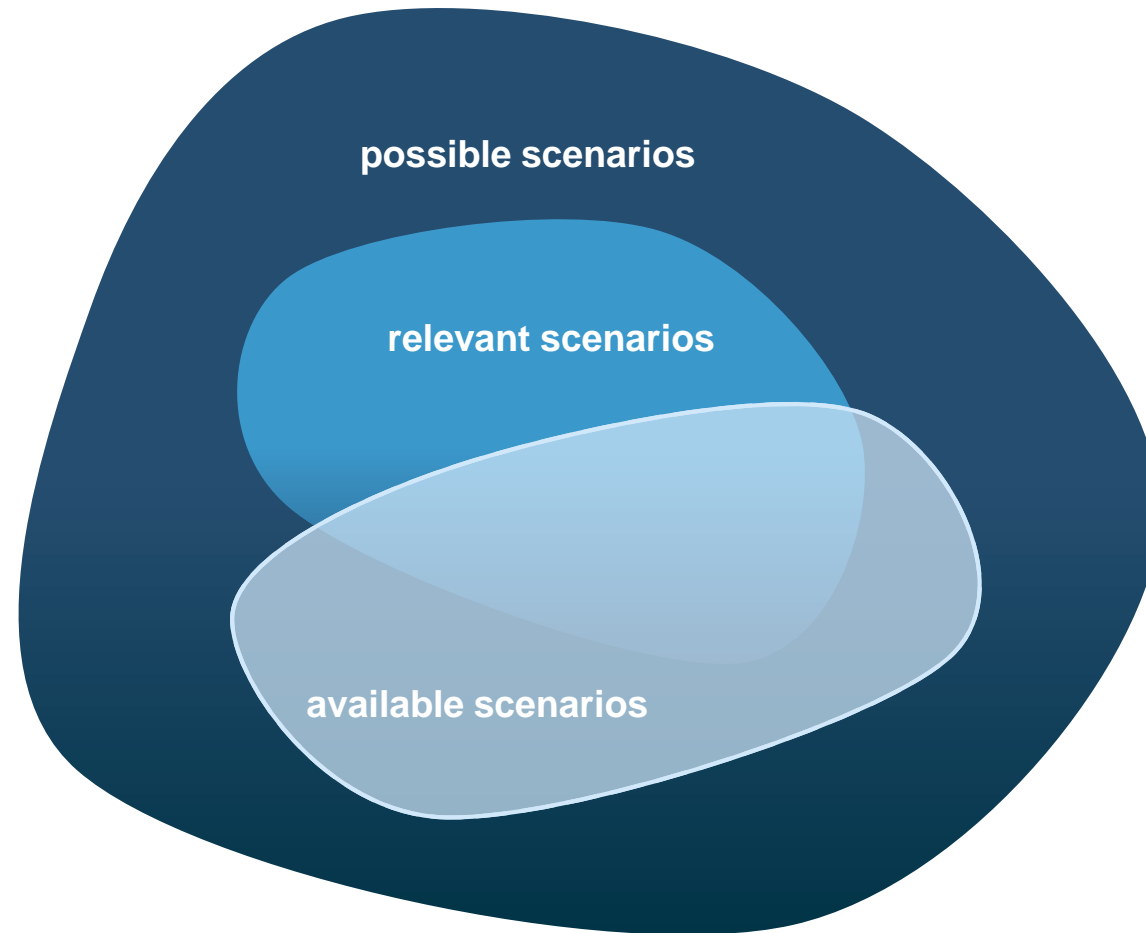


Pedestrian is suddenly crossing the road, nearly invisible due to bad lighting.



# Institute for AI Safety & Security

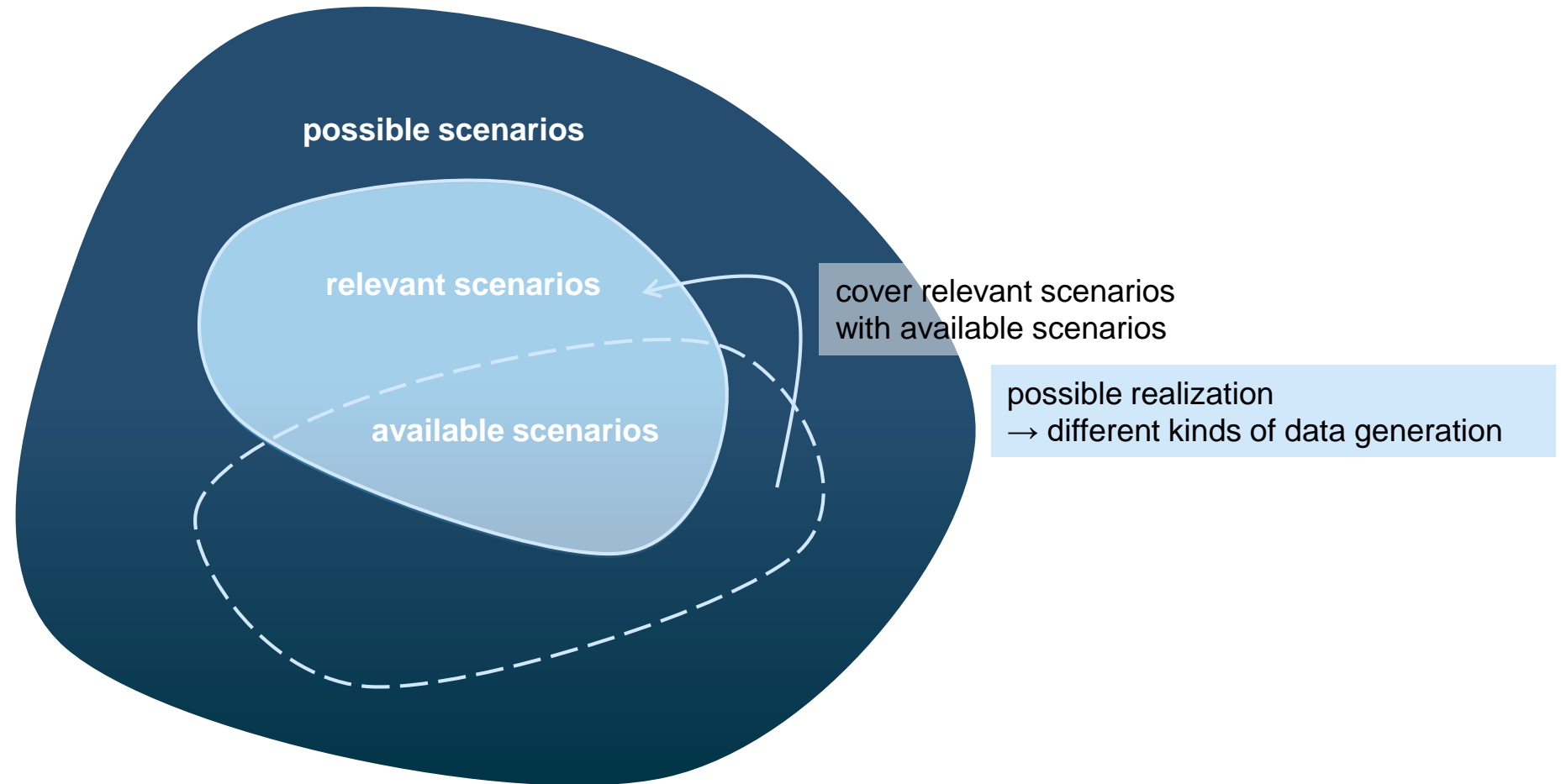
AI Engineering – Discrepancy between relevant scenarios and available scenarios





# Institute for AI Safety & Security

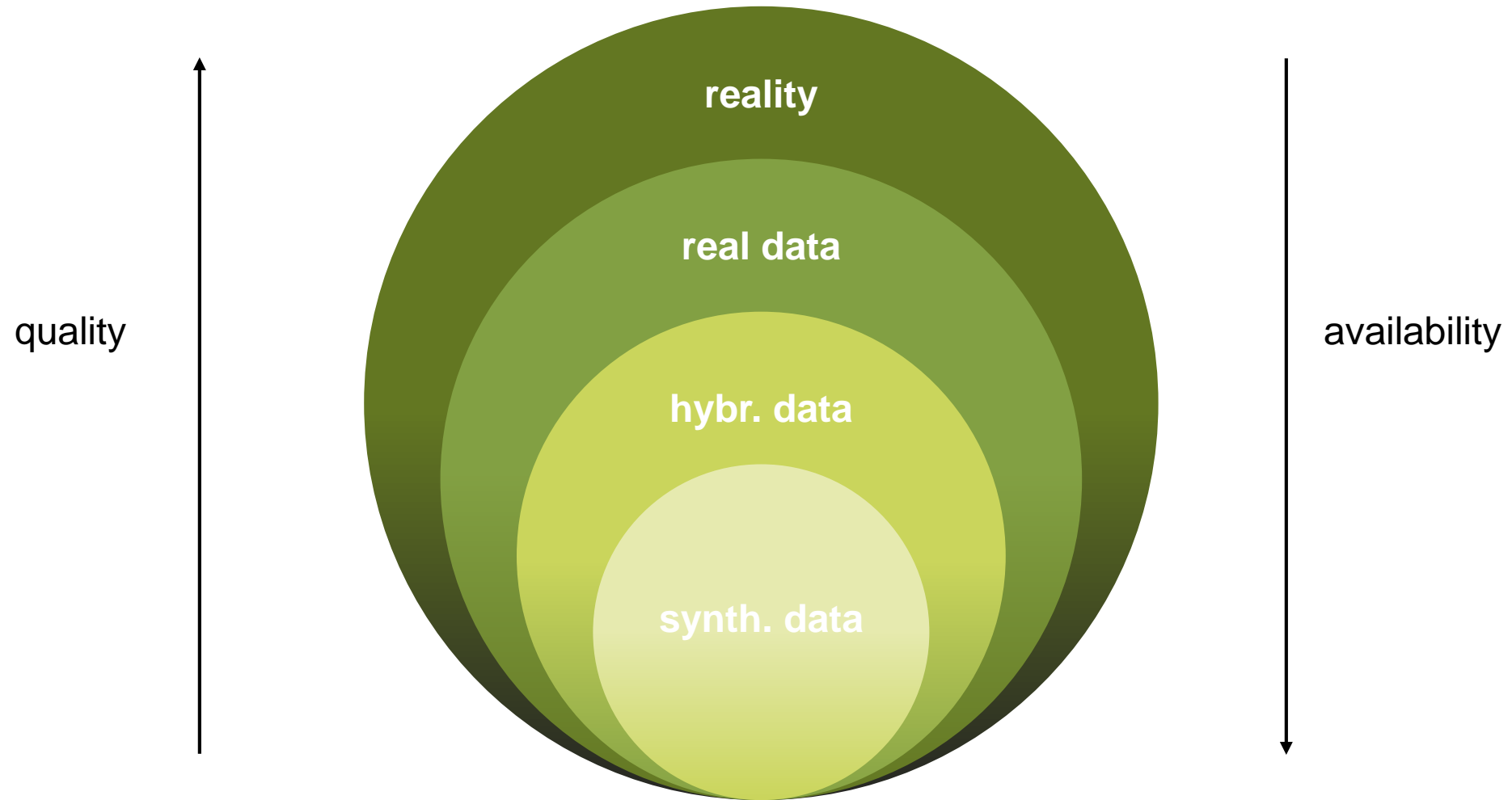
AI Engineering – Available scenarios need to cover relevant scenarios





# Institute for AI Safety & Security

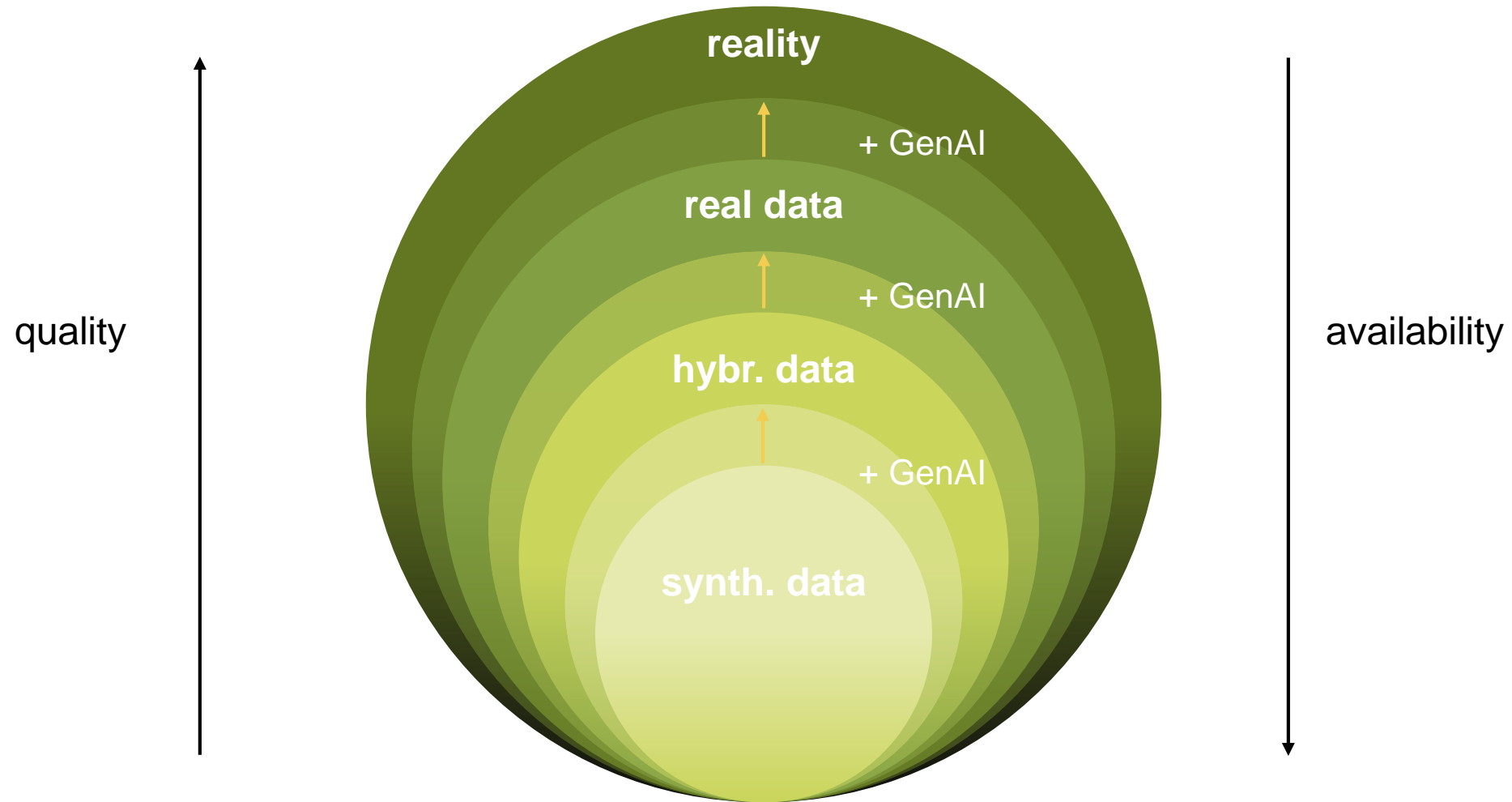
AI Engineering – Trade-off between data quality and accesability





# Institute for AI Safety & Security

AI Engineering – GenAI can close the gap to more data quality





GenAI may be used to enhance datasets by

- Refining images to make them more **realistic**
- Applying **domain adaptation**
- Increase **variation**
- ...



Query



Night



Day



Rain



Sunset



Winter



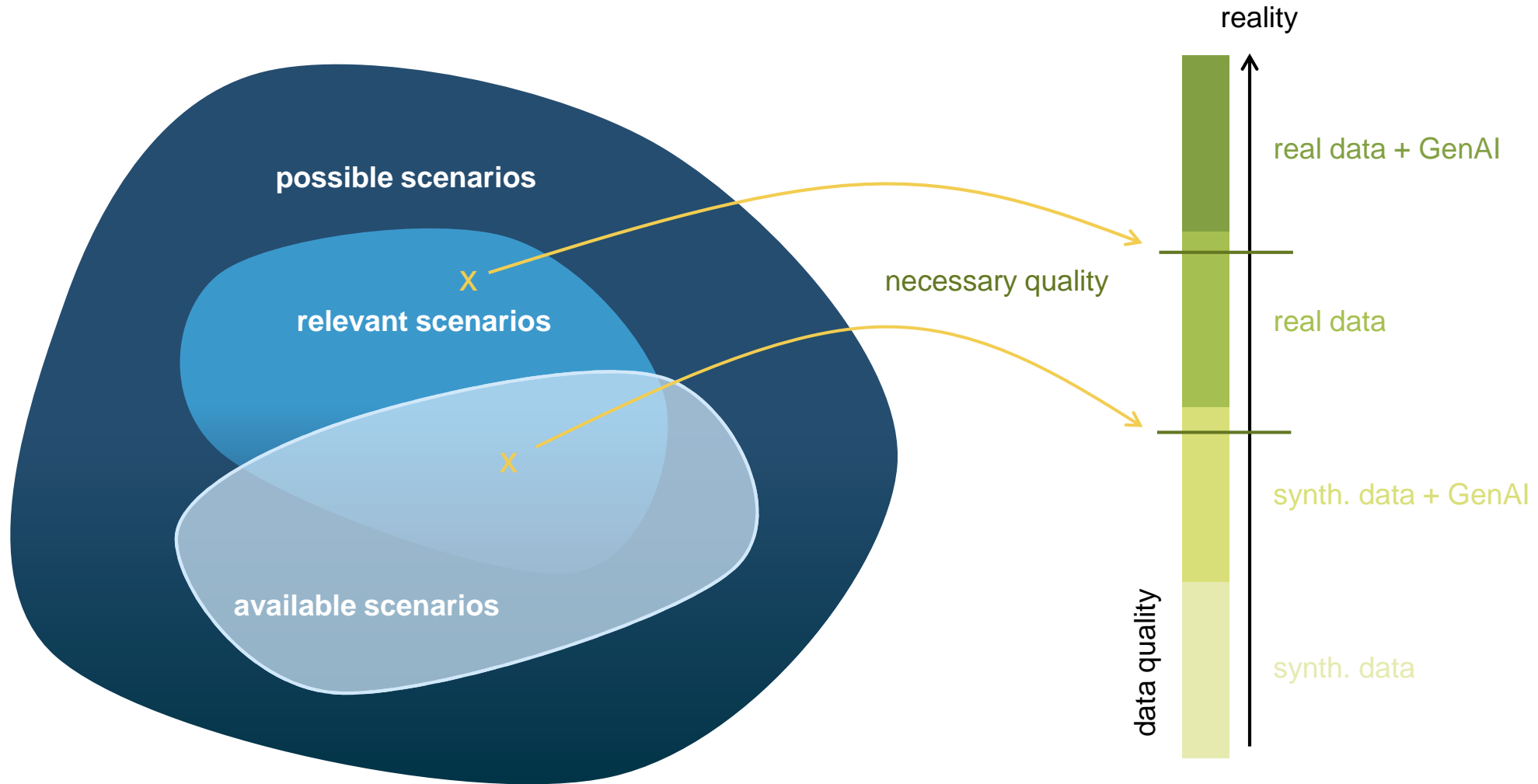
Autumn

[Y. Jila et al.,2024, [arXiv:2312.03048](https://arxiv.org/abs/2312.03048), CC BY-NC-SA 4.0]



# Institute for AI Safety & Security

AI Engineering – For every scenario a necessary data quality needs to be provided



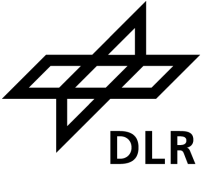




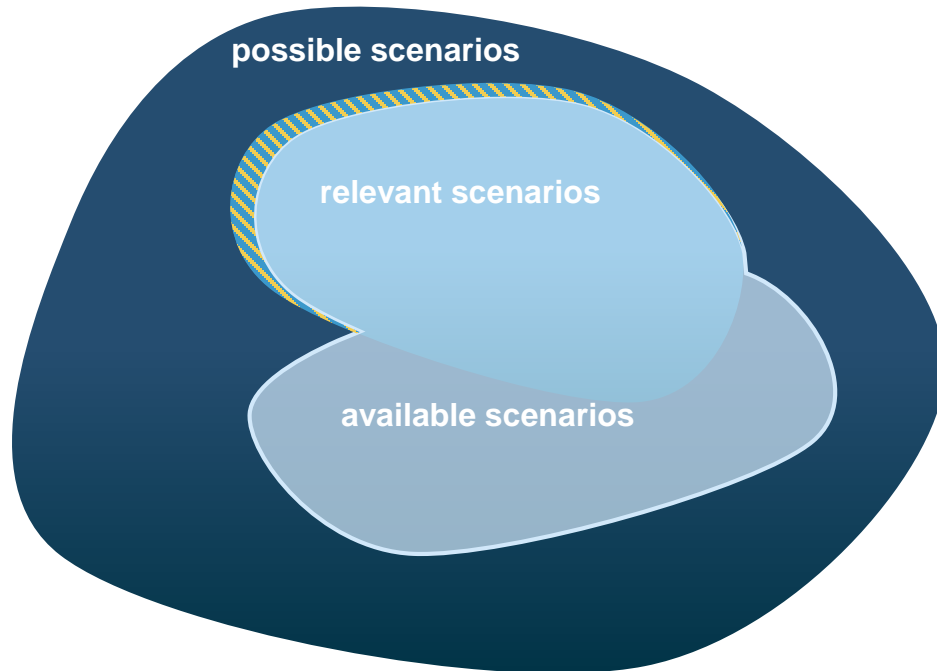
**scenario plane**

**quality plane**

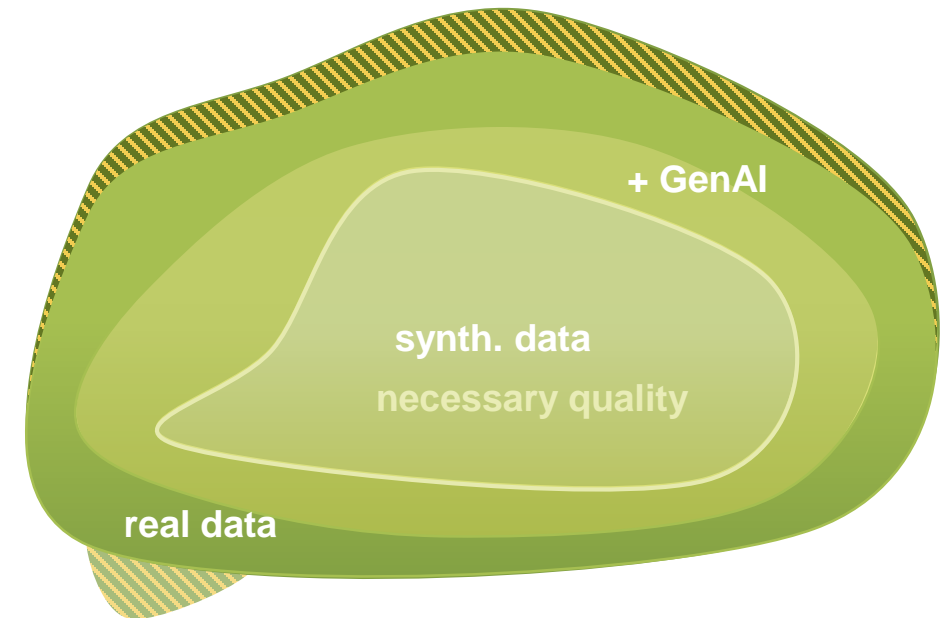




more **scenarios** needed



different **quality** needed



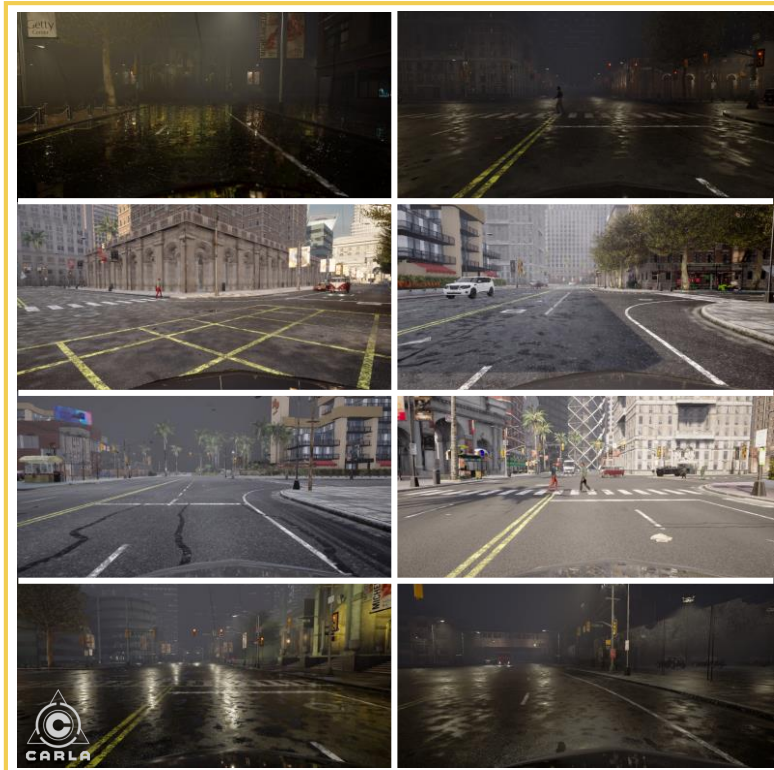


# Institute for AI Safety & Security

AI Engineering – Reaction to insufficient coverage of scenarios/quality



more **scenarios** needed



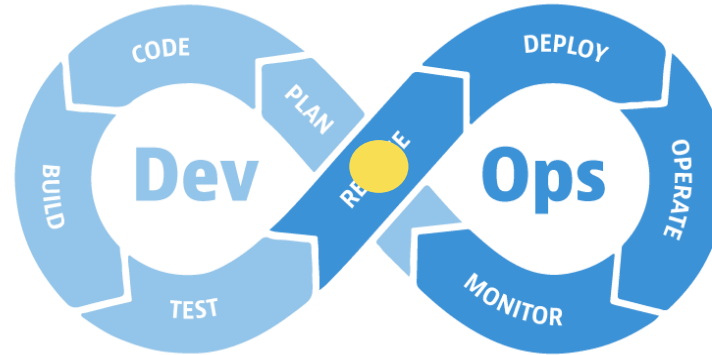
different **quality** needed



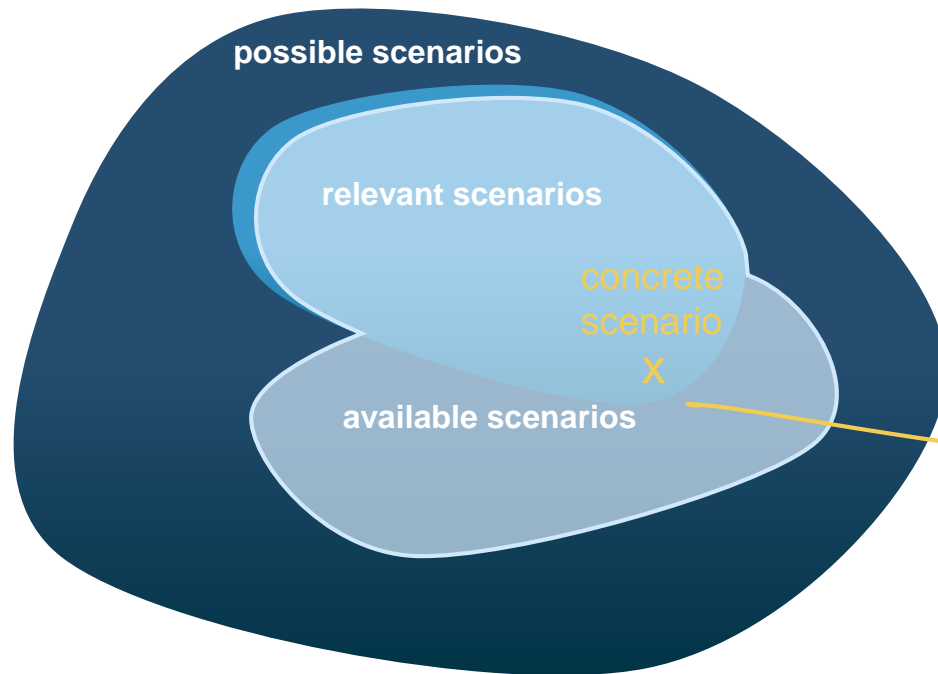
possible scenarios

relevant scenarios

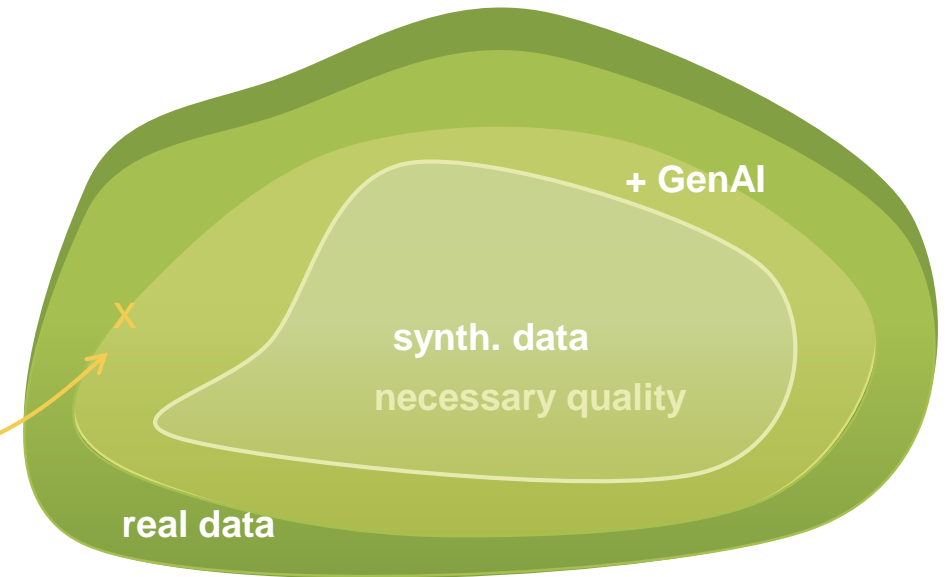
+ GenAI



## scenario plane



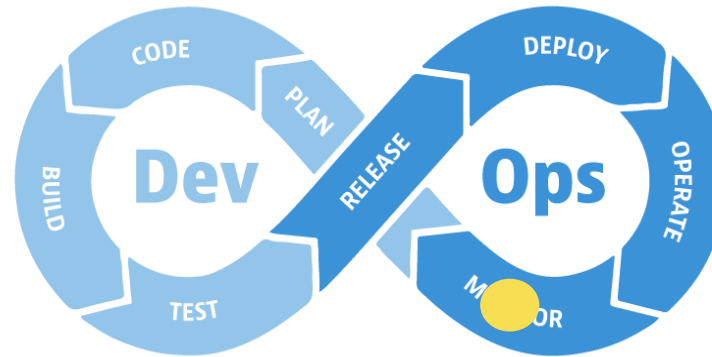
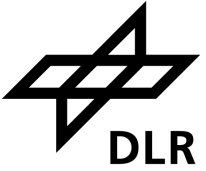
## quality plane



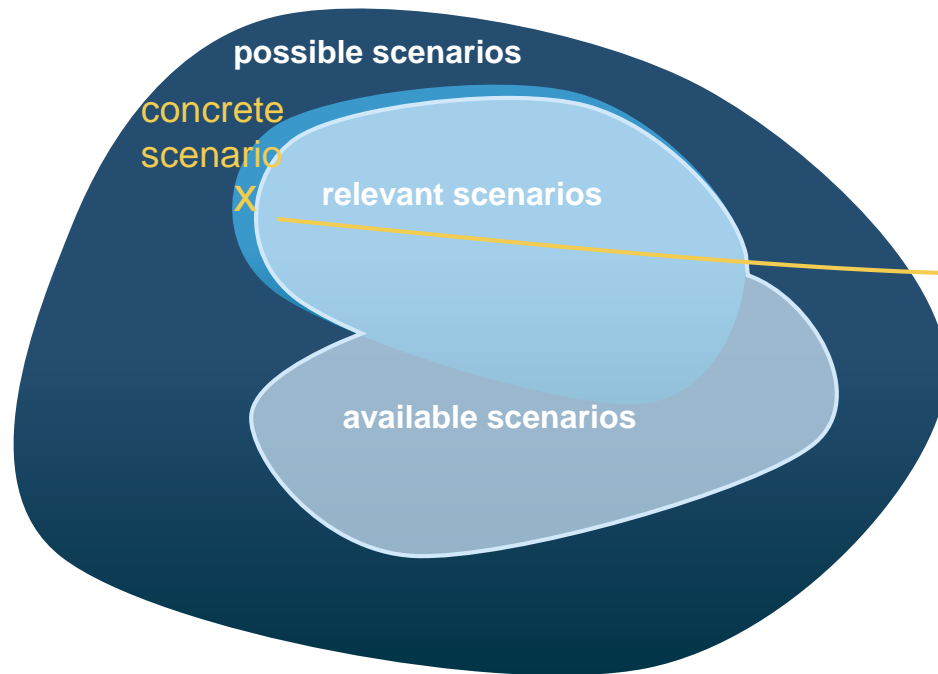


# Institute for AI Safety & Security

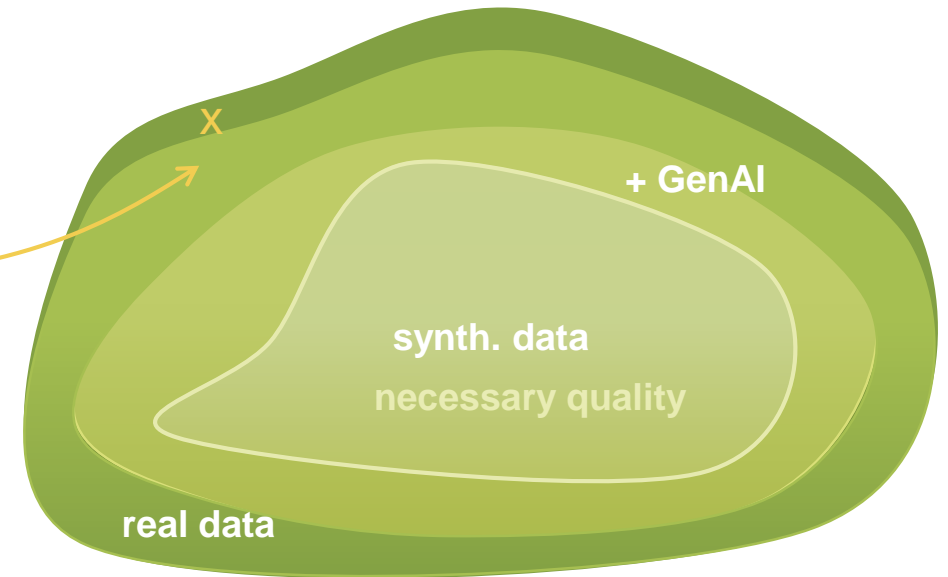
AI Engineering – Monitor operation to update dataset



## scenario plane



## quality plane



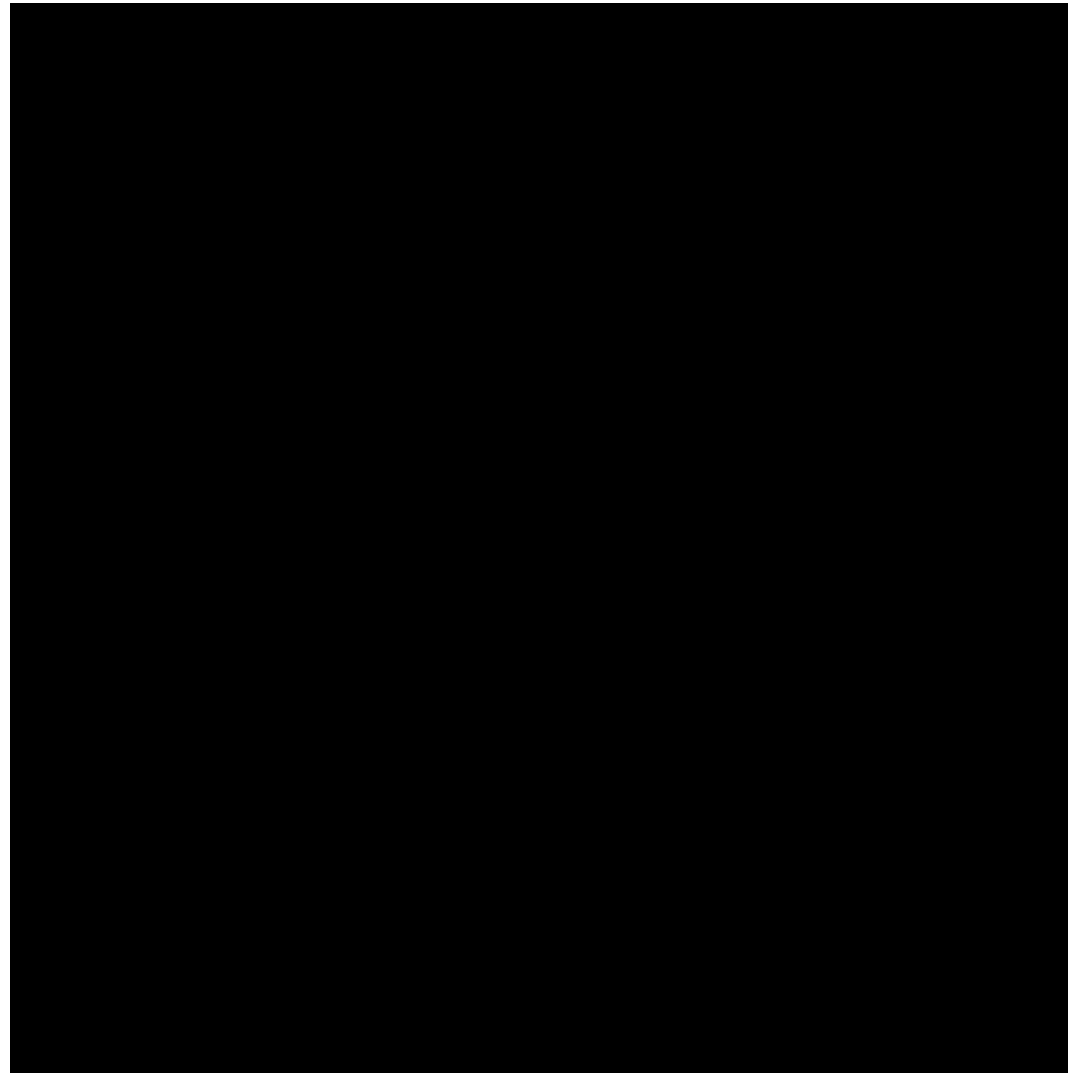
update





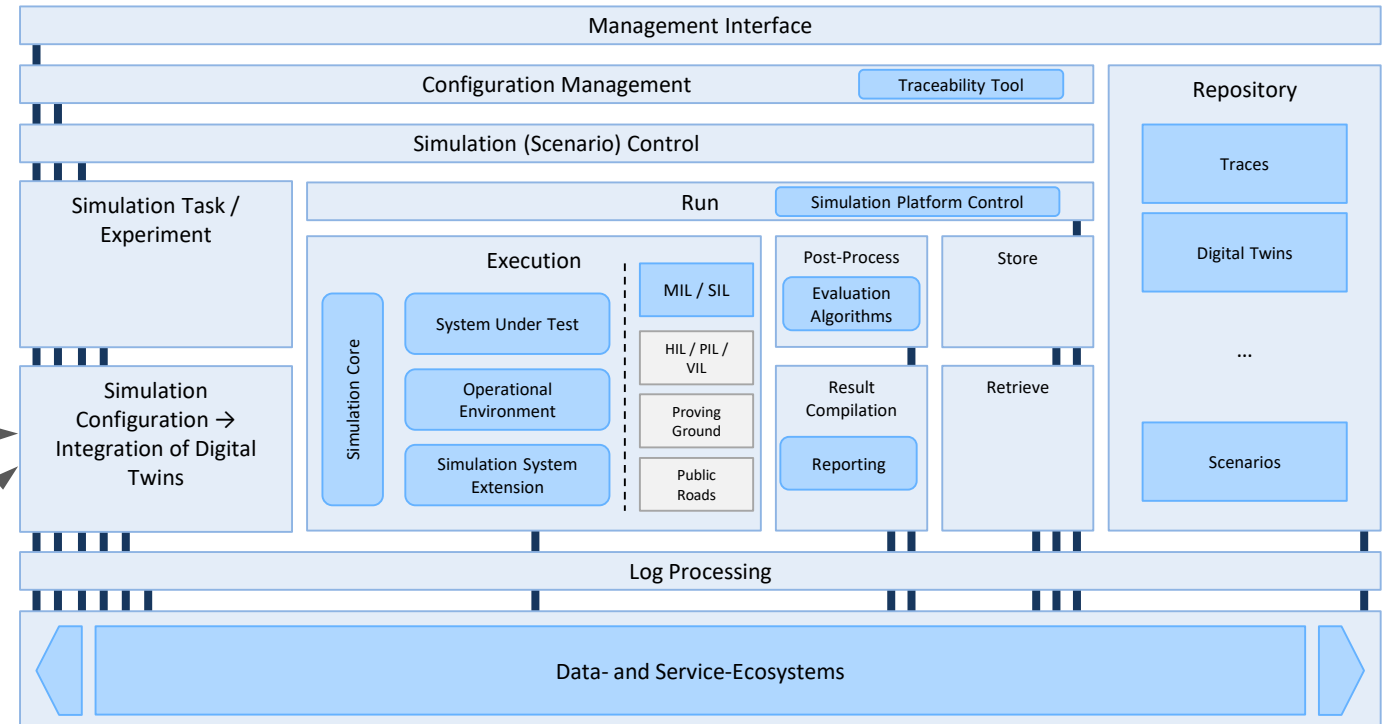
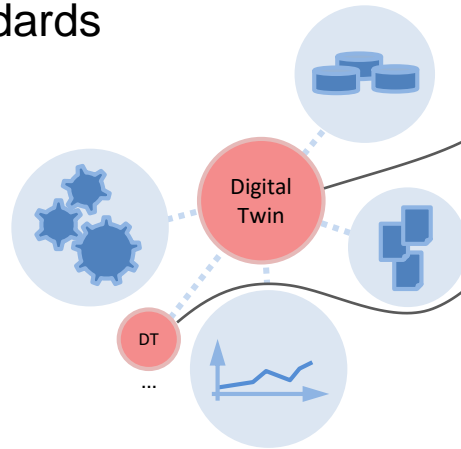
# Institute for AI Safety & Security

AI Engineering – Simulation example





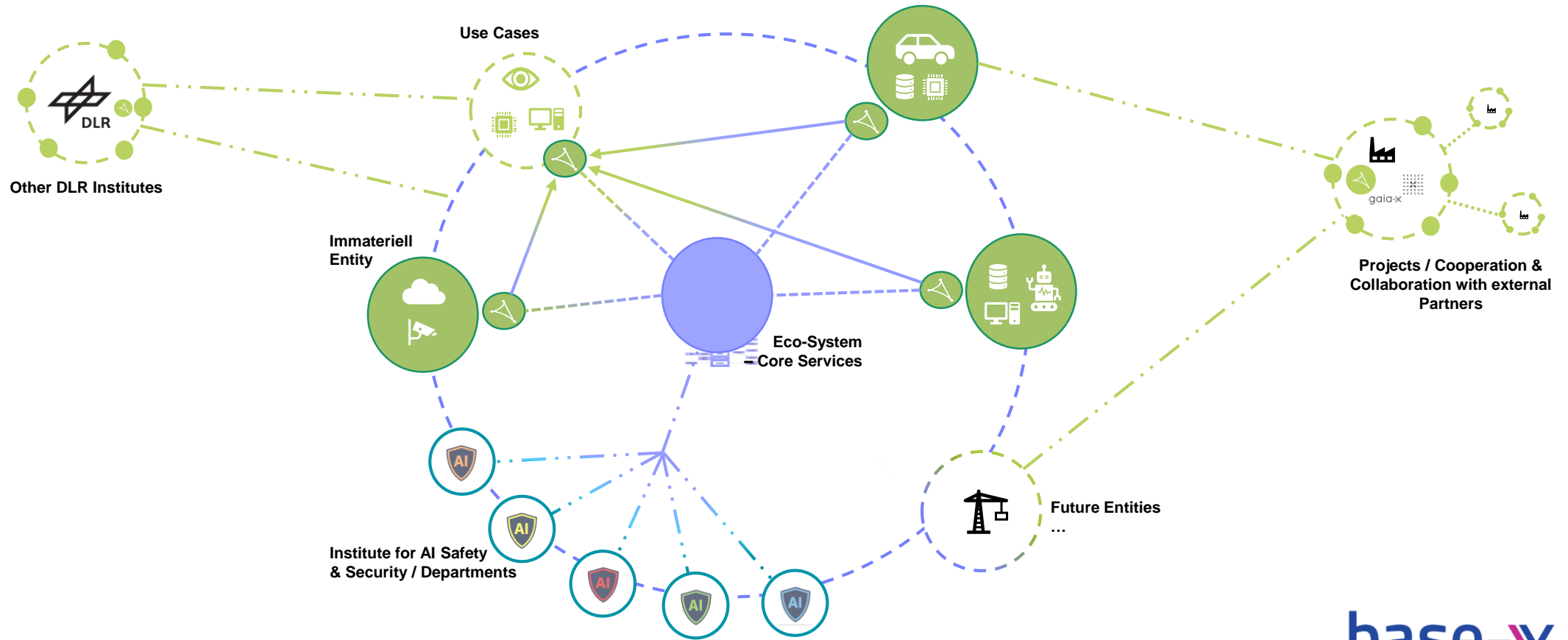
- Generic Open Testing Architecture
- Setup based on Data- and Service-Ecosystems
- Ensures Modularity
- Independent: Domain / System Modeling / Execution Environment
- Modules based on Standards





# Institute for AI Safety & Security

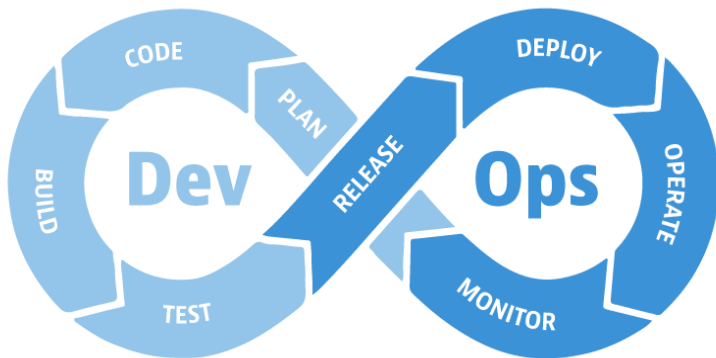
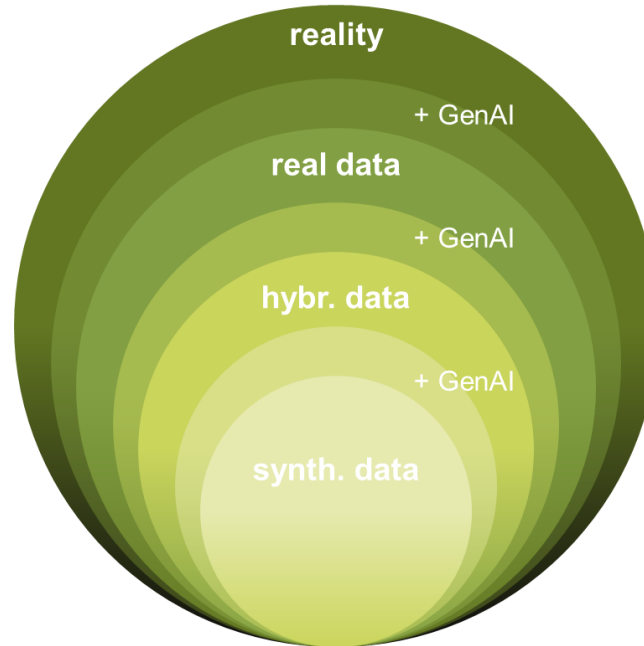
AI Engineering – Simulation-Based Engineering Embedded in Data Spaces and Service-Ecosystems







Combine different techniques to gain sufficient data quality.



Use DevOps to continuously improve your data coverage.

Ongoing: Implementation in CARLA





## Contact

Dr. Elena Hoemann  
German Aerospace Center (DLR)  
Institute for AI Safety & Security  
Sankt Augustin and Ulm  
Germany  
[Elena.Hoemann@dlr.de](mailto:Elena.Hoemann@dlr.de)

Dr.-Ing. Sven Hallerbach  
German Aerospace Center (DLR)  
Institute for AI Safety & Security  
Sankt Augustin and Ulm  
Germany  
[Sven.Hallerbach@dlr.de](mailto:Sven.Hallerbach@dlr.de)