AI for Advanced Applications: Safety and Security is a Must

DLR Research Institute in Sankt Augustin and Ulm

Prof. Dr. Frank Köster

AI@DLR

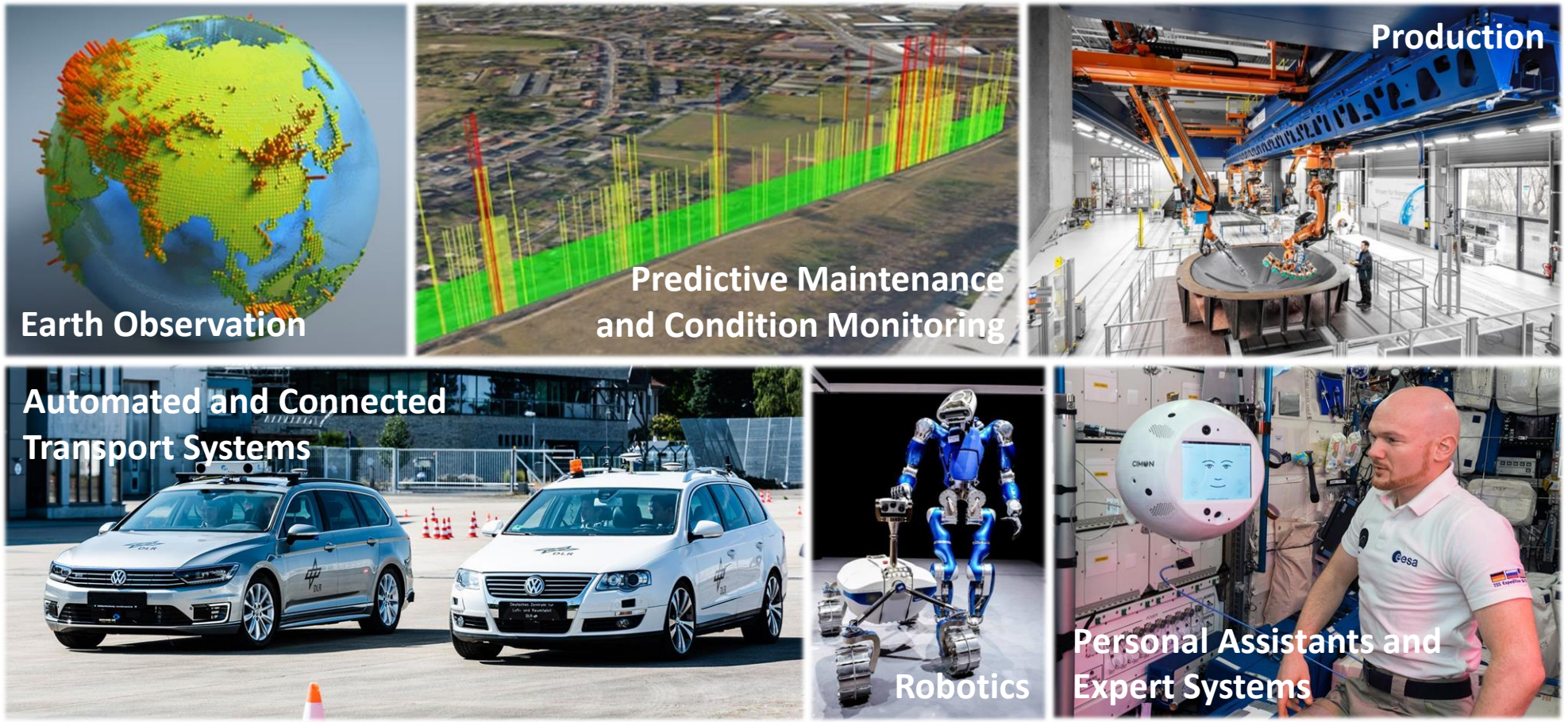


Fig.: DLR

Research for AI Safety and Security – Motivation

Fig.: DLR

Institute for AI Safety and Security

Sankt Augustin and Ulm – Germany – www.dlr.de/ki

Aeronautics | Space | Energy | Transport | Safety | Digitalisation

- AI Engineering
- Algorithms & Hybrid Solutions
- Safety-Critical Data Infrastructures
- Execution-Environments & Innovative Computing Methods

Ethical, Legal & Social Aspects

Science | Industry | Society | Politics

Institute for AI Safety and Security

AI Engineering

Algorithms & Hybrid Solutions

Safety-Critical Data Infrastructures

Execution Environ. & Innov. Comp. Methods

Business Development & Networks

Processes, Methods and Tools

Operation

Test and Approval

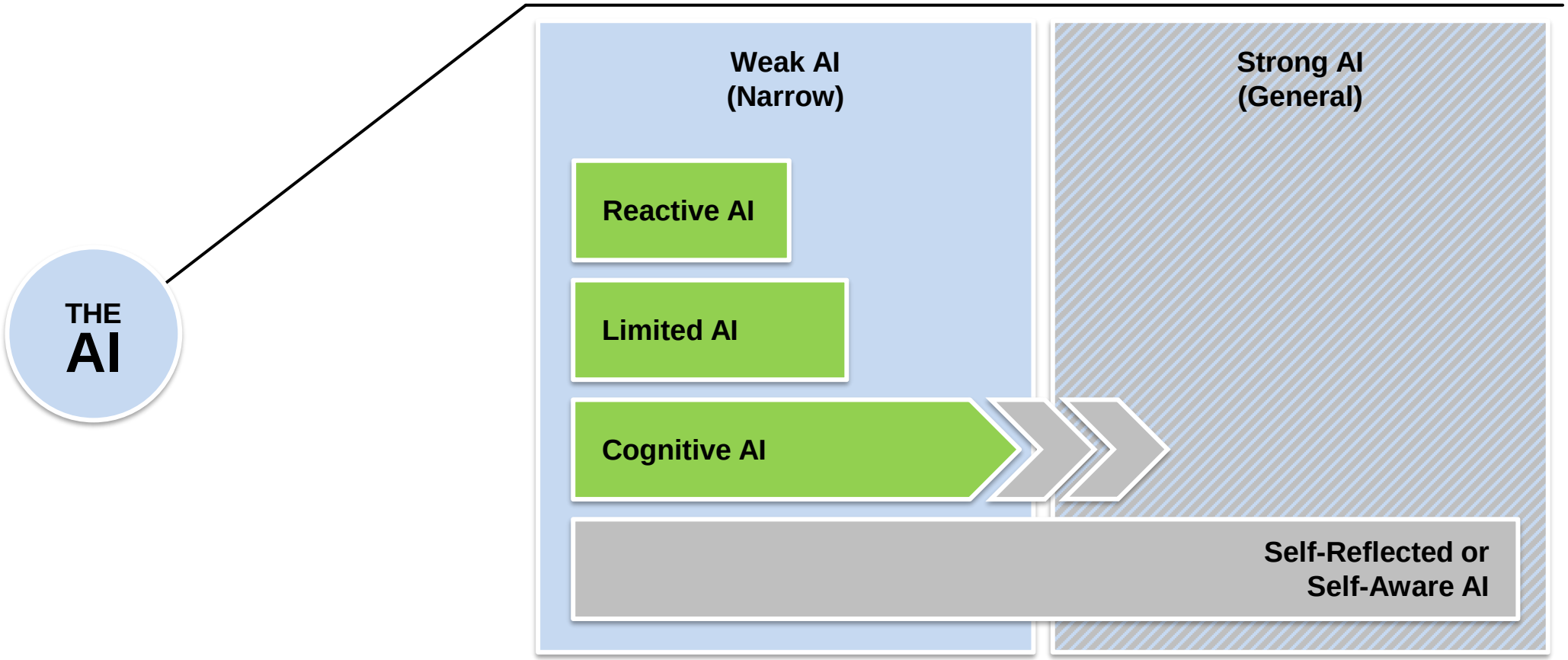
Adequate Data Basis

Trusted and Distributed Data Infrastructures

Ethical, Legal and Social Aspects & Technology Assessment

System Architectures

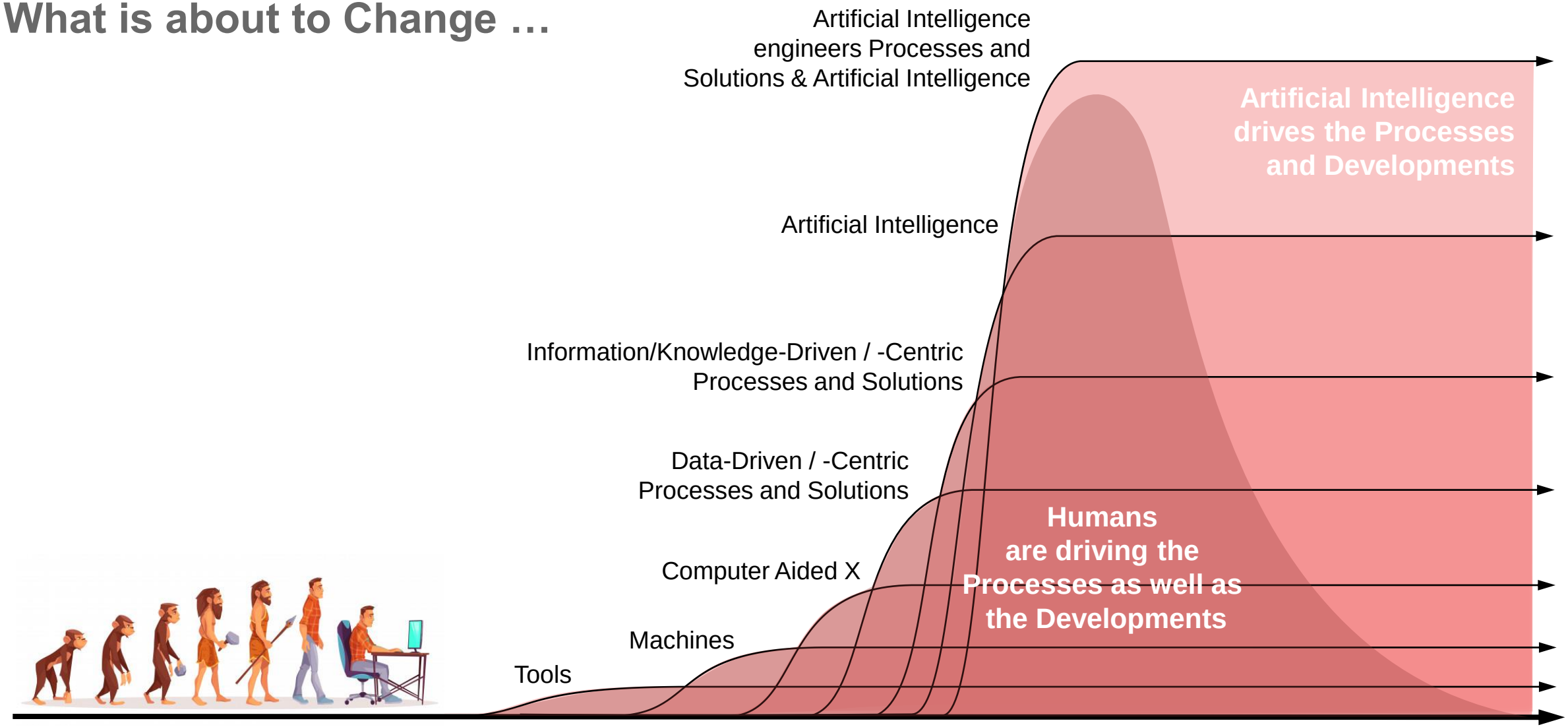
Generic AI-based Functionalities – Types of AI



Generic AI-based Functionalities

- Pattern extraction from data as well as pattern recognition
- Identification of cause and effect relationships and functional relationships
- Classification and categorization of data as well as cluster analysis
- Analysis of networks or graph-structures
- Analysis of image data to e.g. extract and classify objects
- Analysis of video data with a focus on e.g. object extraction, -classification and -dynamics
- Analysis of audio data with regard to language, music and mood/sentiment etc.
- Full text analysis with regard to central statements, lines of argument and mood/sentiment etc.
- Prediction of the development of a system – e.g. object movements and development of patterns
- Action and movement planning as well as control on different system levels
- Manipulation of objects
- Integration and explication of "knowledge"
- …

What is about to Change …



Automated and Connected Driving

Level 0
no Automation

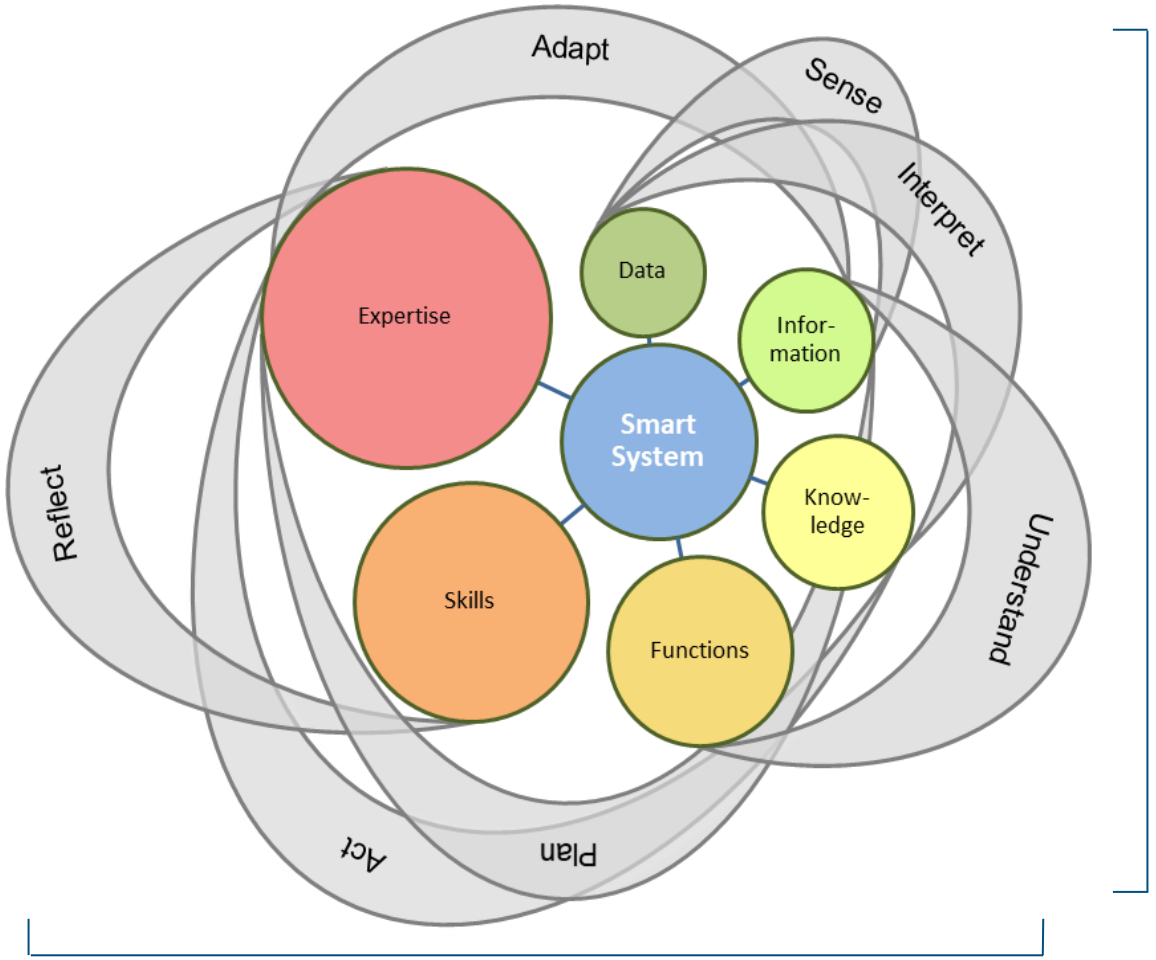
Level 1
Assisted

Level 2
Partial Automation

Level 3
High Automation

Level 4
Full Automation

Level 5
Autonomous



Different Deployment Strategies
and Use-Cases

no cooperation	Level a
provision of function-specific data / information (i.e. handling of data/information by recipient remains open; use without explicit feedback to sender)	Level b
Level b + integration into receiver's situational picture and feedback to sender	Level c
Level c + coop. developm. of a situational picture with the aim of a common situational picture (possibly including joint interpretation or plausibility check)	Level d
Level d + cooperative planning with a fixed goal structure (possibly a differentiation regarding driving tasks could be useful (strategical / tactical / operational))	Level e
Level d + cooperative planning with a flexible goal structure (possibly a differentiation regarding driving tasks could be useful (strategical / tactical / operational))	Level e*

Automated and Connected Driving – Safety & Security by Design

Level 0
no Automation

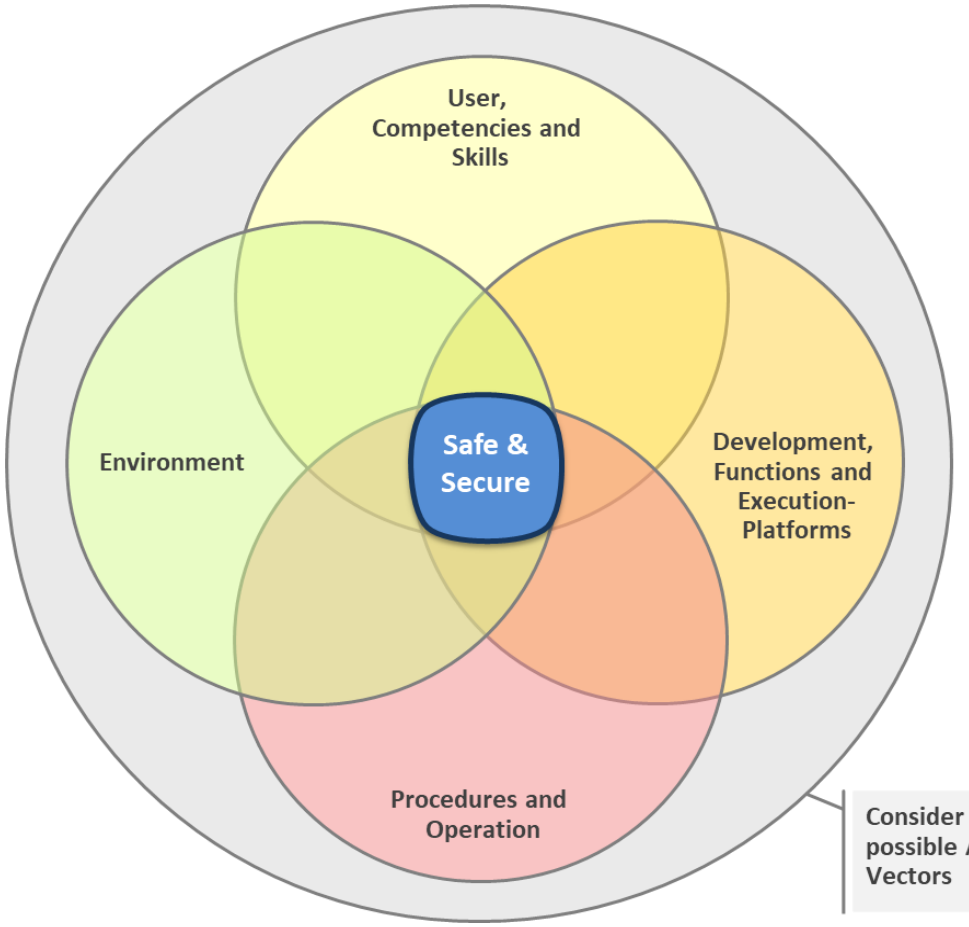
Level 1
Assisted

Level 2
Partial Automation

Level 3
High Automation

Level 4
Full Automation

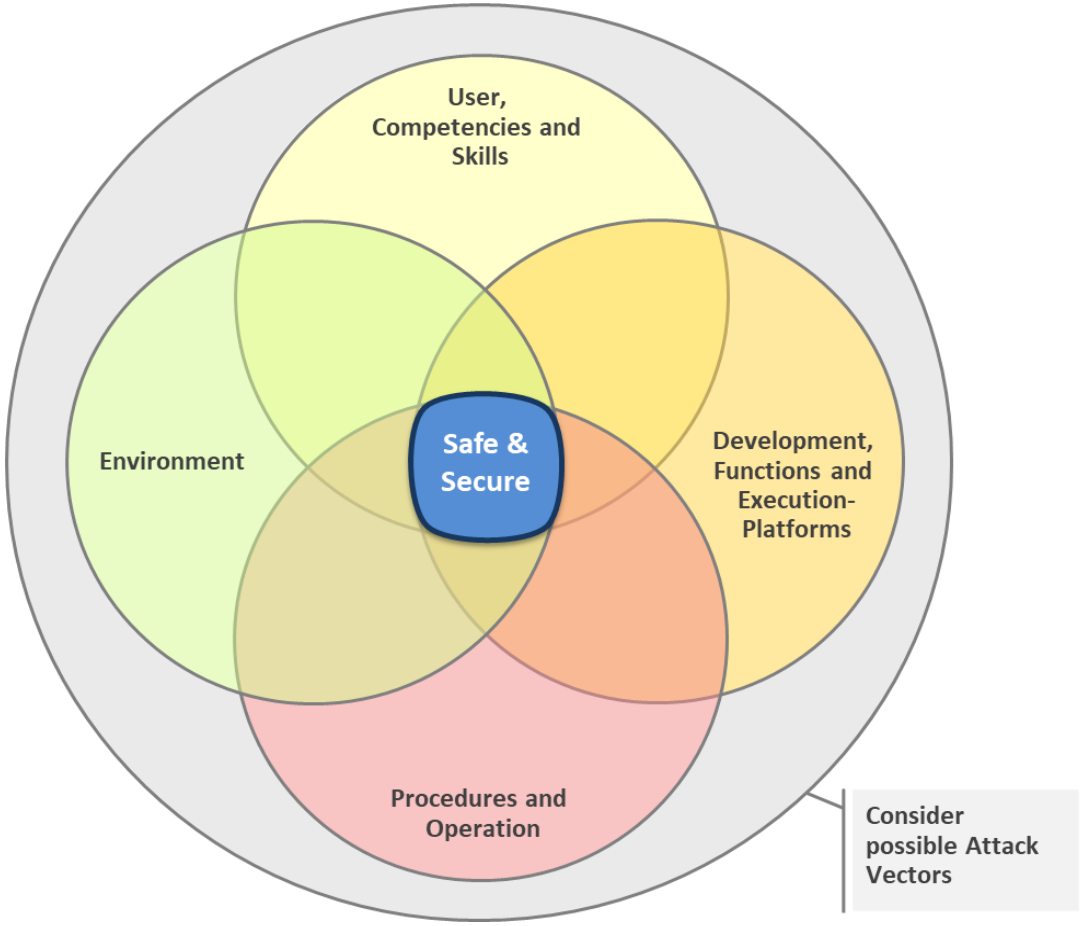
Level 5
Autonomous

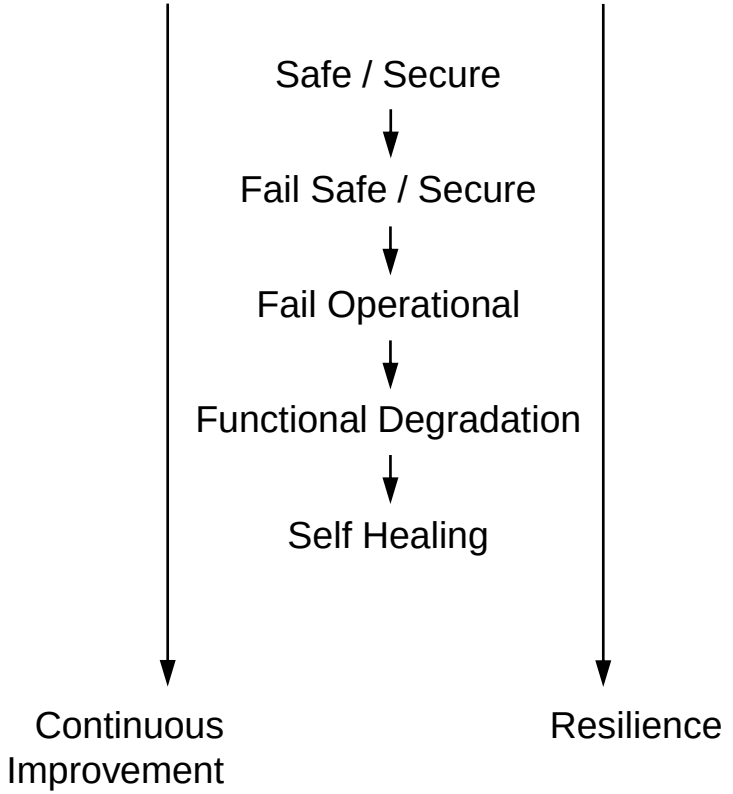


Different Deployment Strategies
and Use-Cases

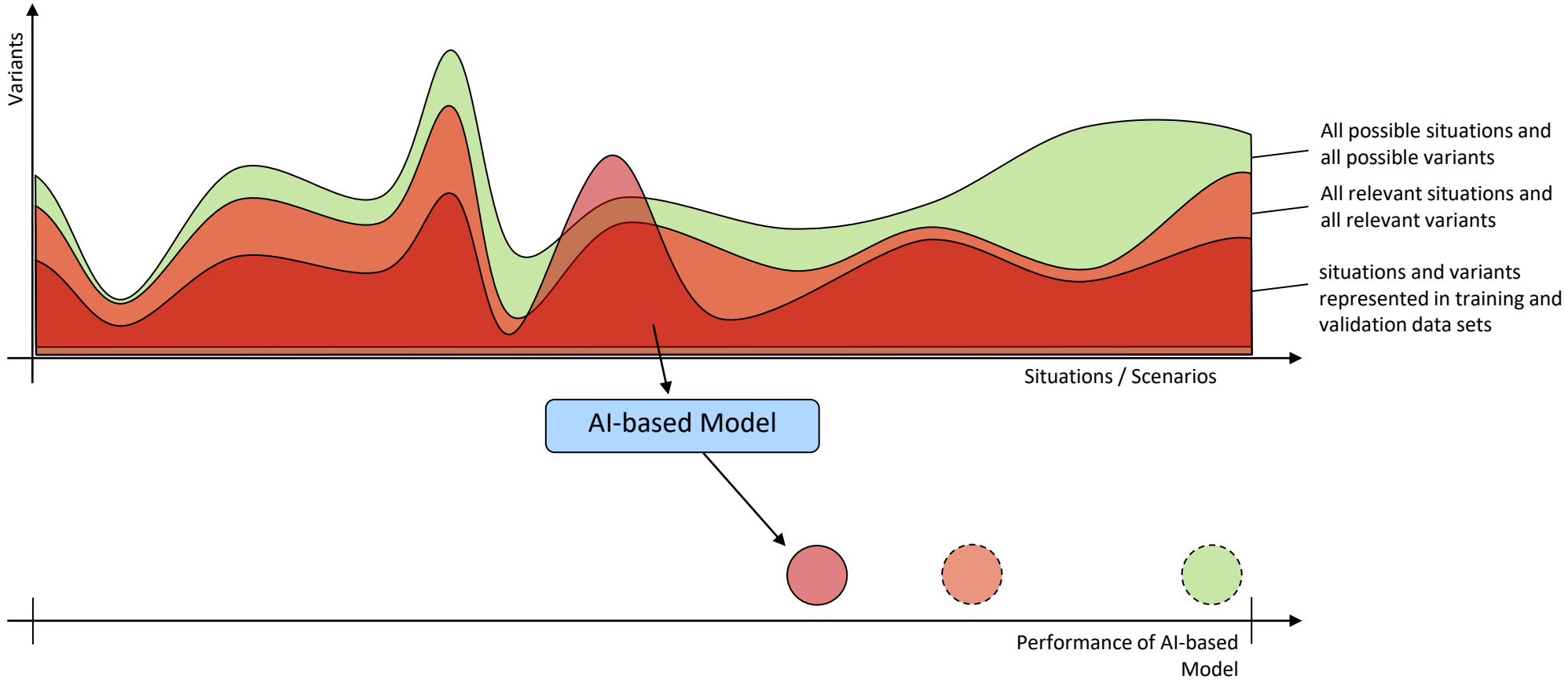
no cooperation	Level a
provision of function-specific data / information (i.e. handling of data/information by recipient remains open; use without explicit feedback to sender)	Level b
Level b + integration into receiver's situational picture and feedback to sender	Level c
Level c + coop. developm. of a situational picture with the aim of a common situational picture (possibly including joint interpretation or plausibility check)	Level d
Level d + cooperative planning with a fixed goal structure (possibly a differentiation regarding driving tasks could be useful (strategical / tactical / operational))	Level e
Level d + cooperative planning with a flexible goal structure (possibly a differentiation regarding driving tasks could be useful (strategical / tactical / operational))	Level e*

Automated and Connected Driving – Safety & Security by Design

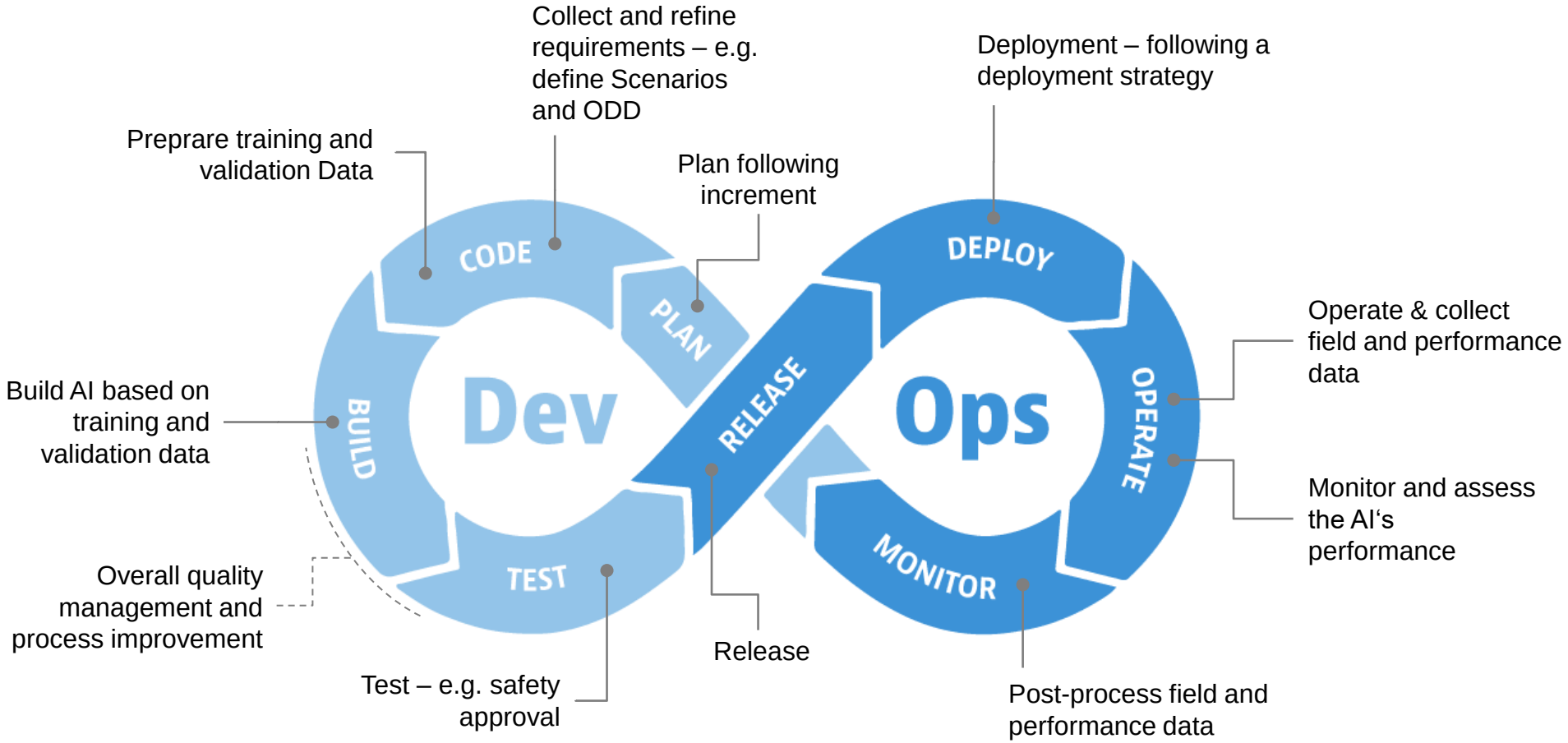




Product Development – Incremental and Iterative → DevOps



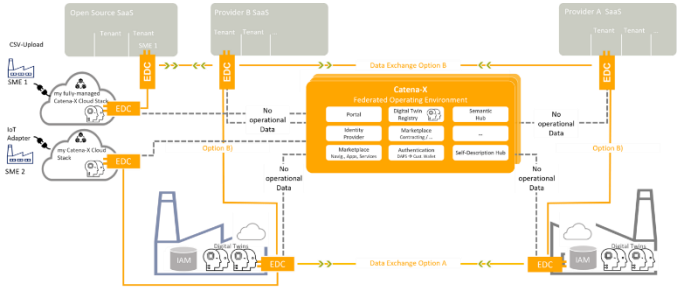
Product Development – Incremental and Iterative → DevOps

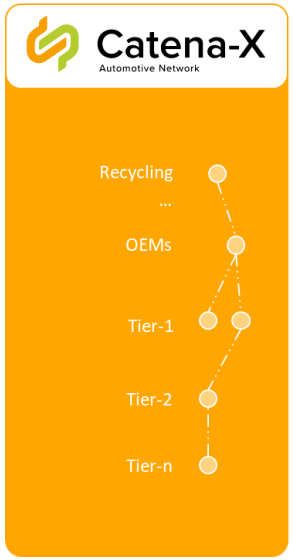


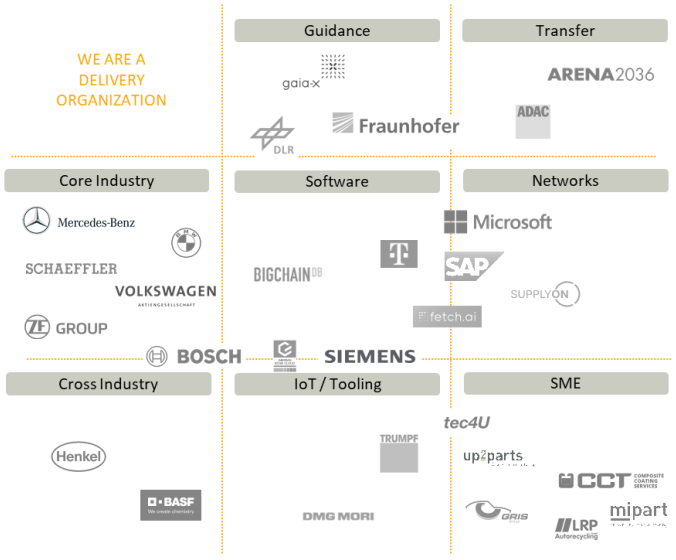
Funded Projects – Catena-X (1a/2)

Catena-X – Joint Mission

- We offer the most user-friendly environment for the **construction, operation and collaborative use of end-to-end data chains along the entire automotive value chain**.
- The resulting data ecosystem makes us **unique and is an important factor** for the sustainable development of the **industrial sector as well as the individual companies**.
- It **rewards all participants** with above-average resilience, innovative strength and profit opportunities.



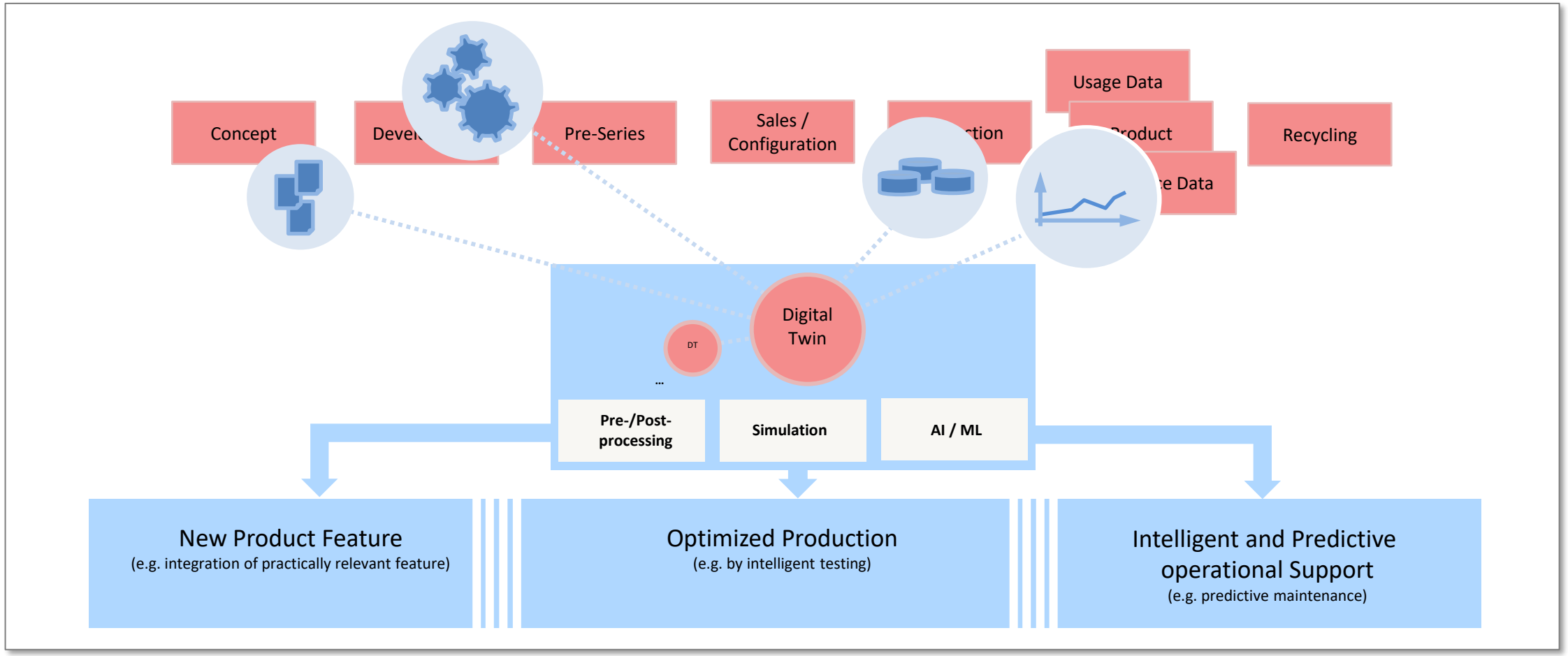




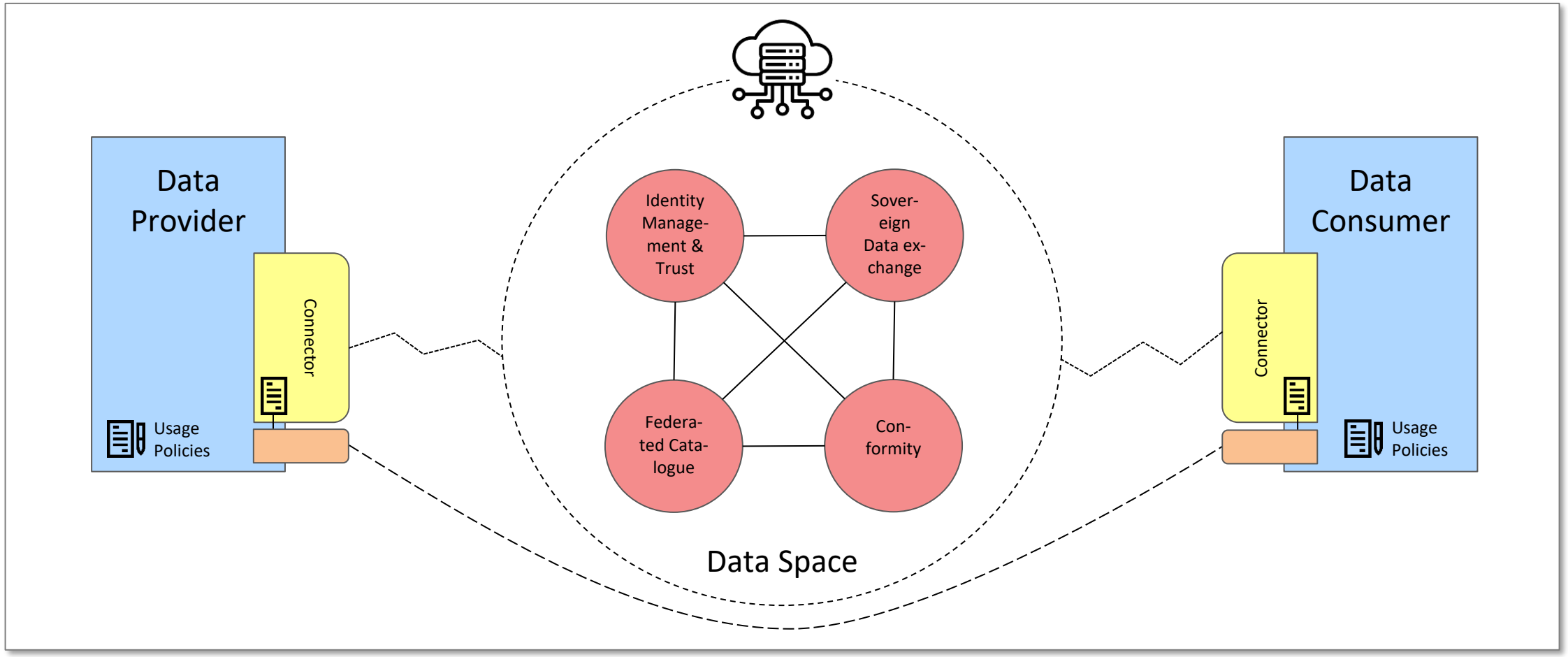
Catena-X Federated Operating System

- Traceability (of product-relevant Artefacts – e.g. Hard- & Software Components)
- Sustainability (e.g. CO2-Footprint and Compliance with Social Standards)
- Circular Economy (e.g. minimizing the CO2-Footprint)
- Quality Management (e.g. Real-Time & Cooperative/Collaborative Quality Mgmt.)
- Demand and Capacity Management
- Business Partner Database
- Data- and Model-Centric Engineering and Operational Support
- Modular Production
- Manufacturing as a Service
- Real-Time Control and Simulation

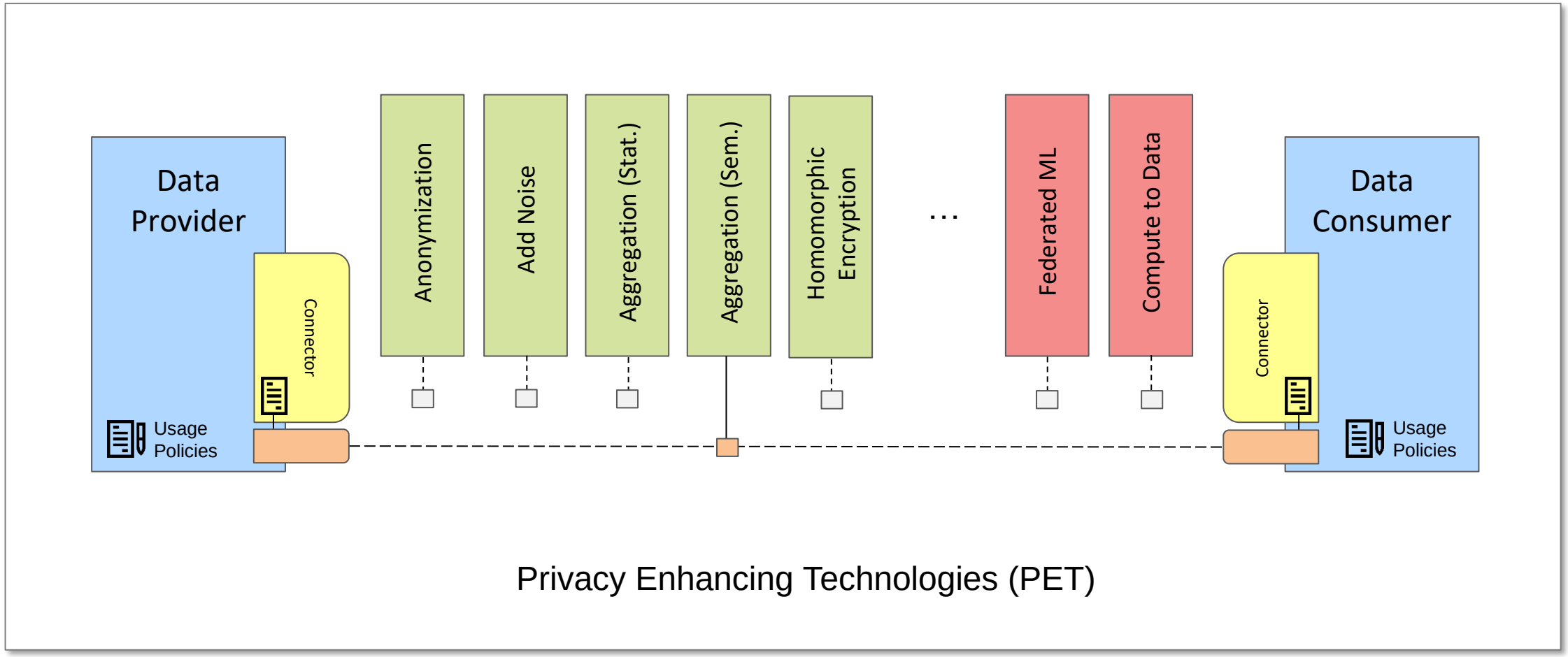
Funded Projects – Catena-X (1b/2)



Data Spaces & Privacy Enhancing Technologies (1/2)

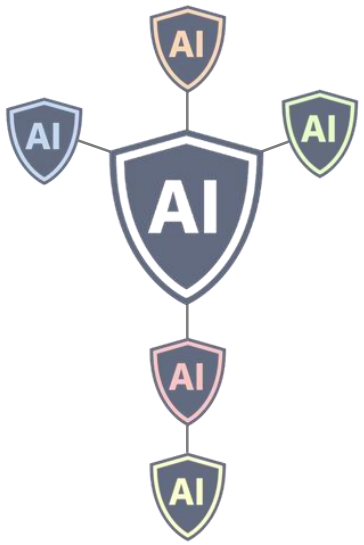


Data Spaces & Privacy Enhancing Technologies (2/2)



Summary & Take-Home Message

– Advances in the area of AI Safety and Security are urgently needed in order to overcome barriers to success for new technological developments in application areas relating to a trustworthy and safe use of AI.



- Industrialize AI for safety-critical systems and systems with high security requirements – proof of safe and secure AI
- Establish a solid foundation for AI-Engineering
- Improve methods and algorithms
- Consider safety and security as an integral aspect
- Provide execution-platforms with respect to innovative computing approaches
- Research on ethical, legal and social aspects in a inter-/transdisciplinary approach

– This is of high relevance for DLR’s established research areas and many other fields of applications beyond the scope of the DLR.

Contact

Prof. Dr. Frank Köster
German Aerospace Center
Lilienthalplatz 7
38108 Braunschweig
Germany

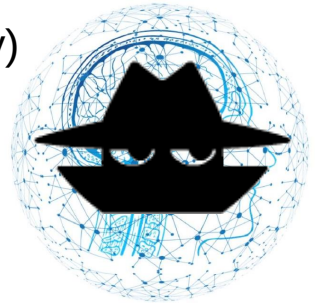
Frank.Koester@dlr.de

AI-based Functionalities and systems – Plausible Expectations

- AI must be aware that there are “bad guys” (everywhere) – possible attack vectors
- AI as a target – direct and indirect, e.g.
  - training and validation data as well as operational data
  - execution platforms (hard-/software) and surrounding system components
  - AI-technologies and -methods, due to abilities and limitations – e.g. by adversarial and denial-of-service attacks
  - malware injected during e.g. maintenance of AI-based components or integration of new knowledge blocks
- AI as a tool

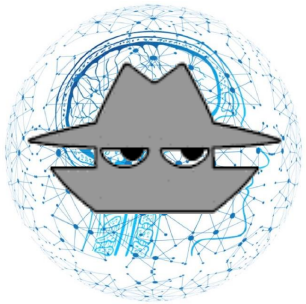
ATTACK

- part of (smart) malware and ransomware
- coordinate attacks and adapt attack-strategies (e.g. identify and use the right moment, raise efficiency)
- mimic behavior – e.g. to deceive anomaly detection and to drive adapting and self-improving phishing attacks (e.g. imitate writing styles or utilize context information from social networks)
- improve fake landing pages with connection to social networks and interests of target persons
- identify vulnerabilities – e.g. identify files with attack-relevant content



PROTECT

- part of (smart) intrusion detection – e.g. to establish anomaly detection on host or network level
- improve detection of phishing attacks und malware
- establish e.g. smart honey-pots or implement smart algorithms to estimate attacker’s identities
- raise efficiency and scale assessments of robustness or vulnerabilities
- basis for continuous authentication



AI-based Functionalities and systems – Plausible Expectations

The human/AI-interaction and human-in-the-loop topics will remain highly relevant

AI must be familiar with the concept of uncertainty and can also deal with uncertainty

AI must be aware that correlations are no causalities

AI must be aware that there are "bad guys" (everywhere)

Management of an AI during the entire AI-lifecycle is essential – including the management of training and validation data as well as operational data

Data provenance is a highly important topic

The importance of synthetic data will increase

Safety and security in the context of AI goes from aspiration to foundational requirement (cf. responsible AI)

Competencies, methods, tools and infrastructures are essential to enable institutions to use/implement AI

Reinforcement Learning will become an increasingly important AI paradigm

AI evaluation and goal-directed tuning will become essential

AI will help us to improve and build AI

AI ethical review boards and codes of conduct on AI and data ethics are needed

Open cloud/edge-based data and service-ecosystems will be the "home" of many AI-based components

Cooperation and Collaboration in the context of AI-development/maintenance will be important

Natural language processing and, in general, the analysis of un-structured audio, image and video data as well as complex structured data, like e.g. digital engineering drawings, will remain major topics

Deep-fakes will be a big problem

Creative AI will begin to generate (mainstream) content

Appropriate guardrails and governance mechanisms will be defined to ensure that the use of AI systems adheres to the principles of trustworthy as well as responsible AI (legislation)

Contact

Prof. Dr. Frank Köster
German Aerospace Center
Lilienthalplatz 7
38108 Braunschweig
Germany

Frank.Koester@dlr.de