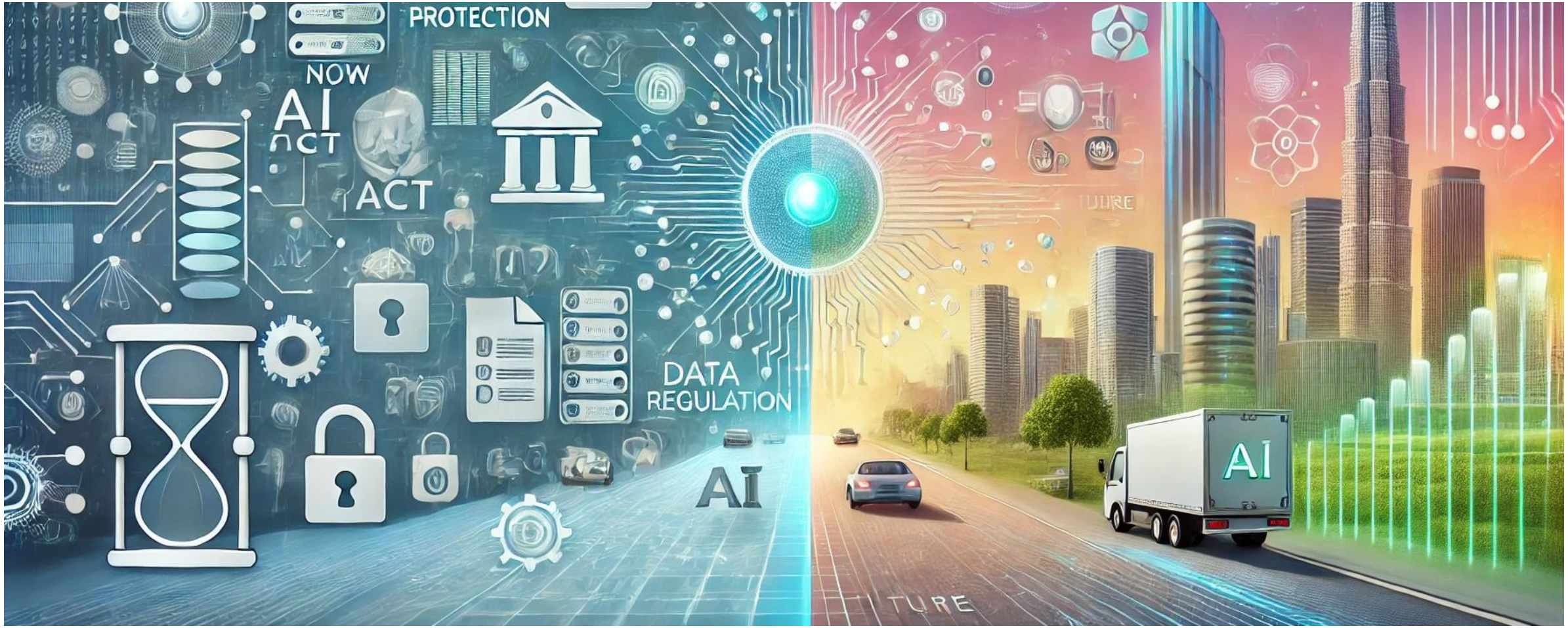
Key Risk Factors of AI Applications

ReMeP Annual Conference 2024

Peter Kieseberg

Major impact

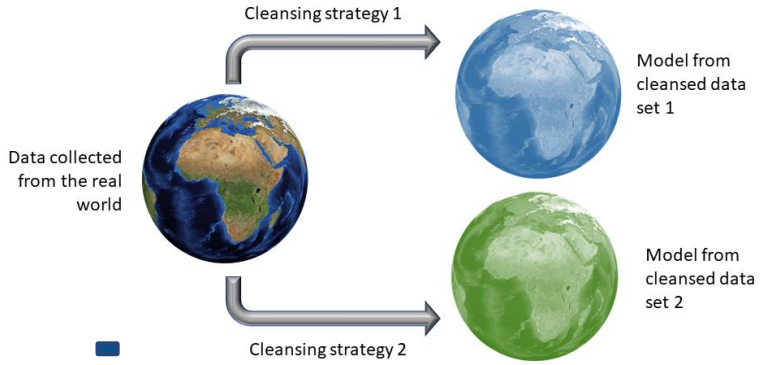
Now and the Future



New security challenges

New attacks & risks

Stability & validity



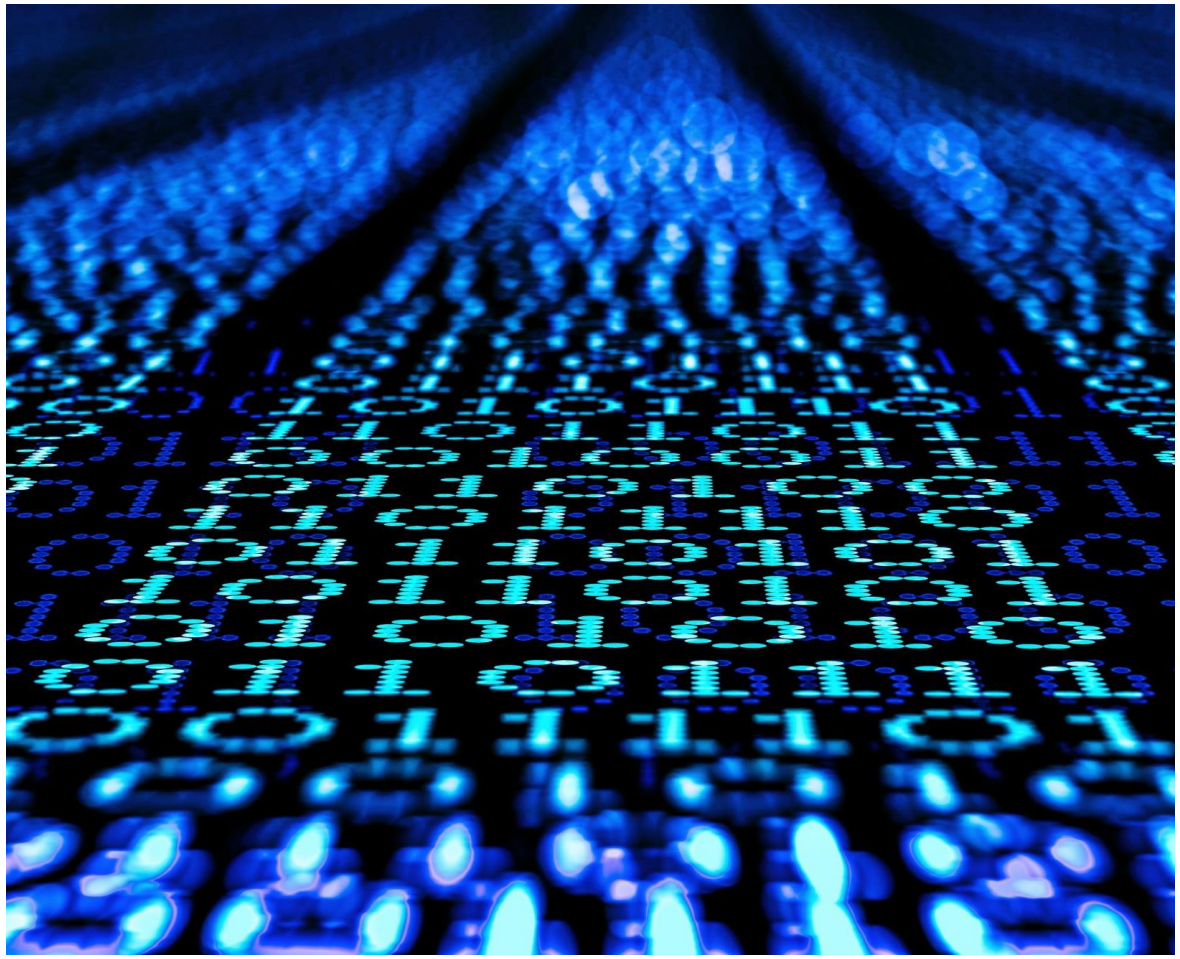
Bias & Data preparation

Privacy

Control over data / models

Procurement

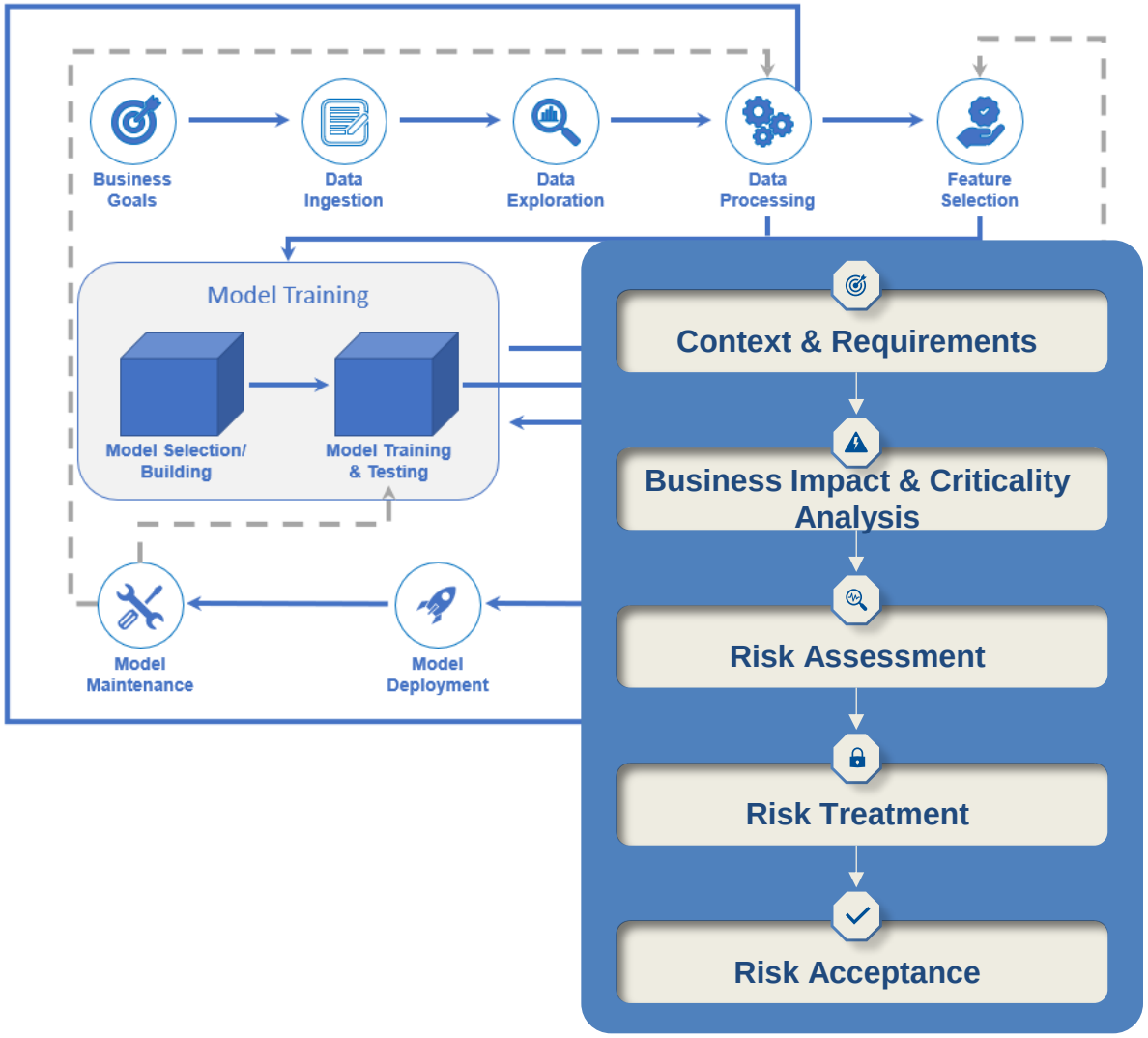
Data Defined Software & Explainability



Ubiquitous AI → What do we actually do?

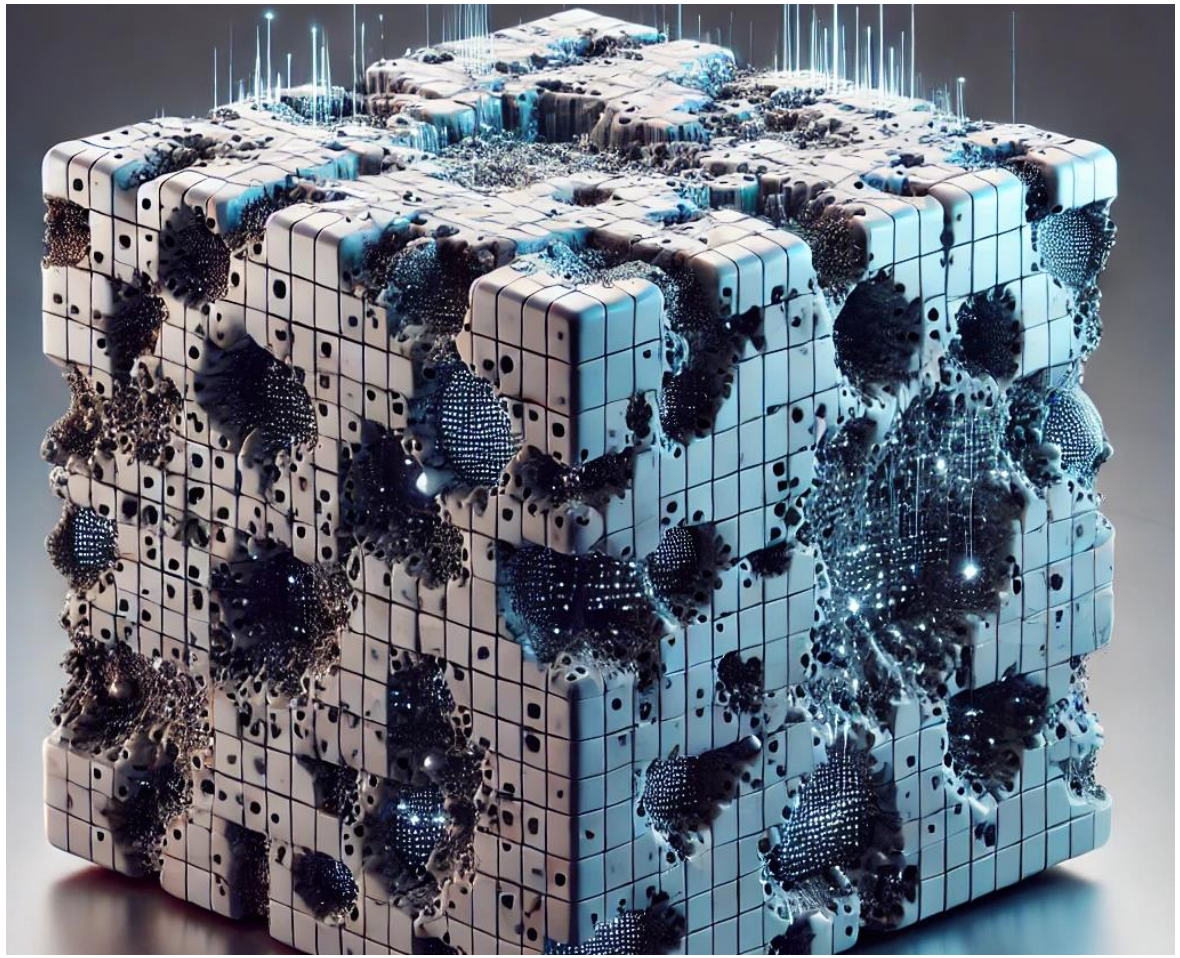
- What does the system do?
- Am I currently using AI?
- What risks arise?
- How do I deal with them?
- How do I make sensible decisions?
- Attribution problem

AI Risk Management



- Not ever AI is the same!
- Black box vs. understanding AI
- Risk & technology assessment
- Threat Intelligence
- Procurement vs. in-house development vs. customisation

Data & Data Preparation

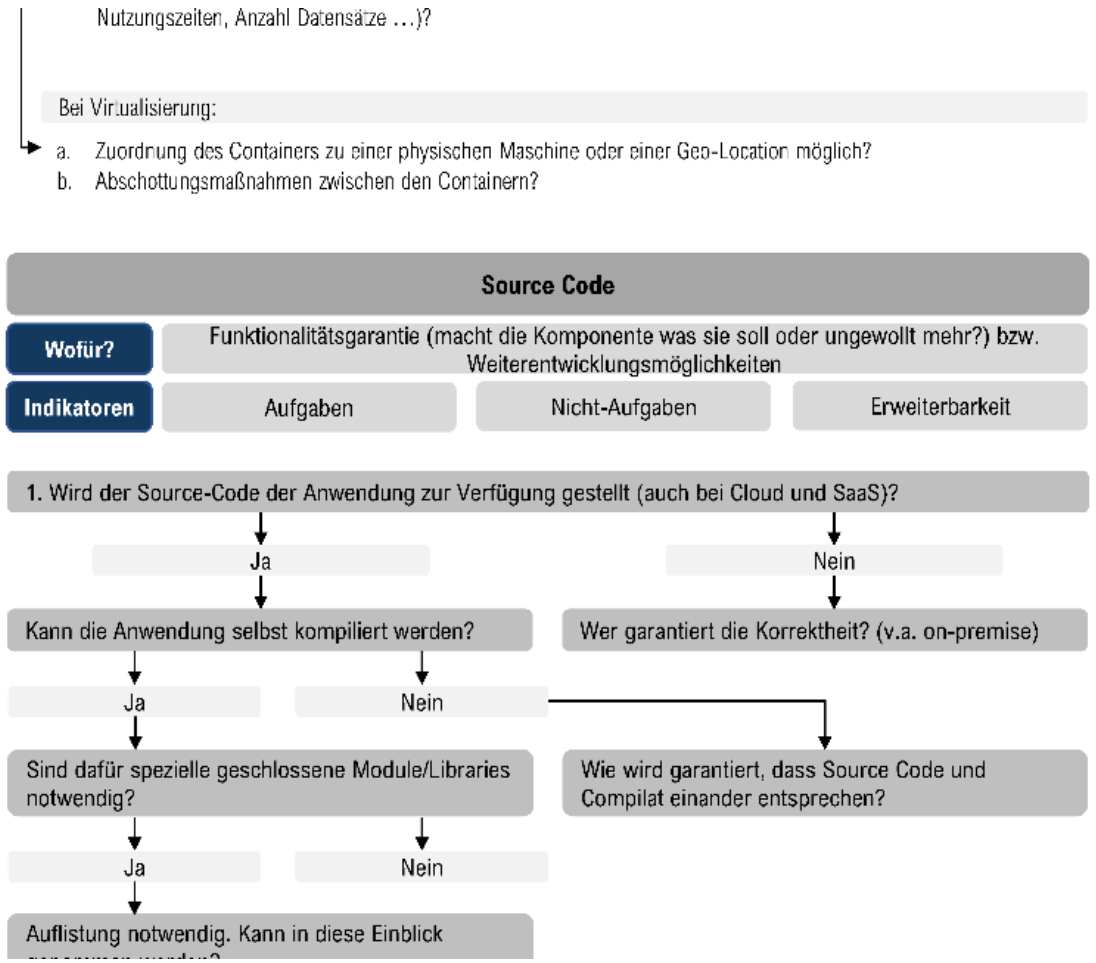


- (Non-) intentional bias, quality
- Data Cleansing
- Data pre-processing
- Reliable sources
- Protecting valuable data
- Dominant data pools / monopolies

Models & transparency

- Transparency
  - Who trains? What/what is trained on?
  - ChatGPT wrapper?
- Patching
  - Data and models
- Reinforcement learning
- Re- & Post-Processing
- Logging and monitoring

Procurement



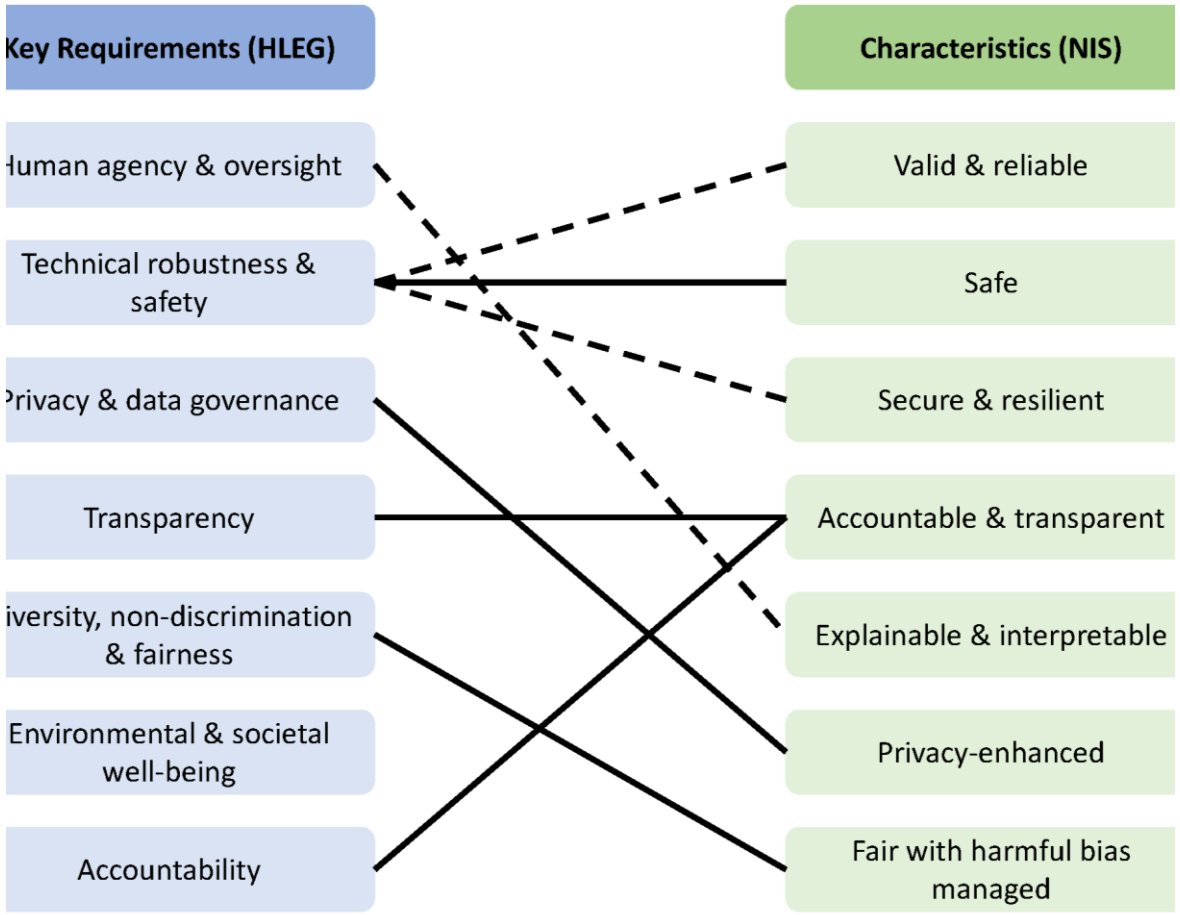
- How much in-house modification possible?
- A slight case of overselling …
  - LLMs everywhere
- What do we buy?
  - Full system, model, AI as a service …
- Who is responsible?

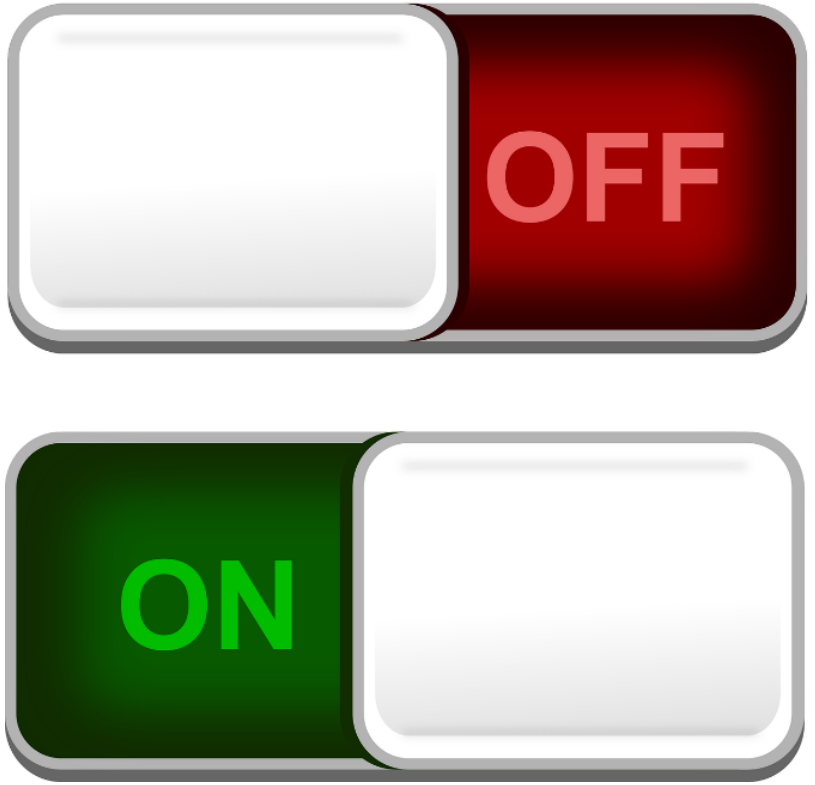
The future - Explorative scenario analysis

Chances

- CISO can benefit
- Time for risk management
- Documentation → YSRYFUM
- Thinking before deployment
- Traceability for data – Questioning data
- Creation of own AI components & AI security industry

Trustworthy vs. Controllable AI





Question of jurisdiction

Thank you very much!