DigiLabs26 RVR Regionalverband Ruhr

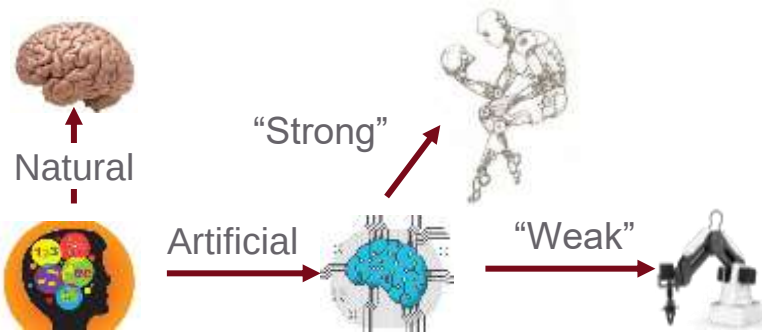
VON DER KI ZUR SUPERINTELLIGENZ

„KI Macht oder Ohnmacht“

Dr. Helmut LEOPOLD
Head of Center for Digital Safety & Security
AIT Austrian Institute of Technology

Essen, 16. Oktober 2026

(v0.6)



ARTIFICIAL INTELLIGENCE …

REKURSIVE SELBSTVERBESSERUNG FÜHRT ZU EINER „INTELLIGENZEXPLOSION“



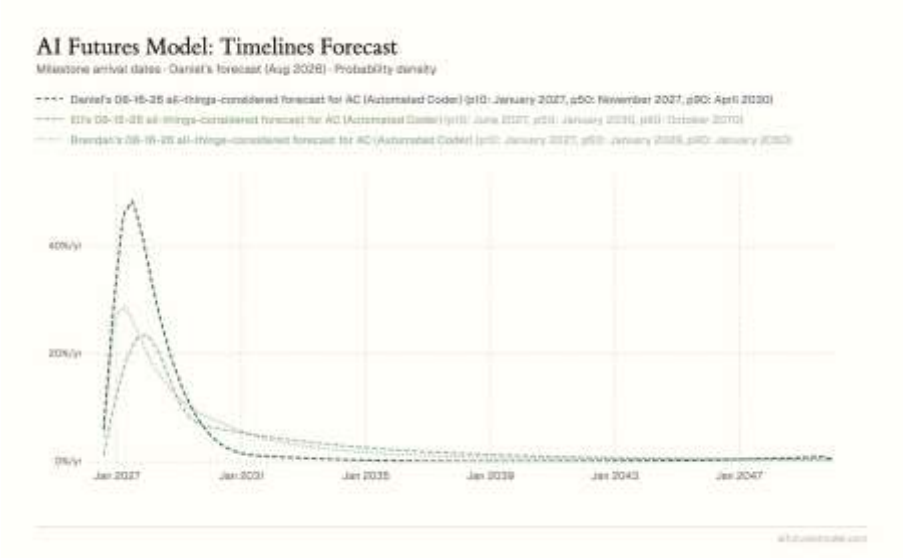
“Strong AI”

- **Definition einer ultraintelligenten Maschine**: “*A machine that can far surpass all the intellectual activities of any human, no matter how clever*”.
- **Die Intelligenzexplosion**: *Because designing better machines is an intellectual activity, an ultraintelligent machine could design even more advanced machines on its own, triggering a runaway "intelligence explosion" that leaves human intelligence far behind.*
- **Die letzte Erfindung**: Good folgert daraus: “*The first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control*".

Quelle: **Irving J. Good**, „Speculations Concerning the First Ultraintelligent Machine“, Advances in Computers Vol. 6, 1966

FORECAST „INTELLIGENZEXPLOSITION KI 2027“

- **2026 – Autonome Agenten**: von Chatbots zu „KI-Mitarbeitern“ (Software-Agenten), die eigenständig Workflows und Codierungsaufgaben auf Computern übernehmen.
- **2027 – Vollständige Automatisierung der KI-Forschung**: KI-Systeme erreichen das Niveau menschlicher Spitzen-Programmierer und -Forscher. Die KI übernimmt ihre eigene Weiterentwicklung.
- **Ende 2027** – Die Intelligenzexplosion: In einer geschlossenen KI-Entwicklungsschleife verbessert sich das System rekursiv selbst – millionenmal schneller als menschliche Entwickler.



Quellen: https://blog.aifutures.org/p/q25-2026-timelines-update-uplift

„AI 2027“-Bericht, ehemaligen OpenAI-Forscher Daniel Kokotajlo; April 2025

TECHNOLOGIEQUALITÄT UND FINANZIER- UND MACHBARKEIT?

KI Code Produktivität

R&D AIT
- Steigerung der Produktivität

Antropic Code:
- > 80% code AI generated

➔ Code Qualität?

KI Energieverbrauch

User-Energie:
- Eine ChatGPT-Anfrage: 0,3-3 Wattstunden (Wh);
- KI-Bilder und -Videos: x 1000

Rechenzentren global:
- 2024 ca. 500 TWh
- 1,5 bis 2 % des weltweiten Strombedarfs.
- 2030: ca 1 – 1,4 TWh!

➔ Energieproduktion?

KI Investitionen

US Tech Konzerne:
- 2026: 2.700 Rechenzentren (mit Hyperscaler), ca. 40 % global
- **2027**: ca. 900 Milliarden - 1 Billion US Dollar
- **2030**: kumulierte KI-Investitionen: ca. **7 Billionen US Dollar!**

➔ Geschäftsmodelle?

➔ Finanzblase?

KI HW Ressourcen

- Nur ein RZ von Elon Musk („xAI“): 220.000 Nvidia GPUs
- JUPITER (D): 24.000 Nvidia-KI-Chips

➔ Globale Chip Supply Chain?

Personal Ressourcen

Pro Rechenzentrum:
- Infrastrukturbau
- mehreren Hundert direkte Arbeitsplätze?
- Kernbetrieb: automatisiert
- externe Dienstleister für Sicherheit, Wartung, Reinigung

➔ Qualität der Arbeitsplätze?

Quellen: https://www.erneuerbare-energie.at/energiefakten/2025/17/01/energiebedarf-ki / https://www.oeko.de/news/pressemeldungen/ki-auf-kosten-des-klimaschutzes-energiebedarf-von-rechenzentren-verdoppelt-sich-bis-2030/ / https://www.gtai.de/de/trade/usa/branchen/elektrizitaet-1829814 / Internationale Energieagentur (IEA) / The Teaser Period: Why the AI Boom Is Hitting a Reset Wall, August 2026, https://www.groundbrkr.com/p/the-teaser-period-why-the-ai-boom, Bank of America, Wall-Street-Analysten

EUROPEAN AI FACTORIES

EU AI Factories





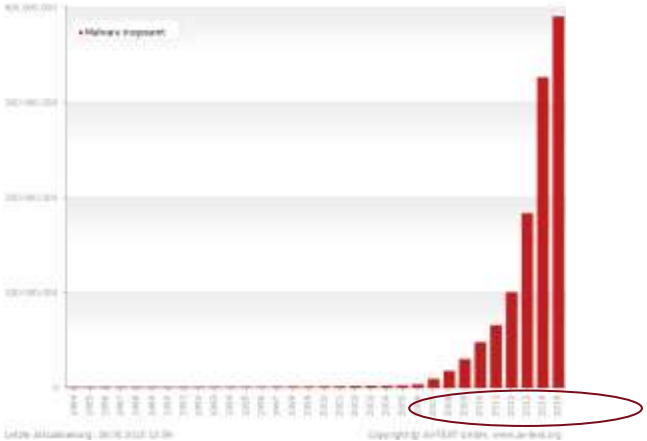
EU Gigafactories



Overall European investment 10-30 billion €!

KI IST EINE SEHR NEUE TECHNOLOGY MIT WENIG ERFAHRUNGEN

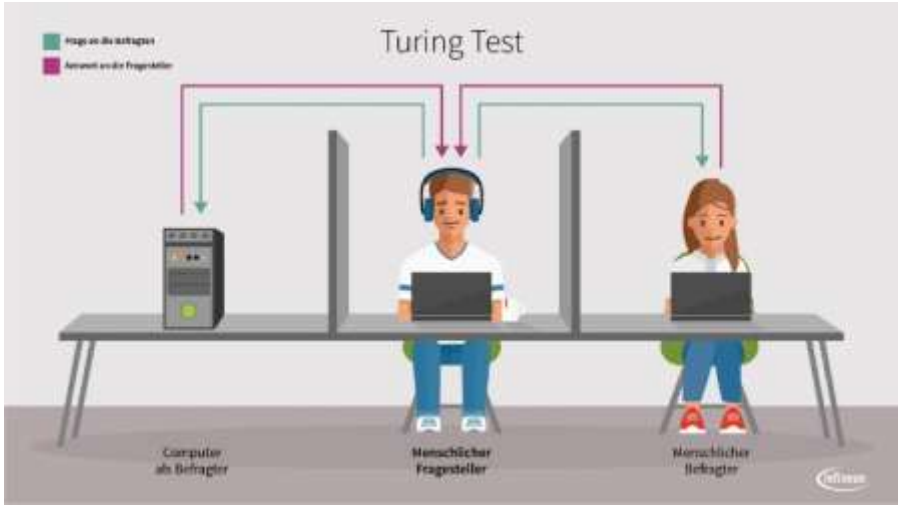
- SW Entwicklung und Testen: Erfahrung seit 30+ Jahre, 100k+ Publikationen
- KI als Technologie – sehr junge Technologie
- ➔ Wir brauchen R&D und Kooperationen!



IST KI INTELLIGENT?

WANN IST EIN SYSTEM INTELLIGENT?

Turing Test



 Bildquelle: Infineon https://www.infineon.com/cms/de/discoveries/definition-kuenstliche-intelligenz/

Intelligentes Leben ist mehr als nur Problemlösung
Intelligenz ist nicht nur eine schnelle „Datenverarbeitung“

Der KI fehlen …

- Werte, Vertrauen, Wahrheit, Freiheit, Freundschaft, Mitgefühl
- Community und soziales Zugehörigkeitsgefühl
- Zeitbewusstsein – Planung, Ziele
- Willen



Emotionen - Trauer, Spass, Liebe, Empathie, …

Intelligenzen: soziale, körperliche, räumliche, soziale, intrapersonale…

Wir brauchen …

- Kompetenzen als Entwickler, Anbieter und als Benutzer
- Standards, Regulierungen
- R&D für eine Technologie- und Datensouveränität

ARTIFICIAL INTELLIGENCE …



Natural



Artificial



"Strong"



"Weak"



Generative AI
Large Language Models (LLM)

Machine Learning (based on Data)

Supervised

Unsupervised

Logic/Rule-based Systems (Symbolic Methods)

Expert Systems / Recommenders / Optimizers
...



Tuning of GenAI

- Domain-specific datasets
- Prompt engineering
- Retrieval Augmented Generation (RAG)
- Fine-tuning
- Orchestration of multiple GenAI
- Agentic AI (agent systems)

GenAI + Symbolic Methods

- GenAI-guided Requirement/ Architecture Design Validation
  - Consistency checking
  - Verification of solutions
  - Ontology-based RAG
- GenAI-improved V&V
  - Spec mining to synthesize invariants
  - Generation of input scenarios

AI SAFETY & SECURITY: A GLOBAL R&D CHALLENGE



- AI - **foundational technology** for critical infrastructures
- **Failures** or **malicious use of AI** can have **severe societal consequences**
  - from service disruptions to threats to human safety and national resilience
- Current **AI systems difficult to interpret, verify, and predict**
- **Building trust in AI** requires **rigorous assurance**
- International collaboration essential to build common **standards**, **certification frameworks** & **trustworthy AI ecosystems**

QUANTEN TECHNOLOGIEN - PHOTONIK

QUANTUM KEY DISTRIBUTION (QKD) IN THE EU

EU Data sovereignty through QKD

AUSTRIAN INSTITUTE OF TECHNOLOGY

enabling market uptake of QKD technologies in the EU

EuroQCI

The EuroQCI initiative aims to build a secure quantum communication infrastructure that will span the whole EU, including its overseas territories.



- EU Industry Strategy - EU Scientific & Industrie Leadership in Quantum Technology
- EU Infrastructure Strategy
- EU Security Strategy

IRIS², Europe's new Infrastructure for Resilience, Interconnection & Security by Satellites



EuroQCI building in particular on the activities of the Horizon 2020 OPENQKD project (AIT Lead)

EuroQCI – Quantum Communication Infrastructure Europe, **2019-2029**

QUANTENSENSORIK

- Medizinische Diagnostik
- Navigation – GPS „Alternative“
- Radarsysteme
- Geophysik
- Kameresensoren „durch die Wand“

Quelle: H. Leopold, Quantum Tech Governance, Springer, Juni 2026, https://www.springerprofessional.de/defencetech-governance/52773768

QUANTUM COMPUTER

IBM Q System One

Google

0 Kelvin

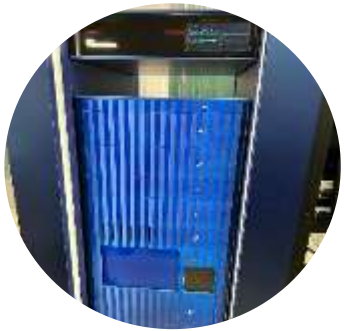


© h.leopold, 2024

XANADU Aurora (CAN)

PsiQuantum (Aus)

Photonics



© h.leopold, 2024

IONQ (USA)

AQT parityqc

Ion-Traps

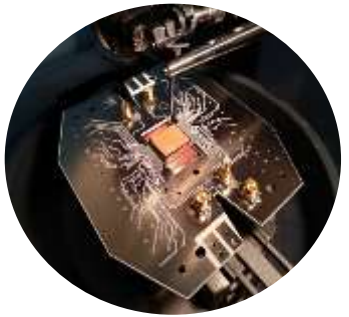


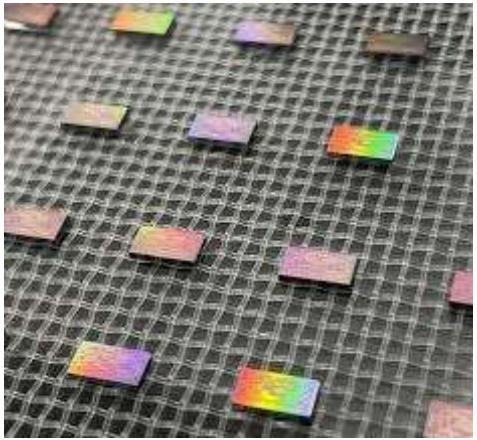
Bild: Universal Quantum
(https://qci.dlr.de/ionenfallen)

Austrian stakeholders have key technology competences

Leading Research & Technology in Quantum & Photonics Chip design

European Trapped Ion Quantum Pilot Line 'CHAMP-ION' coordinated by Silicon Austria Labs with Infineon, AQT, University Innsbruck, Parity QC

PHOTONIC COMPUTING – PHOTONIC TRANSISTOR



Quelle: AIT Center for Digital Safety & Security

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

LLM-SICHERHEITSRISIKEN



FIVE SECURITY PROBLEM AREAS OF GENERATIVE AI SYSTEMS (LLMS)

Ein LLM ist **keine** Datenbank!

- statistisches Modell einer Datenbank
- verlustbehaftete Komprimierung einer Datenbank

Sicherheitsrisiken von KI Systemen:

1. Halluzination → KI erfindet Fakten, Behauptungen, Referenzen, etc.
2. Beeinflussing des Ergebnises durch spezielle “Abfragen” (prompts) → “prompt injection attacks”
3. Zensur/Manipulation der KI ist nicht transparent & nicht nachvollziehbar
4. Kommerzielle APIs
5. Manipulation der KI für kriminelle Aktionen

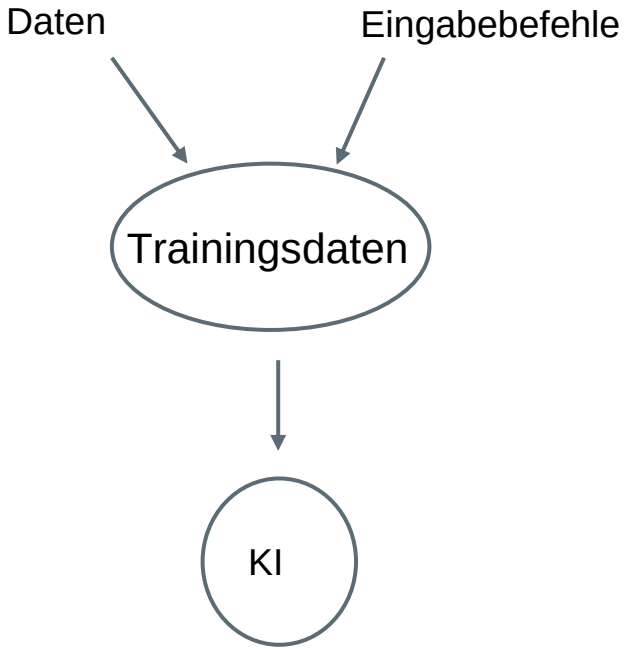
→ Grundlegende Probleme bei der Verwendung von KI Agenten!

LLMS - NICHT VERTRAUENSWÜRDIGE INHALTE

Prompt-Injection

- Das grundlegende Problem besteht darin, dass LLM-Anwendungen **nicht zwischen Daten und Anweisungen unterscheiden können.**

- **LLMs folgen den Anweisungen im Inhalt.**



LLMS SIND KEINE DATENBANK

Erfundene Inhalte - Beispiele:



U.S. NEWS

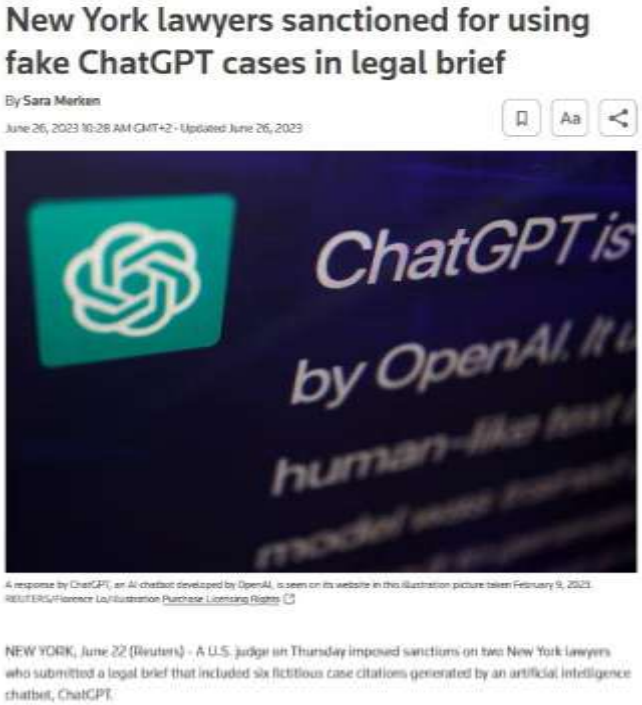
Judge sanctions lawyers defending Alabama's prison system for using fake ChatGPT cases in filings

Updated 12:11 AM GMT+2, July 25, 2025

Share

BIRMINGHAM, Ala. (AP) — A federal judge reprimanded lawyers with a high-priced firm defending Alabama's prison system for using ChatGPT to write court filings with "completely made up" case citations.

https://apnews.com/article/lawyers-judge-ai-prison-alabama-c6a64736cb488cf6379624403d3757ca



New York lawyers sanctioned for using fake ChatGPT cases in legal brief

By Sara Merken

June 26, 2023 10:28 AM GMT+2 - Updated June 26, 2023

Aa

A response by ChatGPT, an AI chatbot developed by OpenAI, is seen on its website in this illustration picture taken February 9, 2023. REUTERS/Florence Lo/Illustration Purchase Licensing Rights

NEW YORK, June 22 (Reuters) - A U.S. judge on Thursday imposed sanctions on two New York lawyers who submitted a legal brief that included six fictitious case citations generated by an artificial intelligence chatbot, ChatGPT.

https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/

5 KI HERAUSFORDERUNGEN DIE R&D AUFWAND VERLANGEN

Modellierbarkeit

Wie bekommen wir die “**Grundwahrheit**” **(ground truth)?**

Wie können wir einen Bias in der Grundwahrheit ausschließen?



Compliance und Ethik

Verhält sich ein KI System nach Regeln und ethisch?



Verifizierbarkeit

Wie testen wir ein funktionierendes Modell?

Wie robust ist ein KI-Modell gegen “**adversarial attacks**”?



Erklärbarkeit

Wie erklären wir das Ergebnis eines KI Modells?



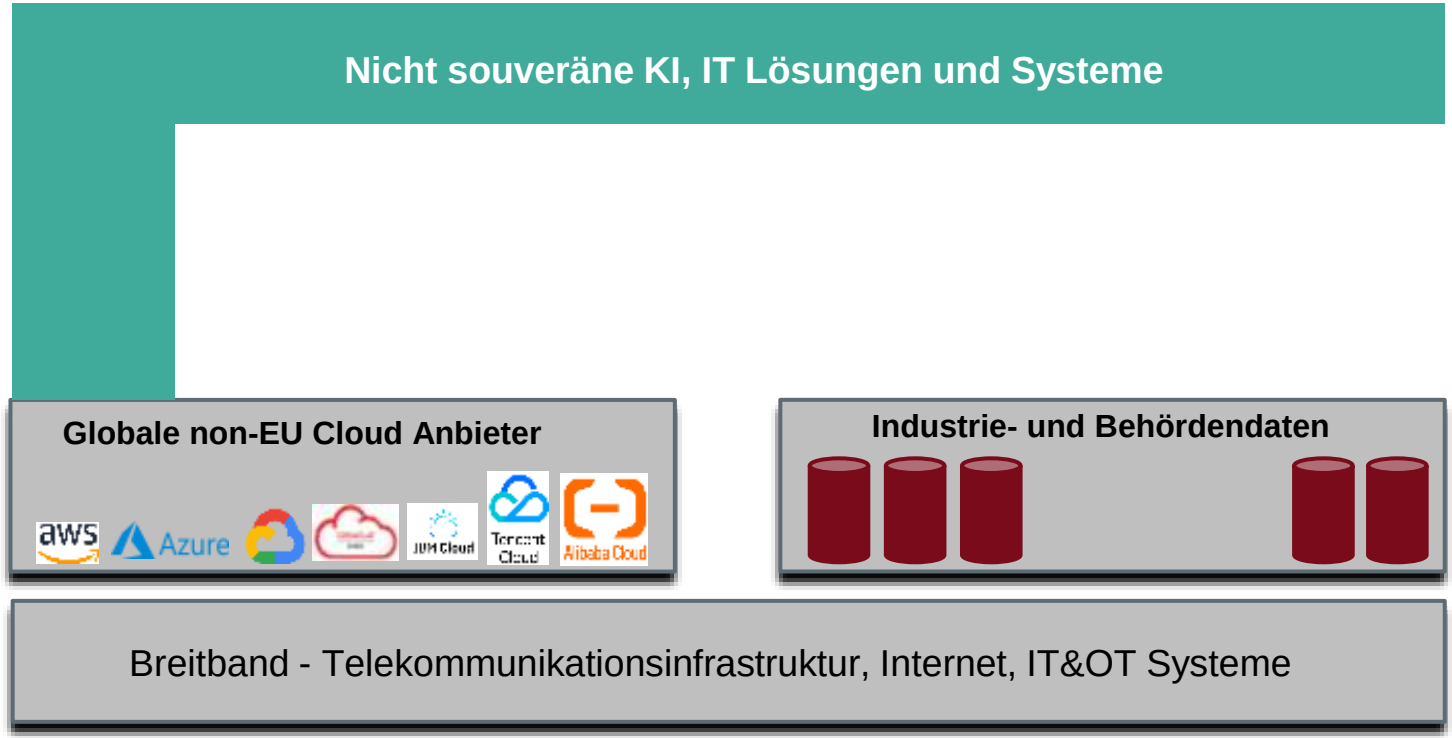
Haftung

Wer trägt die Verantwortung für KI-Systeme?



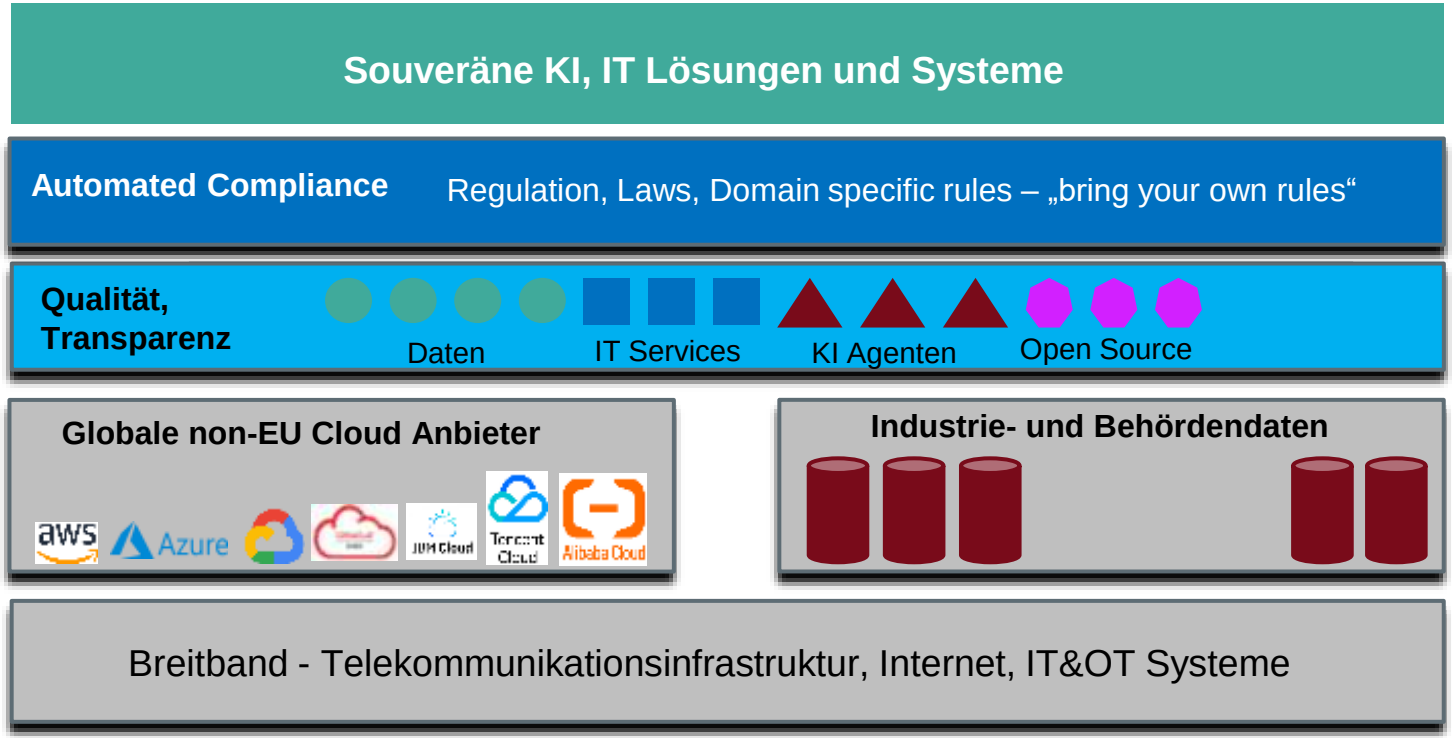
WIE BAUEN WIR VERTRAUENSVOLL UND SOUVERÄNE KI SYSTEME?

VOM INTERNET ZUM SOUVERÄNEN DATENGESCHÄFT UND SOUVERÄNEN IT SERVICES



Daten-speicherung Economy of scale

VOM INTERNET ZUM SOUVERÄNEN DATENGESCHÄFT UND SOUVERÄNEN IT SERVICES



Souvereignty Layer

Daten-speicherung Economy of scale

gaia-x

Gemeinsam Standards & Open Source entwickeln

Gaia-X Summit, 19.11. – 20.11.26, Vienna

Demonstration of Austria as an essential shaping MS in the EU on the topic of data sovereignty

- ✓ **Motto**: Season 2.0 of Sovereign, Trusted AI & Data Ecosystems
- ✓ Gaia-X Hub Austria as host
- ✓ **Europe's Central Conference on Data Spaces, Digital Sovereignty and Trusted Cloud & Data Infrastructures**.
- ✓ **Gaia-X flagship event** for industry, research and politics with best practices, tech sessions and networking on the topic of data sovereignty, data spaces and Gaia-X.
- ✓ **Presentation of Austrian. Innovation & Data Industry**
- ✓ **Programm**: https://gaia-x.eu/event/gaia-x-summit-2026/





AIT KI KOMPETENZ-PORTFOLIO

Counter Cyber Crime



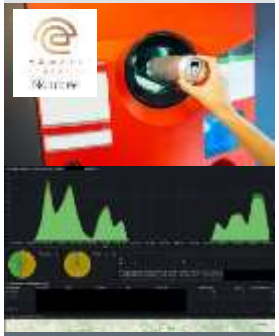
www.fakeshop.at

Security Operation Center (SOC)



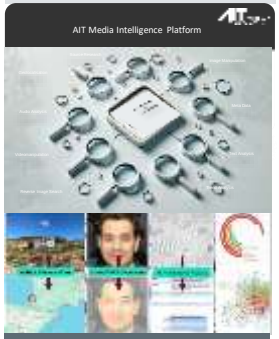
www.taranis.ai

Beverage Packaging Fraud Detection



www.recycling-pfand.at

Counter Desinfo & Terrorism



www.media-intelligence.ait.ac.at

Security by Design



www.media-intelligence.ait.ac.at

Security by Design



www.ai4trees-project.at

Source:

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

DANKE!

The AIT Team for a better global AIT technology

Helmut.leopold@ait.ac.at