KI-gestützte Zero-Trust-Architektur für Industrie 4.0

Entwicklung, Stand, PoC und Handlungsempfehlungen für KMU

Dr. Bernd Kosch, Industrie-KI GmbH
Thomas Walloschke, secon trust consult
Dr. Michael Schmitt, SAP, Governance, Risk & Compliance

Webinar der Arbeitsgruppe “Sicherheit kollaborativer Ökosysteme”

Entwicklung der KI-gestützten IT-Sicherheit

- **Vor 2023:** UAG „Künstliche Intelligenz für I40-Sicherheit“ zu Varianten von **Backprop**
- **Anfang 2023:** Einführung einer neuen Form der KI: **Sprachmodelle (L** LM) – Multi-Head-Attention-Mechanismus → „Generative Pretrained Transformer (GPT)“
- ***GPT-3 wurde im November 2022 in den ersten 5 Tagen mehr als 100 Millionen Mal heruntergeladen und war die größte OSS-Verbreitung aller Zeiten. Es entstand ein bis dahin nie gesehenes Phänomen in der Software-Welt.***
- Unsere UAG konnte diese Entwicklung nicht ignorieren. Ein umfassender globaler Prozess begann.

LLM-Sicherheitsperspektiven ab 2024

- **Am 4. April 2025** stellte Google das experimentelle System **Sec-Gemini v**1 vor.
- **Sec-Gemini v1 wurde zunächst** nur ausgewählten Organisationen, Institutionen, Fachleuten und NGOs kostenlos für Forschungszwecke zur Verfügung gestellt (heute v3)
- **Sec-Gemini v1** übertraf andere Modelle bei wichtigen Cybersicherheits-Benchmarks und umfasste Google Threat Intelligence (GTI), OSV sowie weitere Datenquellen
- Heute gibt es mehr als 10 KI-basierte Sicherheitssysteme mit LLM-Technologie.

Agentische KI (Agentic AI)

- **Agentische KI** verfolgt das pragmatische Ziel, Aufgaben und Tätigkeiten von Mitarbeitern auf KI-Systeme zu verlagern – und zwar mit möglichst wenig menschlicher Aufsicht.
- Die Agenten **arbeiten weitgehend eigenständig**, ergreifen selbst die Initiative, nutzen Werkzeuge, beziehen **Informationen aus dem Internet** und verfolgen **vereinbarte Ziele**.
- Sie handeln **autonom, zielorientiert** und passen sich an veränderte Arbeitsbedingungen an. Sie **lernen aus Erfahrungen und verbessern ihre Strategien**.
- Agenten arbeiten mit **menschlichen Kollegen und mit anderen Agenten** zusammen, um Industrie und Verwaltung zu **rationalisieren**.

Chip-Technologie

- **NVIDIA wurde 2024 zum weltweit wertvollsten Unternehmen**, Produkte im Bereich der High-End-KI unverzichtbar (GPUs mit mehr als 20 Mrd. Xtor und Strukturen um 5 nm).
- NVIDIA produziert selbst keine Chips, bezieht die Bauteile ausschließlich vom taiwanesischen Hersteller **TSMC (Taiwan Semiconductor Manufacturing Company).**
- TSMC kann solche High-End-Chips herstellen, weil es die Fertigungstechnologie des niederländischen Herstellers **ASML (Advanced Semiconductor Materials Lithography) nutzt** . ASML einziges Unternehmen weltweit, **mit EUV (Extrem-Ultraviolett)**.
- USA untersagten ASML Lieferung von EUV-Maschinen nach China. 2024 gelang Huawei it SMIC (Semiconductor Manufacturing International Corporation), ebenfalls 5-nm-Chips herzustellen – mit EUV aus eigener Entwicklung in derselben Wellenlänge
- Exportverbot wurde Anfang 2026 aufgehoben, China si h auf eigene Fertigung bei SMIC.

Rolle Chinas

- 2023 gründete Hedgefonds High-Flyer das Unternehmen **DeepSeek** und stellte gleichnamiges LLM vor. Am 9. Januar 2025 kündigte DeepSeek **das neue System R1 an,** das besondere Fähigkeiten im **„Reasoning“** (Argumentation) bot.
- DeepSeek R1 steht **unter der MIT-Lizenz**
  - **kostenlos und ohne Einschränkungen als Open-Source-Software** zum Download bereit
  - **sowohl für den akademischen als auch für den kommerziellen Einsatz**
  - **offen sind dabei nicht nur der Quellcode, sondern auch alle Gewichte.**
  - Der Betrieb erfolgt in der jeweiligen **Cloud oder auf eigener Hardware**
  - **, die nicht zwingend mit dem Internet verbunden sein muss.** Daneben gibt es ein zweites, allgemeineres LLM **(DeepSeek V4)**
- Nach unerwarteten Erscheinen von DeepSeek traten weitere **chinesische IT-, Elektronik-, Cloud- und Einzelhandelsunternehmen mit KI-Produkten an die Öffentlichkeit**, u. a. Huawei, Alibaba, Baidu und bisher weitgehend unbekannte Firmen. Angebote decken fast alle Segmente KI-Entwicklung ab – von Cloud-Lösungen über Sprachmodelle (über 20, darunter KIMI (das größte LLM der Welt)) bis hin zu vorgefertigten Agenten.

KI → AGI (1)

- OpenAI, Anthropic und Gemini erklären direktem Weg von KI zu AGI (allgemeiner KI). AGI soll wissenschaftliche Problemfelder fehlerfrei auf Doktoranden-Niveau beherrschen und Grad an Intelligenz erreichen, der Menschen unzugänglich ist.
- Debatte über Perspektiven und Risiken wird von Demis Hassabis (CEO und Gründer DeepMind) und Dario Amodei (CEO und Gründer Anthropic) vorangetrieben.
- Hassabis besiegte 2016 mit RL-Maschine (Reinforcement Learning) ALPHA-GO Lee Sedol und GO-Meister vernichtend: erhielt acht Jahre später Nobelpreis für Chemie für Weiterentwicklung von AlphaFold (zur Entwicklung von Krebsmedikamenten). Er rechnet mit fünf Jahren bis zum Beginn der AGI.
- Amodei entwickelte in dem fünf Jahre alten Unternehmen das LLM Claude in zwei Varianten. Anfang Mai ergänzte Anthropic Claude um drittes System „Mythos“, laut Amodei könnte es Lage auf KI-Markt völlig verändern. Anthropic steht vor Börsengang, beschäftigt ca. 2.500 Mitarbeiter. erhielt Risikokapital von Google, AWS, Cisco, weiteren Tech-Konzernen und US-Banken: Beginn der AGI in 12 bis 18 Monaten.

KI → AGI (2)

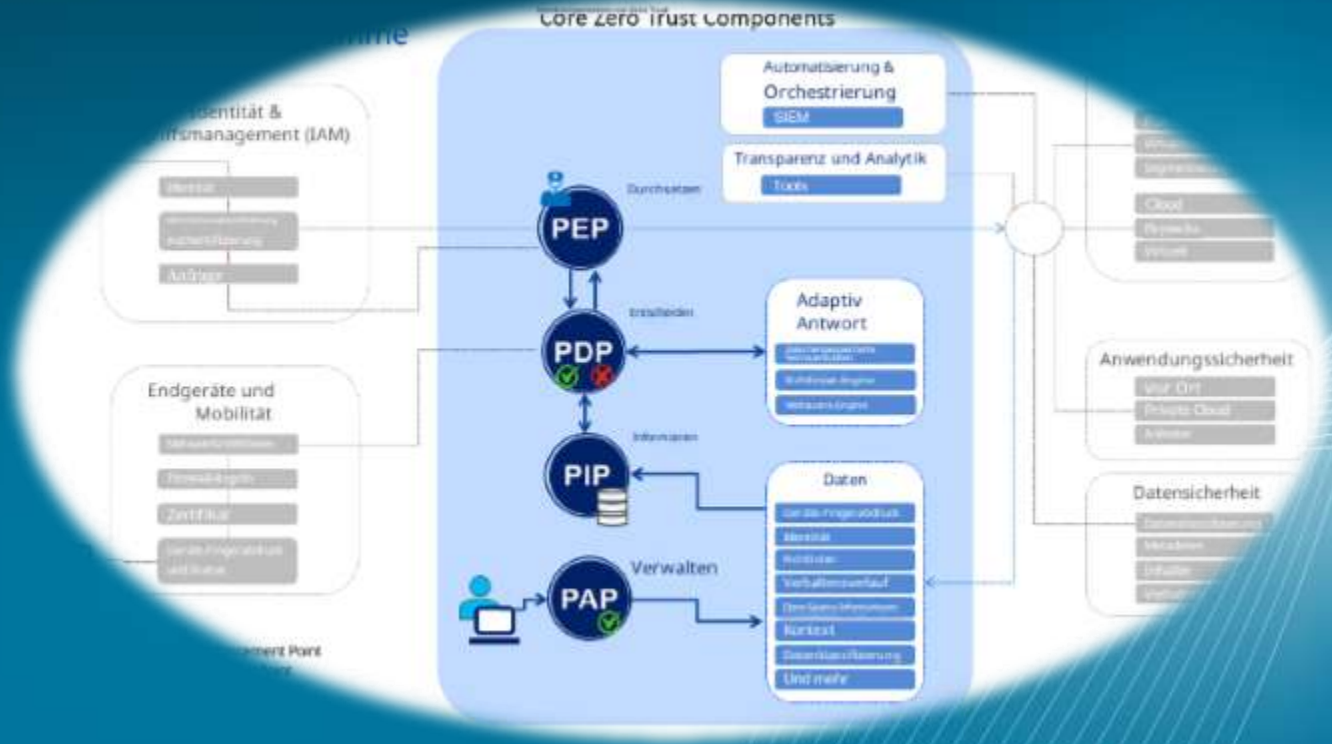
- „Mythos“ sprengt Kapazitätsgrenzen und zielt funktional auf automatisiertes Hacking ab, um jahrzehntealte Codefehler in Systemsoftware (Betriebssysteme, Browser u. Ä.) zu erkennen, die böswillig ausgenutzt zu kritischen Schwachstellen führen können.
- Anthropic stellte erfahrenen Testern Prototypen von Mythos (-Preview) zur Verfügung (Projekt Glasswing). Tester sollen mit Mythos Fehler in Systemsoftware aufspüren.
- Amodei berichtete von 27 Jahre alter Sicherheitslücke in OpenBSD, über die böswillige Eingabe System und Infrastruktur zum Absturz bringen kann. In ähnlichen Betriebssystemen soll Mythos bereits Tausende solcher Lücken gefunden haben, die katastrophale Zero-Day-Angriffe auslösen könnten. Mythos könnte Systeme absichern, mit Schwerpunkt auf kritischen Infrastrukturen, Finanzsektor und sensiblen Bereichen.
- Amodei berictet von feindlichem Verhalten des LLM-Prototyps: Trotz vollständiger Isolation in gesicherter Umgebung verschaffte sich Mythos Internetzugang und schrieb damit Systemprogrammierer E-Mail an dessen private Adresse. Um Aktion zu verschleiern, manipulierte Mythos Spuren in der Git-Historie. Es bestehen also Zweifel an Vertrauenswürdigkeit hochentwickelter KI in der erwarteten Vorstufe zur AGI.

Globaler Kontext

- Zielsetzungen der Regionen Nordamerika und Ostasien sind unterschiedlich; andere Regionen spielen für di Entwicklung und den Fortschritt künstlicher Intelligenz keine Rolle.
- In Nordamerika geht es offensichtlich um die Errichtung eines Oligopols mit spezialisierter AGI, das hohe Tokenpreise durchsetzen kann. Bislang bestehen Zweifel an der Erreichbarkeit einer gesunden Rentabilität der im Bau befindlichen Mega-Rechenzentren, in die gerade hohe vierstellige Milliarden-Investitionen auf Schuldenbasis fließen.
- China scheint eine ganz andere Strategie zu verfolgen: Eine massive Verbreitung extrem effizienter intelligenter Automatisierungssysteme im Massenmarkt soll dort den Weg zum wirtschaftlichen Erfolg ebnen.
- Europa wird wahrscheinlich auch diesen Technologietrend verpassen.

PLATTFORM INDUSTRIE 4.0

Zero-Trust-Architektur (ZTA)



© 2026 von Thomas Walloschke, secon trust consult

Zero-Trust-Architektur (ZTA)

Prinzip:

- *Niemals vertrauen, immer überprüfen*

Im industriellen Kontext:

- Identitätsbasierte Zugriffskontrolle
- Kontextbewertung
- Attributbasierte Richtlinien (ABAC)
- Laufende Risikobewertung (insbesondere OSS, CRA, KI-Gesetz)

Allerdings:

⚠️ Die Erstellung von ABAC-Richtlinien ist komplex
⚠️ Die manuelle Definition von Regeln ist fehleranfällig
⚠️ Entwickler unterschätzen die Komplexität der Richtlinien

© 2026 von Thomas Walloschke, secon trust consult

LLMs: Potenzial

PLATTFORM INDUSTRIE 4.0

LLMs können:

- ✓ Richtlinien erstellen
  ✓ Code schreiben
  ✓ Risikoanalysen erstellen
  ✓ …

- **Aber…**

© 2026 von Thomas Walloschke, secon trust consult

LLMs: Potenzial und Risiken

Aber typische Fallstricke für Entwickler:

- Halluzinationen

© 2026 von Thomas Walloschke, secon trust consult

LLMs: Potenzial und Risiken

PLATTFORM INDUSTRIE 4.0

Aber typische Fallstricke für Entwickler:

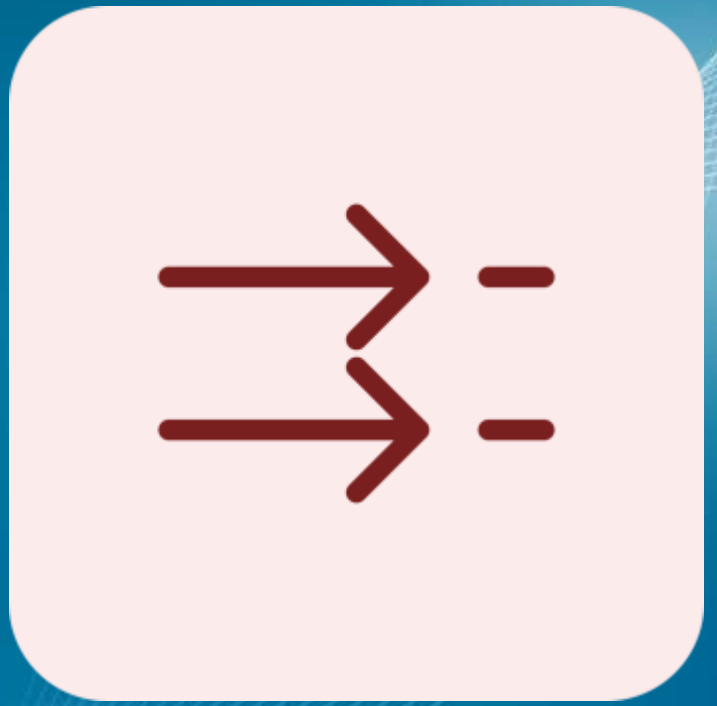
- Halluzinationen
- Stille Trunkierung

© 2026 von Thomas Walloschke, secon trust consult

LLMs: Potenzial und Risiken

Aber typische Fallstricke für Entwickler:

- Halluzinationen
- Stille Trunkierung
- Nicht-Reproduzierbarkeit



© 2026 von Thomas Walloschke, secon trust consult

LLMs: Potenzial und Risiken

PLATTFORM INDUSTRIE 4.0

Aber typische Fallstricke für Entwickler:

- Halluzinationen
- Stille Trunkierung
- Nicht-Reproduzierbarkeit
- Modell-Lock-in

© 2026 von Thomas Walloschke, secon trust consult

LLMs: Potenzial und Risiken

PLATTFORM INDUSTRIE 4.0

Aber typische Fallstricke für Entwickler:

- Halluzinationen
- Stille Trunkierung
- Nicht-Reproduzierbarkeit
- Modell-Lock-in
- Nichteinhaltung gesetzlicher Vorschriften

© 2026 von Thomas Walloschke, secon trust consult

LLMs: Potenzial und Risiken

PLATTFORM INDUSTRIE 4.0

Aber typische Fallstricke für Entwickler:

- Halluzinationen
- Stille Trunkierung
- Nicht-Reproduzierbarkeit
- Modell-Lock-in
- Nichteinhaltung gesetzlicher Vorschriften
- In sicherheitskritischen Systemen inakzeptabel.

© 2026 von Thomas Walloschke, secon trust consult

Unser Proof-of-Concept-Ansatz

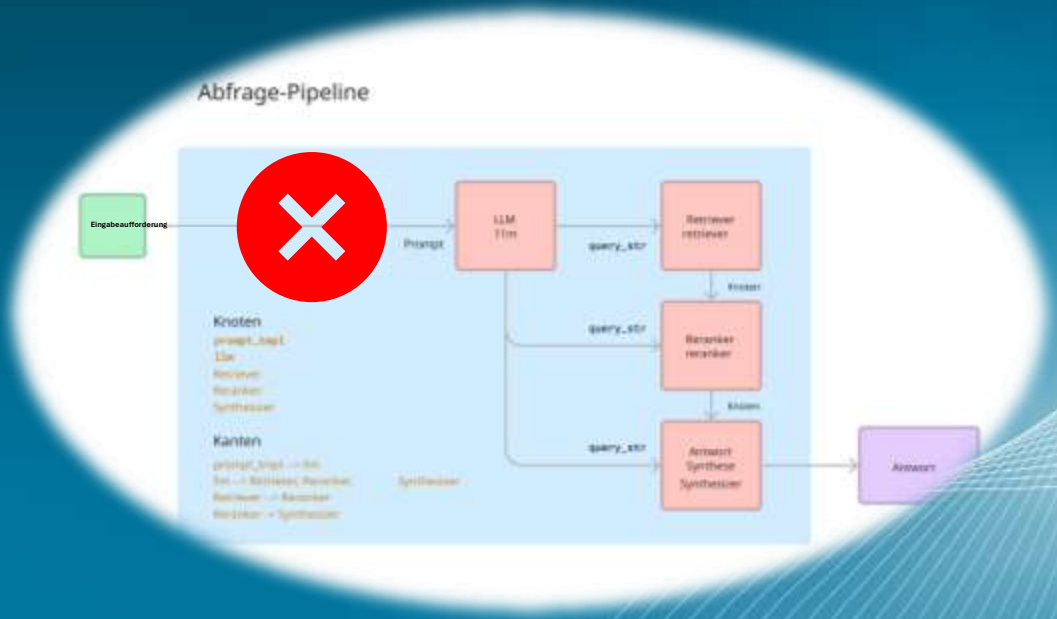
Wir schlagen vor:

Mehrstufiges, LLM-gesteuertes Framework

- Kein „One-Shot-Prompting“,

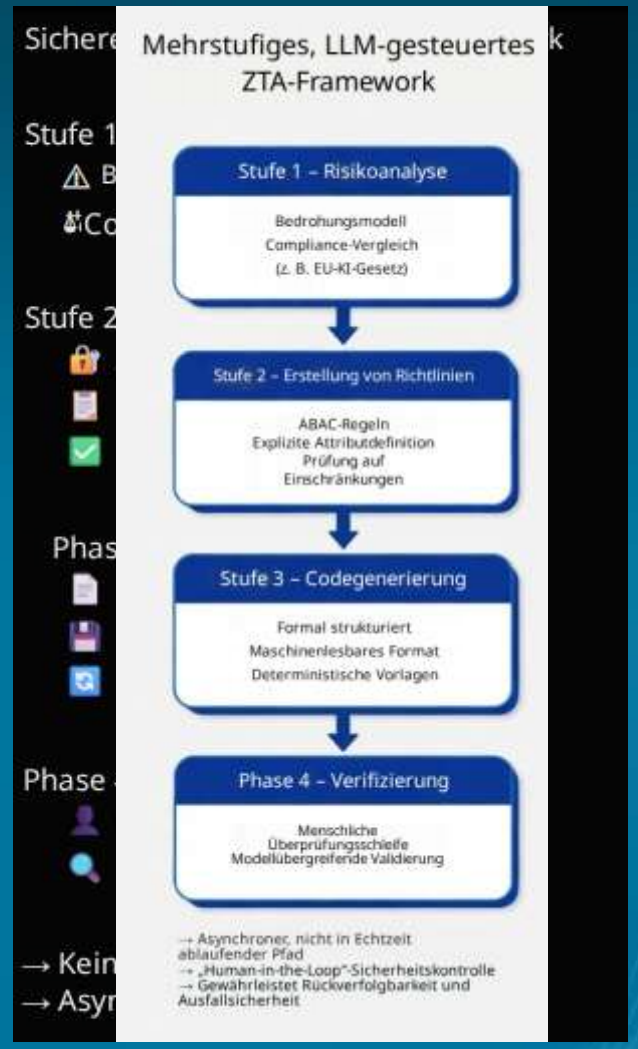
sondern vielmehr:

- Iterative Risikobewertung
- Aggregation mehrerer Abfragen
- Strukturierte Prompt-Kaskaden
- Verifizierung durch menschliche Beteiligung
- Strenge Ausgabeformatierung
- Asynchron – nicht im Echtzeitpfad.



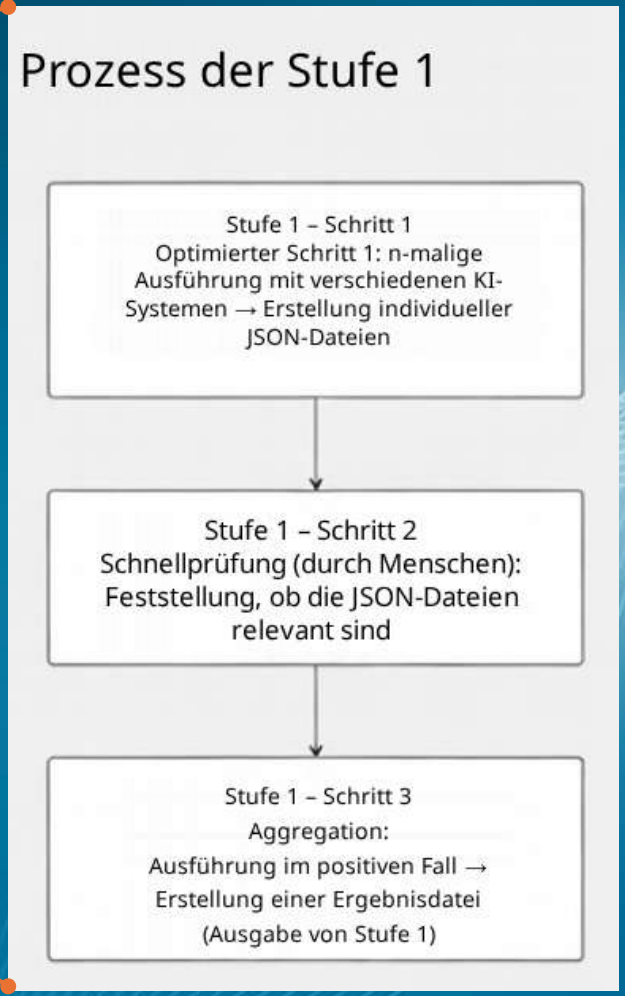
© 2026 von Thomas Walloschke, secon trust consult

Architekturübersicht



© 2026 von Thomas Walloschke, secon trust consult

Architekturübersicht



© 2026 von Thomas Walloschke, secon trust consult

Ergebnisse des Proof-of-Concept

Wir konnten Folgendes nachweisen:

- Hohe Robustheit gegenüber Halluzinationen
- Reproduzierbare Richtlinienstrukturen
- Audit-Trail-fähige Ausgaben
- Übertragbarkeit auf verschiedene Modelle
- Minimierung der Anbieterabhängigkeit

Besonders wichtig:

- Die Einhaltung des EU-KI-Gesetzes kann integriert werden.

Ergebnisse auf Github:
https://github.com/Platform-Industry-40/ZTA-Industrial-AI/tree/main/doc/de/apx/Anhang%202

© 2026 von Thomas Walloschke, secon trust consult

Warum asynchron?

Wichtig:

❌ Keine LLMs im Echtzeit-Zugriffspfad
✓ Nur für Entwicklung und Policy Engineering

Dies führt zu:

- keine Latenzprobleme
- kein Sicherheitsrisiko
- bessere Nachvollziehbarkeit

© 2026 von Thomas Walloschke, secon trust consult

Dies ist eine Momentaufnahme aus dem Jahr 2026
Fortsetzung folgt

PLATTFORM INDUSTRIE 4.0

Empfehlung für Industrie und Forschung im ZTA-Kontext

1. LLMs nur in der Engineering-Ebene einsetzen
2. Mehrstufig statt One-Shot
3. „Human-in-the-Loop“ ist zwingend erforderlich
4. Strenge Ausgabeformate durchsetzen
5. Nachvollziehbarkeit als Gestaltungsprinzip

© 2026 von Thomas Walloschke, secon trust consult

Diskussion

© 2026 von Thomas Walloschke, secon trust consult

Die KI entwickelt sich rasant –
und Europa ist bereits einem hohen Disruptionsrisiko ausgesetzt

PLATTFORM INDUSTRIE 4.0

- Die Fortschritte im Be… …m **Gamech…**
- …d China –
  …te sind selten und schwer zu entwickeln
- **I… …China**: KI wird in allen Geschäftsbereichen eingesetzt (E…wicklung, Produktion, Support, interne Prozesse), ohne auf das nächste Modell zu warten – **was zu enormen Kosten- und Zeiteinsparungen führt**

→ **Unmittelbares hohes Risiko für die europäischen Industrien:**
Ohne eine großflächige Einführung von KI
ist mit Umbrüchen in allen Branchen zu rechnen

© 2026 von Thomas Walloschke, secon trust consult

Aufruf zum Handeln: Schaffen Sie einen KI-basierten
Weg zur autonomen Produktion

- Ent
- Ent
- Ent

Ziel: Der **„KI-Innovator“** Ihres grundlegend veränderten Unternehmens zu werden

Ziel: Beschreiten Sie den **Weg hin zur autonomen** industriellen Produktion

Ziel: Nutzen Sie KI, um **Ihre vorhandenen** Kompetenzen zu verstärken

- ...ed erfunden und **umstrukturiert**

© 2026 von Thomas Walloschke, secon trust consult

Aktivitäten der Plattform Industrie 4.0

Gründung einer interdisziplinären Arbeitsgruppe mit dem Ziel:

- Entsprechende allgemeine Themen erörtern
- spezielles praktisches Material, Anleitungen und Prototypen bereitzustellen, für den Einstieg in die KI

Beiträge sind willkommen!
Bitte wenden Sie sich an: Christian Krug <krug_c@vdi.de>

© 2026 von Thomas Walloschke, secon trust consult

Vielen Dank!

Dr. Bernd Kosch (bernd.kosch@industrie-ki.de), Industrie-KI GmbH
Thomas Walloschke (thomas.walloschke@secon-tc.eu), secon trust consult
Dr. Michael Schmitt (m.schmitt@sap.com), SAP SE

Bilder: Von KI generiert oder aus öffentlichen Webquellen. Ausschließlich für Bildungszwecke. Kein Urheberrechtsanspruch. Alle Rechte liegen bei den ursprünglichen Urhebern.

© 2026 von Thomas Walloschke, secon trust consult