

Von GenAI zu Agentic AI

Was das für die Versicherungsbranche bedeutet –
jenseits des Hype

Andreas Gillhuber · &act

Versicherungsmath. Kolloquium, LMU München, 15.06.2026

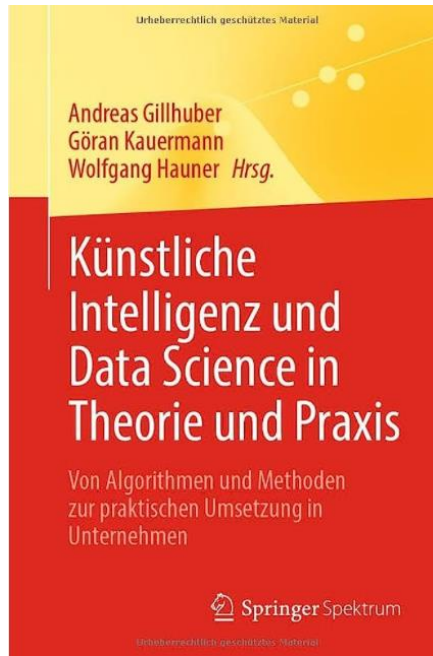




Andreas Gillhuber

CCO & Co-Founder, &act

KI-Prozessautomatisierung für den Mittelstand



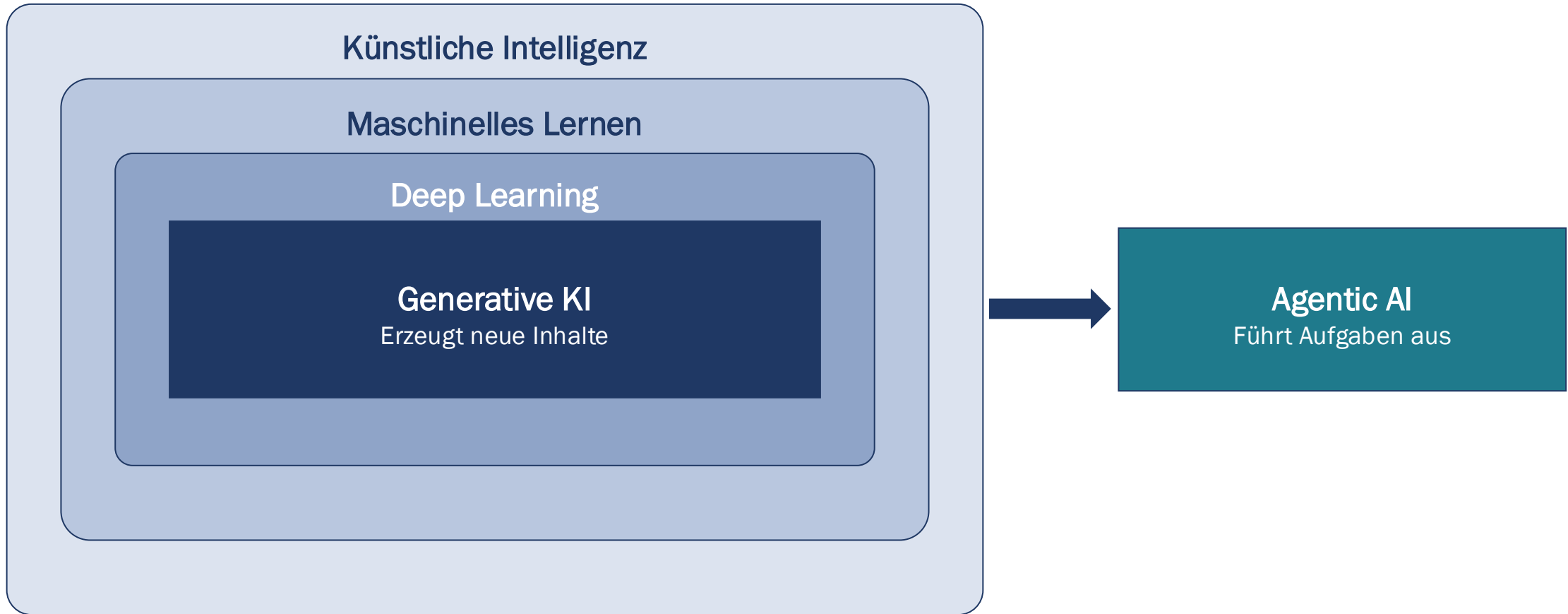
Senior Advisor Beratungen & Private Equity · CFO German Data Science Society e.V.

Herausgeber/Autor Springer · Gastdozent (Constructor Univ., Scheer School, DBU)

>30 Jahre IT & KI: IBM, Nokia/Siemens, RWE, BMW · Co-CEO Alexander Thamm

Dipl.-Ing. Elektrotechnik & IT, TU München · Exec. Program IMD Lausanne

KI, ML, Deep Learning, GenAI – und Agentic AI



KI fährt nicht gegen die Wand

Bildmodelle

präzise & kontrollierbar

Coding

längere Aufgaben

Agenten

arbeiten über Stunden

Kosten

Faktor 10–100
günstiger

Die Entwicklung hat sich im letzten Jahr beschleunigt – nicht verlangsamt

Reifegrad: von der Ambition zur Aktivierung

66 % berichten messbare Produktivitäts- und Effizienzgewinne

34 %

tiefgreifende Transformation

30 %

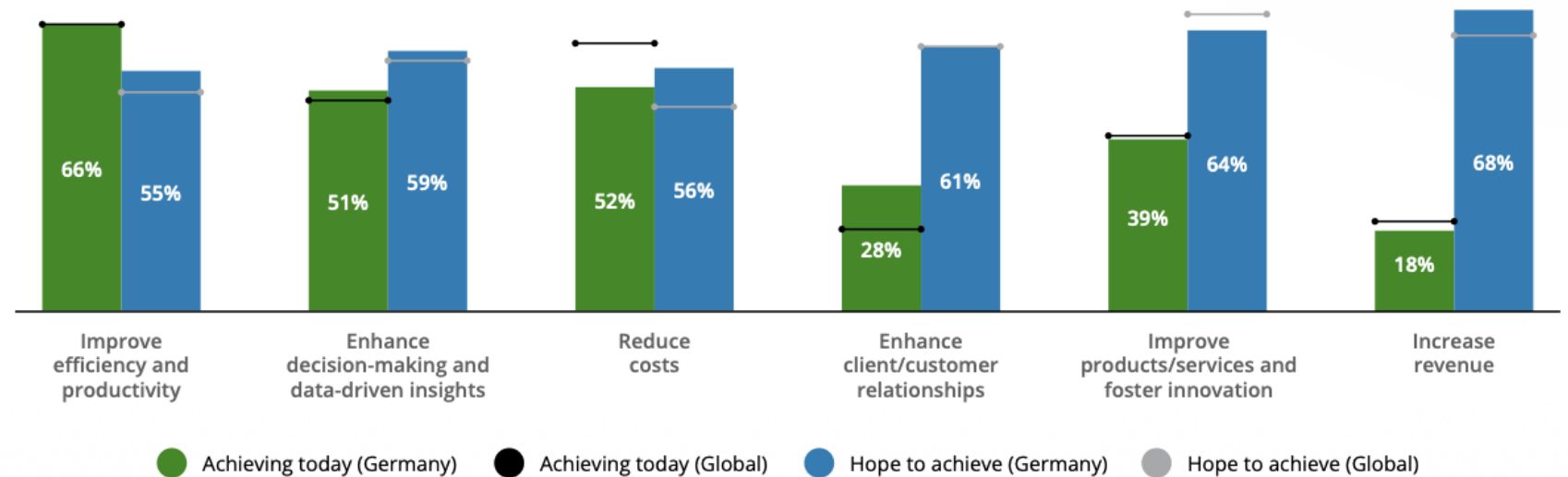
Kernprozesse umgestaltet

37 %

bleiben an der Oberfläche

Figure 2: AI value is concentrated in efficiency, while ambition is in growth

Question: With regards to benefits from your AI efforts: Which benefits are you achieving today? Which benefits do you hope to achieve?



KI-native vs. etablierte Organisationen

Etablierte Organisationen

bauen KI mühsam in bestehende Prozesse ein

KI-native Organisationen

bauen viele Prozesse gar nicht mehr
manuell auf

Produktivität läuft auseinander – kleine Teams leisten mehr, Unicorns mit weniger Personal

DACH/EU vs. USA/China

USA

Spitzenmodelle
Kapital

China

Effizienz
Hardware-Dominanz

Europa / DACH

Anwendung & Branche
Regulierung & Vertrauen

~25 % sagen: GenAI transformiert mein Geschäft bereits · 42 % strategisch vorbereitet –
Lücken bei Infrastruktur, Daten, Risiko, Talent

LLM-Leaderboard: Spitze aus den USA

Stand 06/2026

9 von 10 Modellen aus den USA

Anthropic, OpenAI und Google

China aktuell etwas weniger vertreten (Alibaba), Deepseek #20

Europa: Mistral #54, AlephAlpha

Modelle sind austauschbare Bausteine geworden

Vorsprünge schmelzen im Quartalstakt

Special Purpose Modelle aus Deutschland zB Black Forrest für Bildgenerierung, Prior Labs für SAP

Relevant: nicht „welches Modell“, sondern „welches System drumherum“

Model ↑↓	Features ↗		Intelligence ↗		Price ↗	Speed ↗	Latency ↗	End-to-End Response Time
	Context Window ↑↓	Creator ↑↓	Artificial Analysis Intelligence Index ↓		Blended USD/1M Tokens ↑↓	Median Tokens/s ↑↓	Latency First Chunk (s) ↑↓	Total Response (s) ↑↓
 Claude Fable 5 (with fallback) ♀	1M	 Anthropic	65		\$7.70	64	147.11	154.95
 Claude Opus 4.8 (max) ♀	1M	 Anthropic	61		\$3.85	60	22.76	31.04
 GPT-5.5 (xhigh) ♀	922k	 OpenAI	60		\$4.35	73	79.61	86.46
 GPT-5.5 (high) ♀	922k	 OpenAI	59		\$4.35	63	27.91	35.90
 Claude Opus 4.7 (max) ♀	1M	 Anthropic	57		\$3.85	54	17.52	26.85
 Gemini 3.1 Pro Preview ♀	1M	 Google	57		\$1.74	136	23.17	26.85
 GPT-5.5 (medium) ♀	922k	 OpenAI	57		\$4.35	59	9.38	17.89
 Qwen3.7 Max ♀	1M	 Alibaba	57		\$1.43	196	2.54	17.34
 Gemini 3.5 Flash ♀	1M	 Google	55		\$1.31	171	18.65	21.58
 Gemini 3.5 Flash (medium) ♀	1M	 Google	55		\$1.31	186	16.05	18.74

Der China-/USA-Effekt

~4 \$

US-Spitzenmodelle – je 1 Mio. Tokens

~0,20 \$

China Deepseek u.ä. – je 1 Mio. Tokens

~0 \$

China Open-Weight – je 1 Mio. Tokens

Quelle: Artificial Analysis 06/2026

Pressemitteilung zur Abschaltung der
Anthropic-Modelle Fable 5 und Mythos 5

**Keine Überraschung,
kein Weckruf.
Sondern eine
Quittung.**

 KI BUNDESVERBAND

“Was wir gestern erlebt haben, ist keine Überraschung, es ist die Quittung für Jahre des Zögerns. Die US-Regierung hat bewiesen, was wir seit 2021 sagen: KI ist strategische Infrastruktur. Wer sie nicht selbst kontrolliert, wird von anderen kontrolliert. Wir fordern keinen weiteren Strategieprozess. Wir fordern einen Gipfel. Jetzt.”



Rasmus Rothe
Vorstandsvorsitzender, KI Bundesverband

 KI BUNDESVERBAND

Souveräne KI-Infrastruktur in Europa wird real

5

geplante AI-Gigafactories
(bis 100.000 Chips)

~20 Mrd. €

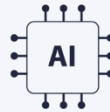
öffentliche Investition

10.000

Blackwell-GPUs – Industrial AI
Cloud (Dt. Telekom)

EuroHPC: 19 AI Factories ·
JUPITER (Jülich) als erstes
Exascale-System
EU-Verordnung 2026/150
ermöglicht die Gigafactories ·
FR/IT/DE/UK bündeln 3.000
Exaflops

EuroHPC Key Facts & Figures



19 + 13

AI Factories + Antennas



14

Supercomputers



10

Quantum Computers



62

Research & Innovation Projects



38

Participating States



8,2

Billion EUR Total Budget

Mistral: Europas produktionsreife Option



Vibe (ehem. Le Chat)

Modelle + Open-Weights,
Hosting in FR (DSGVO)

Mistral Forge

Training auf eigenen Daten,
on-premise

SAP-Partnerschaft

souveräner KI-Stack für
DE/Europa

Daneben: **Teuken 7B (Open GPT-X)** und **OpenEuroLLM** – noch experimentell



Souveränität: der wichtige Vorbehalt

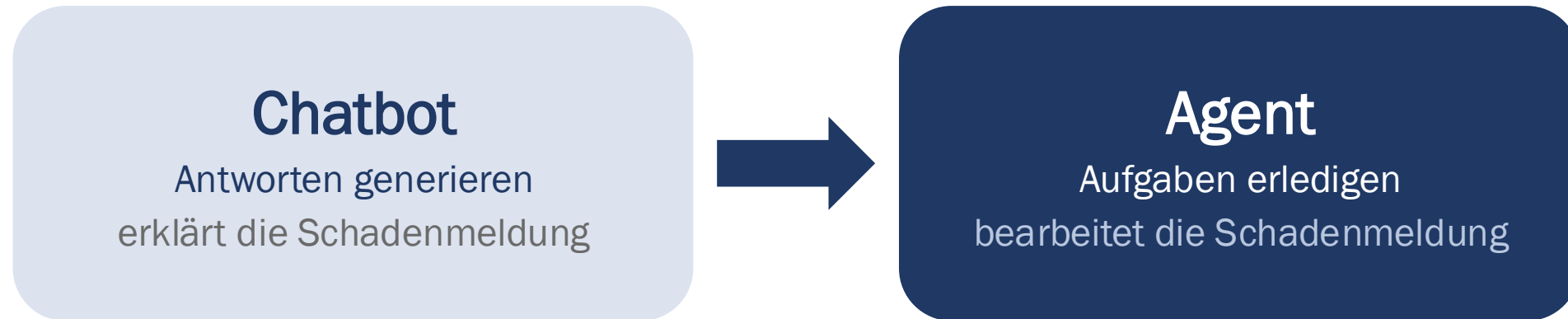
Standort ≠ Unabhängigkeit

„Souveränität“ meint 2026 oft nur den Standort der Rechenleistung – nicht echte architektonische Unabhängigkeit (Lock-in-Risiko)

Datenhoheit unter Druck

Das EU-US Data Privacy Framework steht politisch unter Druck – Datenhoheit wird wichtiger

Vom Chatbot zum Agenten



„Parallelökonomie: kleine Teams arbeiten hyperproduktiv mit Agenten-Teams.“ – Klöckner, OMR 2026

Reality-Check: Agenten zwischen Hype und Alltag

17 %

haben KI-Agenten im Einsatz

>60 %

planen es binnen 2 Jahren

„Mehr Piloten als Lufthansa – gebraucht werden Operatoren, nicht Piloten.“

Die meisten Deployments sind eng umrissen; volle Autonomie ist noch nicht reif

Warum jetzt? Treiber und Bremsen

Rechenzentren

Investitionswelle (Compute-Capex)

Energie

Energiehunger belastet Volkswirtschaften

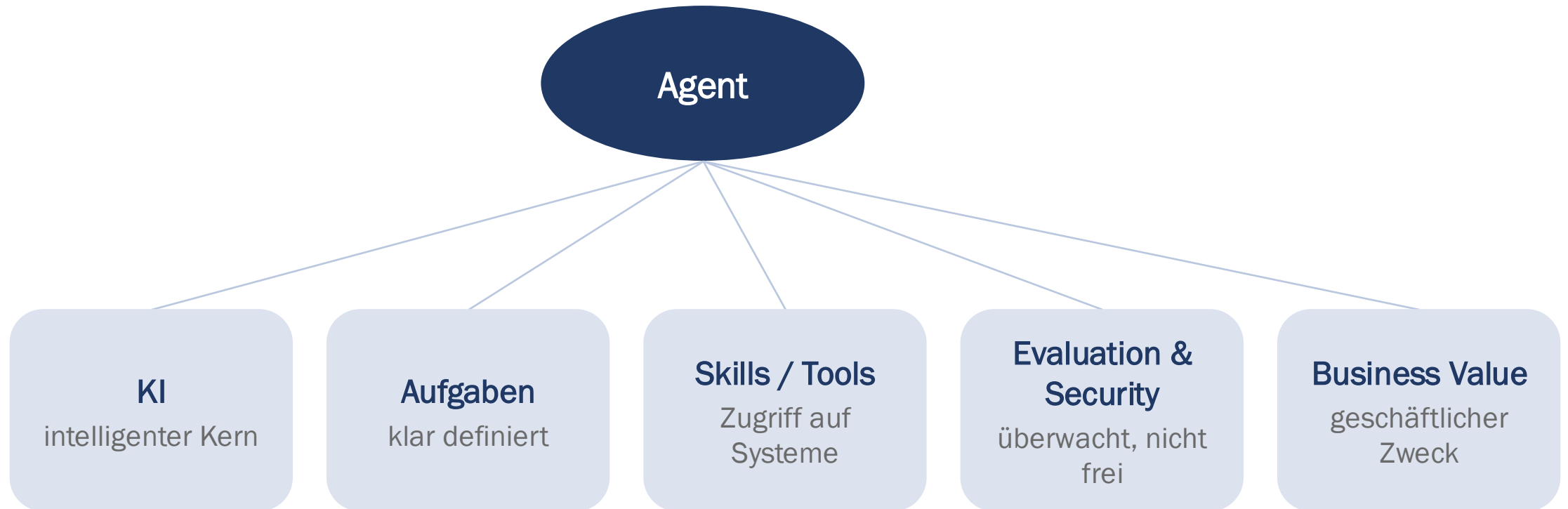
Chips

Abhängigkeit von Taiwan

Arbeit

Routinetätigkeiten unter Druck

Was ist ein Agent? Fünf Bausteine



Der &act-Assistent: von der Anfrage zur Antwort



Läuft vollständig in der EU - in der Versicherung: Anfrage = Antrag/Schaden, ERP = Bestands-/Schadensystem

Praxis (1/3): Ein Agent reicht nicht

Ein Angebotsagent

Businesslogik
+ Formatierungsanweisungen
+ Tool-Anweisungen

Ausführung verweigert

Neue Regeln ignoriert

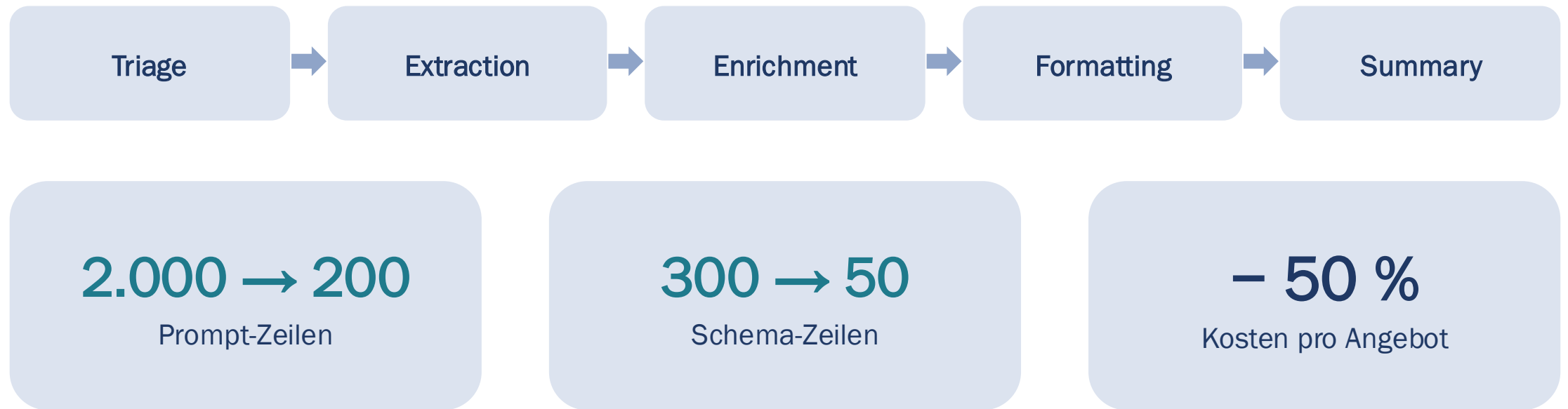
Tools wiederholt aufgerufen

Kosten & Latenz steigen

Ursache: Context Pollution

Praxis (2/3): Ein Team spezialisierter Agenten

Aufgabe aufteilen – bis zu 8 spezialisierte Agenten



Genauigkeit hoch, auch bei komplexen Anfragen

Praxis (3/3): Die wichtigsten Lektionen

Günstige Modelle

für viele Teilaufgaben gut genug

Früh evaluieren

Logik-, Regel-, LLM-Prüfung

Versionierung

jede Komponente reproduzierbar
-> Revisionssicherheit

Fallback & Retry

APIs allein nicht zuverlässig

Elektrotechnik:

DEHN SE verringert Gesamtaufwände für die Angebotserstellung um 60%

Anwendungsfall

- Täglich dutzende Preisabfragen und Anfragen zur Angebotserstellung aus dem Außendienst
- Anfragen enthalten unterschiedliche Datei-Formate und werden in verschiedenen Sprachen gestellt

Das Problem

- Angebote werden manuell in SAP angelegt, was bis zu 30 Minuten je Angebot dauert
- Über 50.000 Produkte sowie Varianten und komplexe Preisregeln müssen berücksichtigt werden

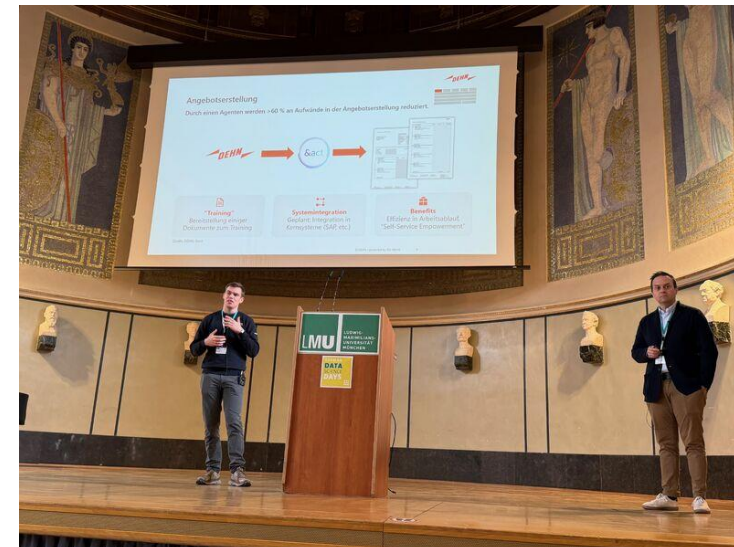
Die Lösung

- &act bearbeitet eingehende Anfragen, erstellt passgenaue Angebote und synchronisiert diese mit SAP
- Gemeinsam definierte Regeln stellen höchste Genauigkeit sicher, stellen nötigenfalls Rückfragen und erläutern Annahmen (keine „Black Box“)

Highlights:

- 60% Zeitersparnis
- Umsetzung in 3 Wochen
- Einsatz in 3 Ländern / 4 Sprachen
- Roll-out in 70 Länder und weitere Sprachen geplant

DEHN



Unser Gründer Kai Sahling zusammen mit Daniel Nowakowski, Director Data & AI bei Dehn SE, bei den German Data Science Days 2026

Retail: POS-ELECTRONIC Außendienst erstellt Angebote per Spracheingabe aus dem Auto heraus

Anwendungsfall

- Mitarbeiter im Vertrieb sind hauptsächlich im Außendienst bei den Kunden
- Unproduktive Zeit während der Autofahrten

Das Problem

- Aufgaben wie die Angebotserstellung müssen am Abend abgearbeitet werden
- Langwierige, manuelle Arbeit in Word und Excel

Die Lösung

- Per Sprachnachricht können Mitarbeiter aus dem Auto heraus Angebote erstellen und bearbeiten
- Zuvor ungenutzte Reisezeit wird jetzt produktiv genutzt

P•O•S-ELECTRONIC GmbH ELEKTRONISCHE ZAHLUNGSSYSTEME

Highlights:

- Angebote und weitere Dokumente aus dem Auto heraus per Sprachnachricht erstellen
- Umsetzung innerhalb von 2 Wochen



Öffentliche Verwaltung: Markt Weisendorf erstellt mit &act Ausschreibungen und analysiert die erhaltenen Angebote

Anwendungsfall

- Neu- und Wiederbeschaffungen in den Kommunen und Städten müssen ausgeschrieben werden
- Diese Öffentlichen Ausschreibungen sind gesetzlich geregelt

Das Problem

- Die für die Ausschreibungen relevanten Gesetze und Vorschriften für die Ausschreibungen sind umfangreich, unterliegen permanenter Änderung
- Hoher manueller, administrativer Aufwand

Die Lösung

- Massive Entlastung der Verwaltungsangestellten
- Automatisierung der Ausschreibungserstellung und Analyse der erhaltenen Angebote auf Basis der Bewertungsmatrix; Sicherstellung Compliance



Markt
Weisendorf

Highlights:

- Workflow-gestützte Ausschreibungs-Erstellung entlastet Verwaltung
- automatisierte Angebotsanalyse auf Basis Bewertungsmatrix und relevanter Gesetze
- “human in the loop“: der Mensch bleibt in der Verantwortung und Entscheidungshoheit



Unsere Gemeinde

Bürgerinfo

Leben in Weisendorf

Weisendorf erleben



Gastronomie/Catering: Weihenstephan erstellt Angebote per Sprachnachricht parallel zum Tagesgeschäft



Anwendungsfall

- Monatlich 50 Anfragen zu Veranstaltungen per Email, Telefon oder vor Ort

Das Problem

- Menüs werden individuell mit dem Kunden geplant
- Gleichzeitig organisiert die Restaurantleitung das Tagesgeschäft

Die Lösung

- Email-Anfragen werden automatisiert bearbeitet und Angebotsentwürfe erstellt
- Mitarbeiter erstellen / aktualisieren Angebote „zwischen Tür und Angel“ per Spracheingabe

Highlights:

- 12 Stunden Zeitersparnis pro Woche
- Umsetzung innerhalb von 2 Wochen



Was davon in Ihre Welt passt

Angebot / Underwriting

Anfrage → Tarifierung → Entwurf

Antrag

Dokumente lesen, prüfen, anlegen

Schaden

Meldung + Anhänge → Ersteinschätzung

Korrespondenz / Service

Eingang klassifizieren, Antworten

GenAI ist Tischeinsatz – der Vorteil entsteht durch verlässliche Agentensysteme mit Regeln und EU-Betrieb

Neue Cyberrisiken & Schatten-KI

Angriffe

Phishing glaubwürdiger,
skalierbarer

Schatten-KI

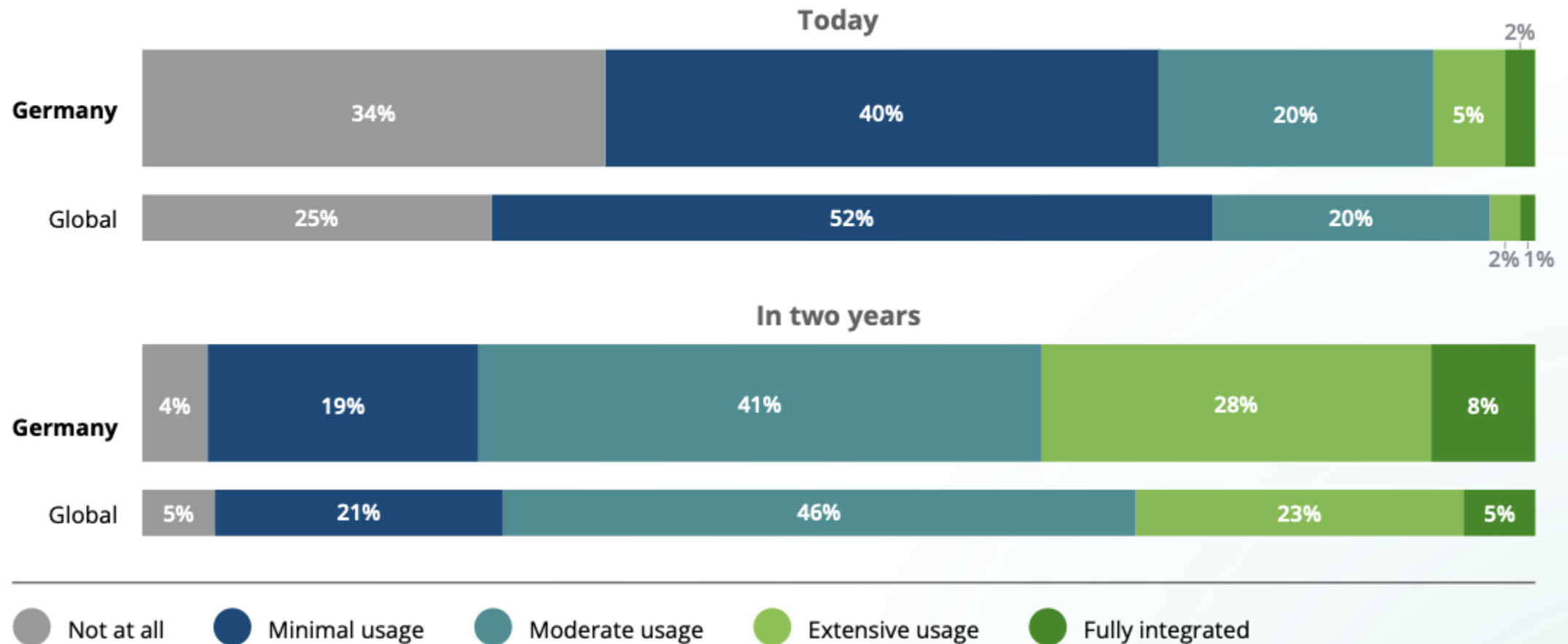
Daten in externe Tools ohne
Regeln

Verteidigung

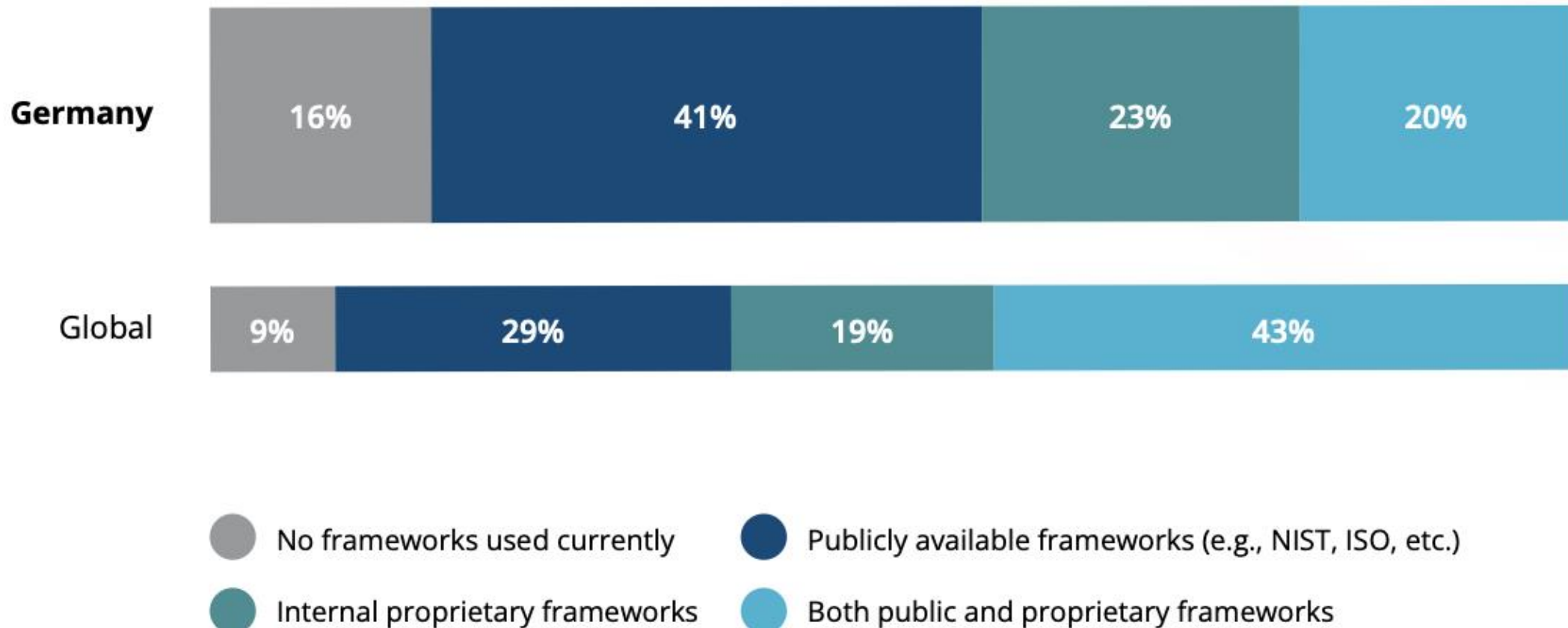
KI hilft bei Log-Analyse &
Priorisierung

Antwort: Governance – erlaubte Tools, Datenklassen, Freigaben, menschliche Kontrolle

Agentische KI zieht rasch in die dt. Unternehmen ein...



... aber die nötige Governance ist noch nicht etabliert



EU AI Act & Versicherung

Hochrisiko (Annex III)

Annahme Leben, Prämien,
Schadenbewertung

Pflichten

Risikomanagement, Data Governance,
Nachvollziehbarkeit, Aufsicht

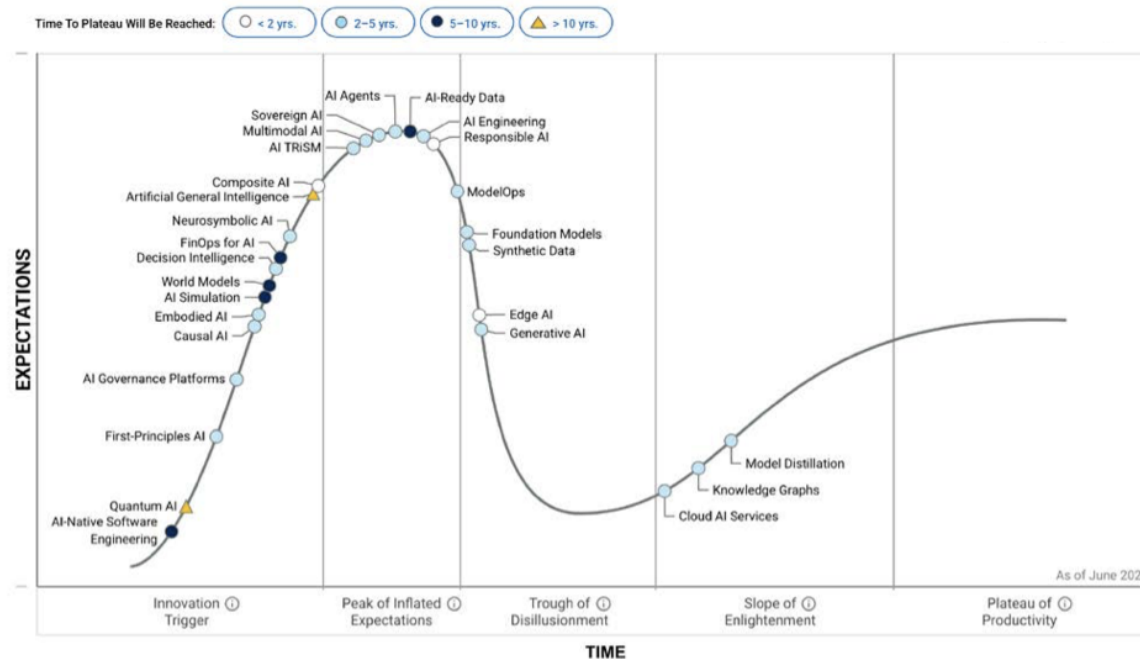
&act-Fit

Regelwerk, Human-in-the-Loop,
Versionierung, EU-Betrieb

Frist: ursprünglich 2. Aug. 2026 – Digital Omnibus verschiebt Annex-III-Pflichten ggf. bis Dez. 2027
Stand vor dem Termin prüfen – Zeitplan ist in Bewegung

Gartner Hype Cycle für KI, 2025

Hype Cycle for Artificial Intelligence, 2025



Source: Gartner

About this Hype Cycle

AI investment remains strong, but focus is shifting from GenAI hype to foundational innovations like AI-ready data, AI agents, AI engineering and ModelOps. This research helps leaders prioritize high-impact, emerging AI techniques while navigating regulatory complexity and operational scaling.

Our top 3 positions

1. On AI Agents

By 2029, over 70% of enterprise agentic AI initiatives will fail due to widespread agent-washing missing the importance of use-case optimization, user risk tolerance, compliance and audit requirements.

2. The Foundations of Generative and Agentic AI

In generative AI, speed amplifies what's underneath it. Without AI-ready data and solid AI engineering foundations, fragility grows faster than value.

AI-ready data for GenAI and agentic AI demands urgent evolutions in data management, architecture, platforms, skills and processes.

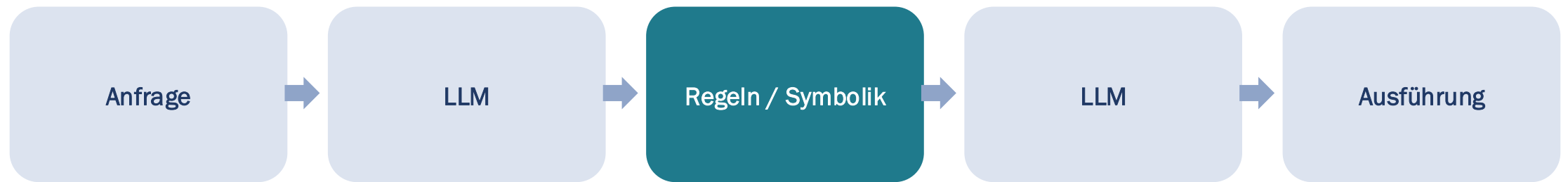
3. AI and the Workforce

AI won't trigger a jobs apocalypse — it will unleash job chaos. Starting in 2028-2029, AI will create more jobs than it eliminates, yet over 32 million roles will be transformed significantly each year.

Gartner Top 10 Strategische Technologietrends 2026



Semantic AI / Neurosymbolik: keine Black Box

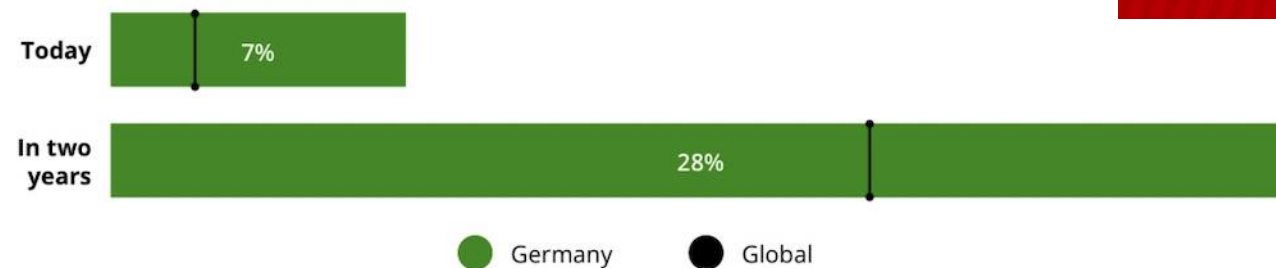


Physical AI / Robotics

~80 %

Physical-AI-Adoption binnen 2 Jahren

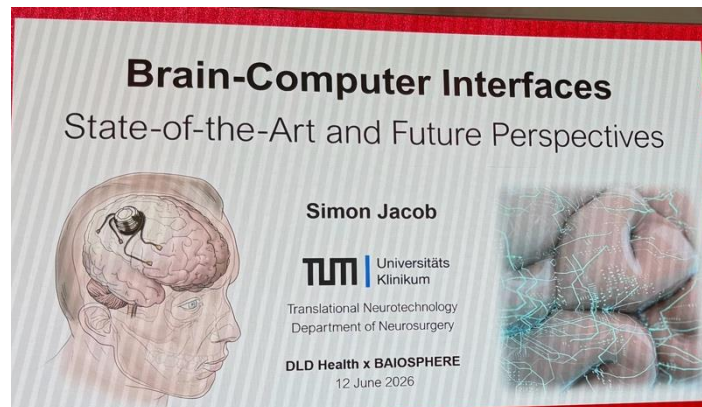
Figure 3: Germany's Physical AI lead is expected to widen



Question: To what extent is your organization utilizing physical AI (e.g., robotics, automated machinery) in its operations?
Advanced usage combines extensive and fully integrated use of Physical AI.



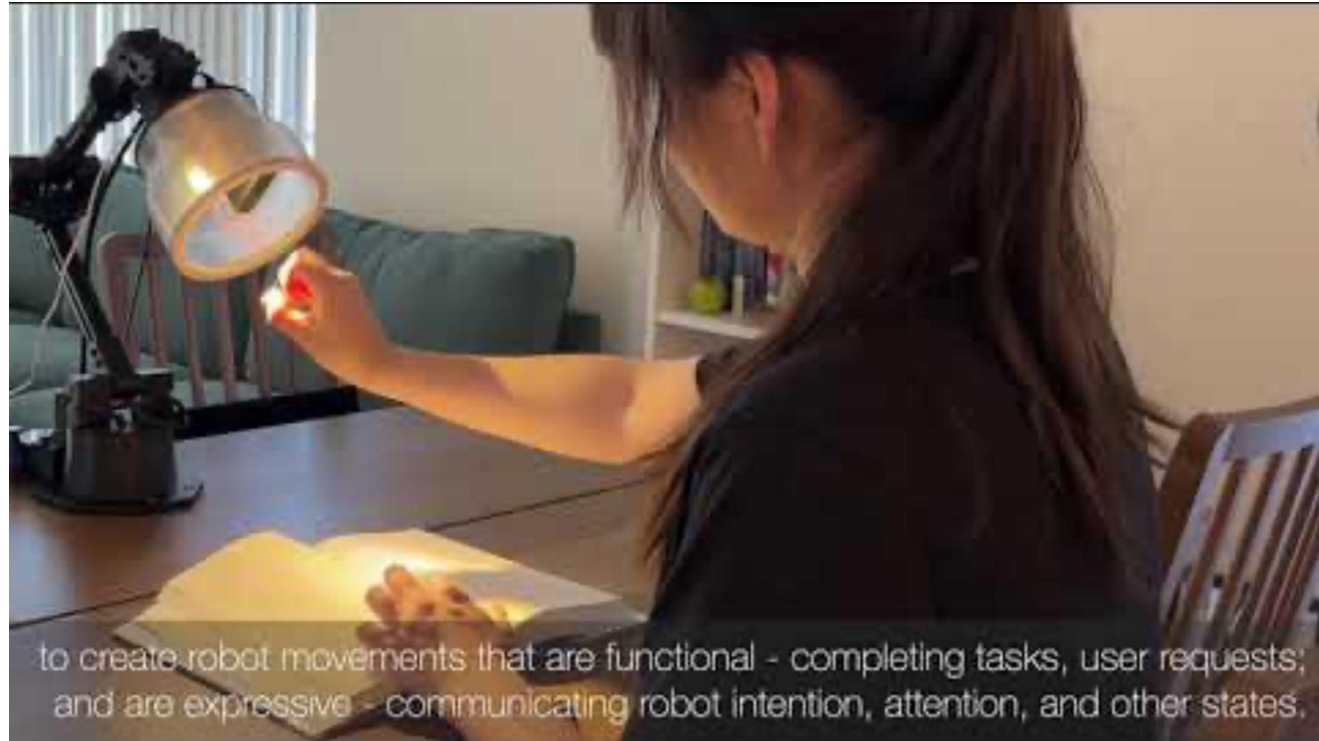
Health Tech @Bayern



Die „expressive“ Lampe

Lampenartiger Roboter:
Bewegung mit Funktion
UND Ausdruck

Studie: ausdrucksstarke
Bewegung steigert
Engagement – v. a. bei
sozialen Aufgaben



to create robot movements that are functional - completing tasks, user requests;
and are expressive - communicating robot intention, attention, and other states.

Die offenen Fragen



Energie

Energiehunger der KI



Hardware

Chinas Dominanz



Europas Rolle

im Technologierennen

Europas Karte = regulierte, vertrauenswürdige Anwendung = Ihr Spielfeld

Zum Mitnehmen

„Weniger Piloten, mehr Operatoren.“

„Architektur und Regeln schlagen das größte Modell.“

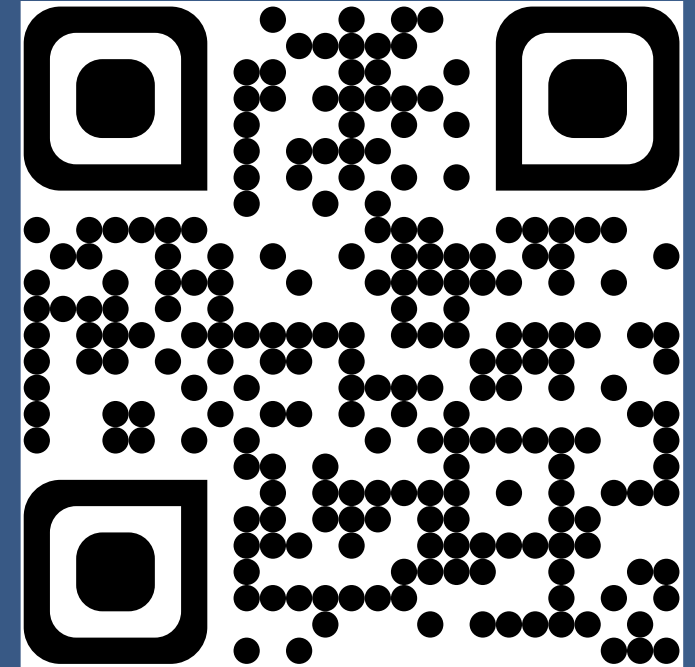
„Der Mensch entscheidet – die Maschine erledigt.“

Vielen Dank für Ihr Interesse!



Andreas Gillhuber
Co-Founder & CCO

andi@andact.io
+49 160 5302420
[linkedin.com/in/agillhuber](https://www.linkedin.com/in/agillhuber)



**Scannen und Demo
vereinbaren**