



KI-Reporting im Finanzbereich: Was heute geht – und was (noch) nicht

06. Oktober 2026

Vortragsteam



Dr. Marko Helwich
Generali Deutschland



Dr. Alexander Raisch
Generali Deutschland



Dr. Christian Jönen
B&W Deloitte

Deloitte.

Agenda

- 1 | Wo steht Generative KI heute?**
- 2 | Use Case Smart Reporting: Vom Konzept zur Umsetzung**
- 3 | Nächste Evolutionsstufe: Agenten, Automation und Datenplattformen**
- 4 | Ausblick & Diskussion**

Künstliche Intelligenz entfacht ein enormes Potenzial und ist ein Wachstumstreiber

Wir sind umgeben von KI und der Einfluss wird immer größer.

TECH

Nvidia's 70% growth forecast puts it on track to become tech's No. 2 company by revenue

PUBLISHED WED, AUG 26 2026-8:20 PM EDT | UPDATED THU, AUG 27 2026-1:11 AM EDT

Kif Leswing
@KIFLESWING

SHARE [f](#) [x](#) [in](#) [e](#)

KEY POINTS

- Nvidia shocked the AI world on Wednesday when, in the middle of its second-quarter earnings report for fiscal 2027, said that it forecast 70% growth in the company's fiscal 2028.
- That was significantly higher than Wall Street analysts were expecting for the period ending Feb. 1 in 2028.
- Based on the consensus projection for revenue of \$396 billion for fiscal 2027, which ends in January, sales next year would hit \$673 billion. That would put Nvidia ahead of Apple and Alphabet, according to Wall Street projections, and behind only Amazon — still primarily a retailer — among U.S. tech companies.

TRENDING NOW

U.S. Gre
Indi
Mo

[Nvidia 70% growth forecast puts it on track to be tech No. 2 company](#)

BUSINESS INSIDER

Elizabeth Holmes Doc | AI Feud | Trump's \$500 Checks | Small Business Insider | Sign up | The Long Play

TECH

Google says 75% of the company's new code is AI-generated

By Hugh Langley [+ Follow](#)



Google CEO Sundar Pichai and DeepMind CEO Demis Hassabis, Raj K Rajhansdan Times via Getty Images

Apr 22, 2026, 3:17 PM GMT+2

Three-quarters of new code created inside Google is now generated by AI and reviewed by human engineers, the company said Wednesday.

[Google Says 75% of the Company's New Code Is AI-Generated - Business Insider](#)

Versicherungsbote

News | Schwerpunkt | Dienste | Vergleichsrechner | Rechner | Über uns | Fachmagazin | Newsletter | Podcast | Videos

NETZWELTEN

02.08.2024

„Auch 2030 werden noch Menschen Daten eintippen, weil nicht überall Dunkelverarbeitung funktioniert“

Digitalisierung und Dunkelverarbeitung werden auch 2030 nicht so weit sein, wie sie sein sollten, meint Digitalisierungs-Experte Dr. Sebastian Schulz. Im Exklusiv-Interview spricht der ehemalige Sach-Vorstand der VHV-Gruppe über Schadensprozesse und -meldungen der Zukunft und Spracheingabe als ganz normale Kundenschnittstelle.

[„Auch 2030 werden noch Menschen Daten eintippen, weil nicht überall Dunkelverarbeitung funktioniert“ - Netzwelten - Versicherungsbote.de](#)

Sachschäden

Neuronale Netze spüren Versicherungs-Betrüger auf

Versicherer automatisieren ihre Abläufe immer stärker. KI-Technologie sorgt dabei für mehr Effizienz und erkennt Manipulationen. Doch noch setzt die Regulierung einige Grenzen.

Thilo Großer
16.09.2026 - 10:43 Uhr

9 Min. [u](#) [p](#) [b](#) [g](#) Bei Google bevorzugen

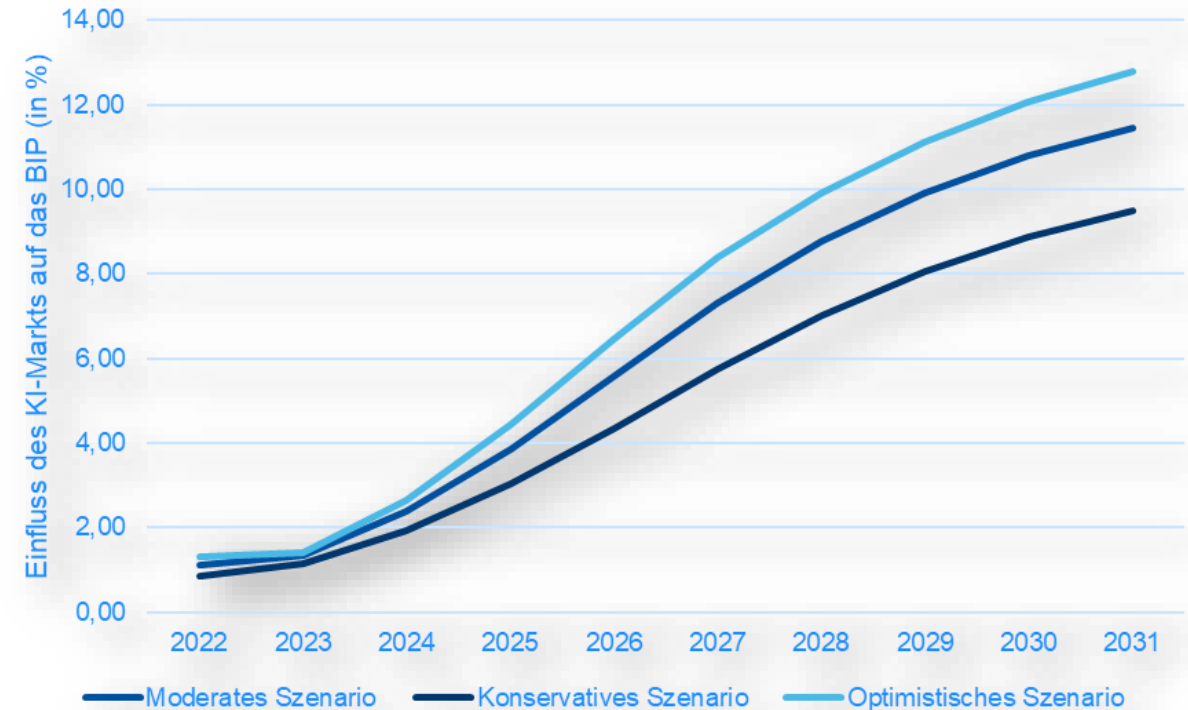


Sturmschaden: Versicherer sparen bei der Schadenregulierung mit KI viele Kapazitäten. Foto: Getty Images

[Sachschäden: Neuronale Netze spüren Versicherungs-Betrüger auf](#)

Der Markt für künstliche Intelligenz wächst kontinuierlich

KI ist eine neue Basistechnologie – KI ist gekommen, um zu bleiben ...



<https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide?currency=EUR>

Schlüsselfaktoren für die Einführung und das Wachstum von KI

KI-Anwendungen werden durch verschiedenste Entwicklungen und Enabler getrieben

Technologischer Fortschritt

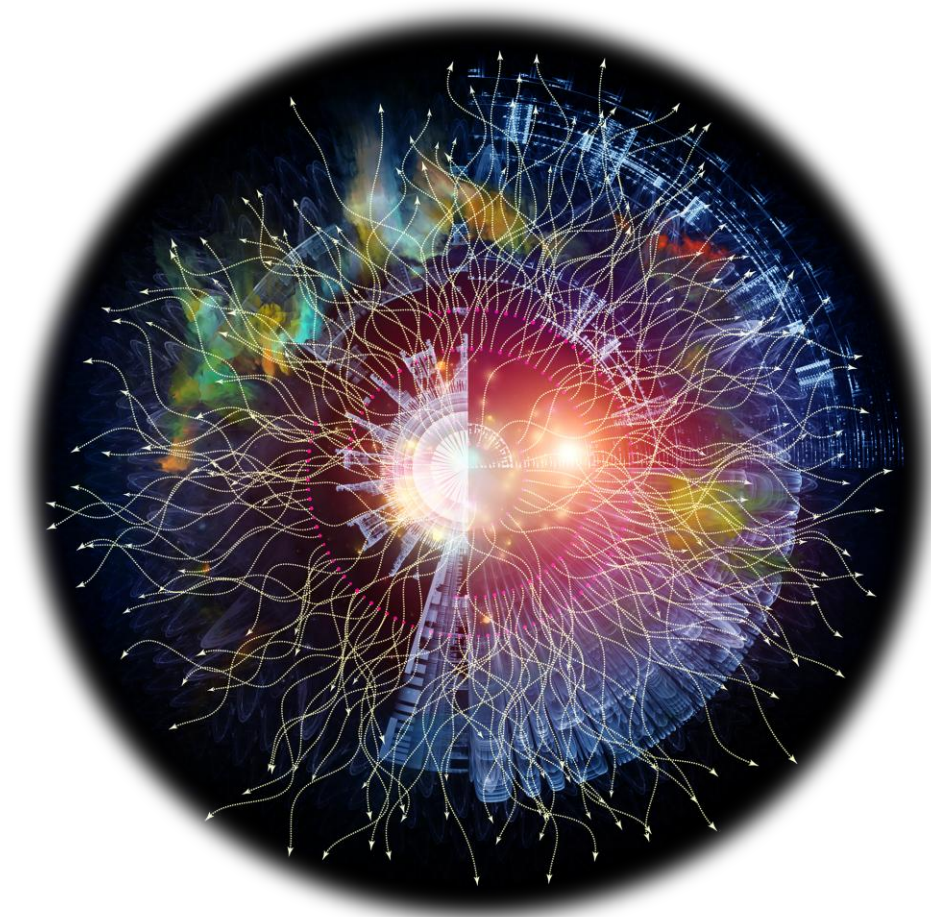
Steigerungen der **Rechenleistungen** und **effizientere Datenspeicherung** ermöglichen das Potenzial von großen Modellen und der Verwendung von großen Datenmengen zu entfalten.

Verfügbarkeit von Big Data

Die starke Zunahme der **Verfügbarkeit von Daten** bildet die Basis für die Entwicklung datengetriebener Ansätze und Modelle.

Öffentliche Wahrnehmung und Ausbildung

Die **Bevölkerung wird mit KI** vertrauter und es werden mehr und mehr Ausbildungen und Schulungen angeboten.



Cloud Computing

Cloud-Plattformen bieten skalierbare Ressourcen, um kosteneffizient KI-Anwendungen schnell bereitzustellen.

Breite Anwendung & Förderung

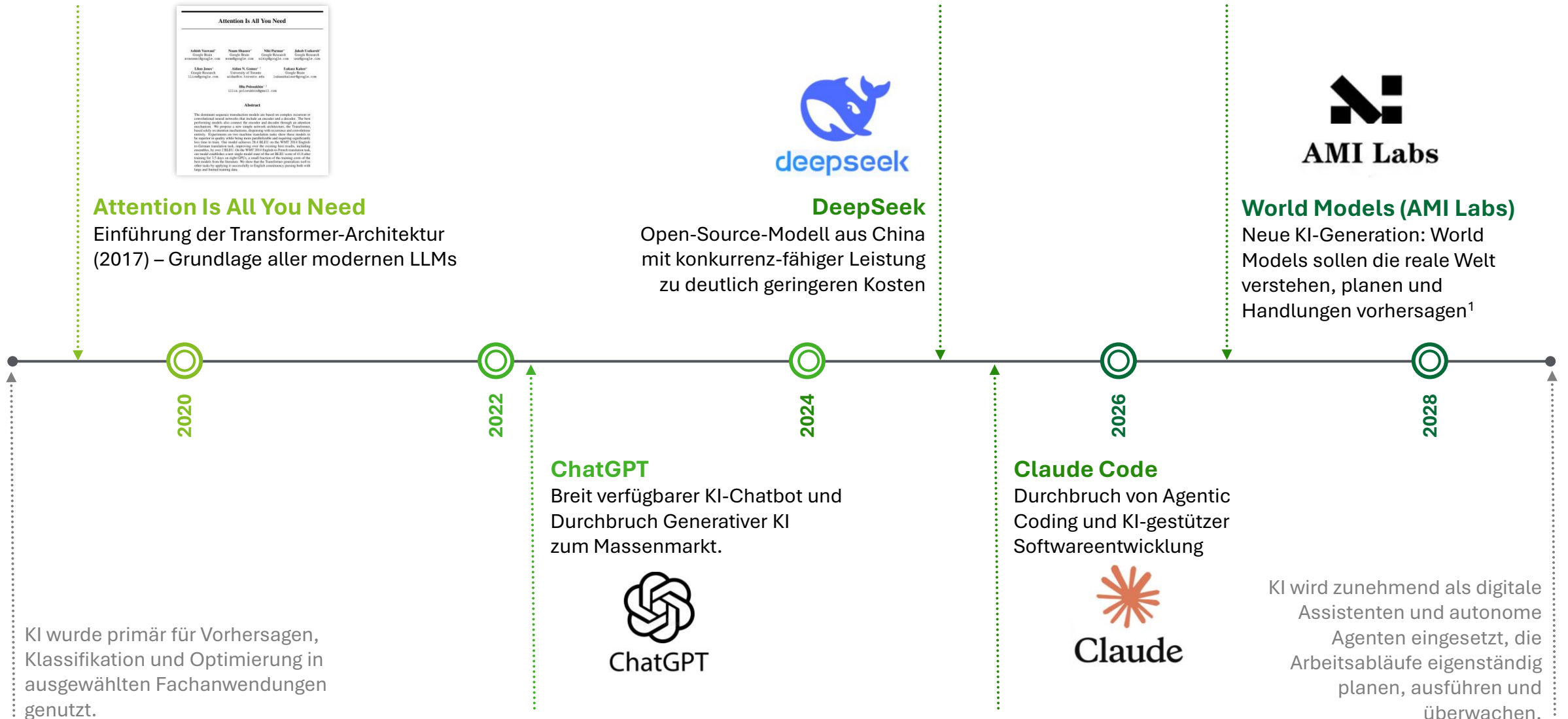
Verschiedene Industriezweige sowie private und öffentliche **Finanzierungen** fördern die Umsetzung von KI-Lösungen, wodurch „Business“-getriebene Anwendungen aber auch Forschung & Entwicklung stark gefördert werden.

Regulatorische Unterstützung

Regierungen haben das Potenzial von KI erkannt und schaffen **Richtlinien und Rahmenbedingungen**, um die vertrauensvolle Entwicklung und Integration zu unterstützen (vgl. EU AI Act).

Die Evolution der Künstlichen Intelligenz

Von Transformer-Modellen über Generative KI bis hin zu autonomen KI-Agenten und World Models.



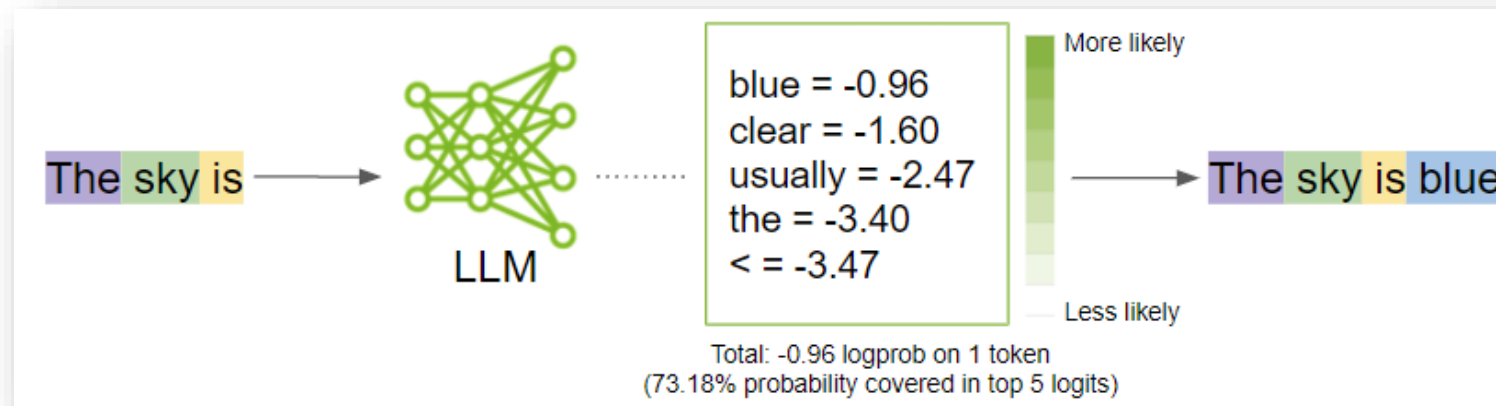
¹ 1,03 Mrd. USD Seed bei 3,5 Mrd. USD Bewertung

Was sind LLMs?

Es ist hilfreich zu wissen, wie ein LLM funktioniert.

Large Language Models (LLMs) sind...

- **Leistungsstarke Modelle**, die darauf ausgelegt sind, Sprache zu verstehen und zu generieren.
- **Größe:** Sie bestehen aus Milliarden von Parametern (lernbare Gewichte im neuronalen Netz), z.B. GPT4 ~1,76 Billionen Parameter
- **Training:**
 - LLMs werden auf Textdaten aus dem Internet, Büchern, wissenschaftlichen Artikeln, Foren, Code, ... trainiert.
 - **Ziel ist es**, Wahrscheinlichkeiten für das nächste Wort vorherzusagen – dadurch lernen sie Grammatik, Stil, Fakten und sogar logisches Denken.



[How to Get Better Outputs from Your Large Language Model | NVIDIA Technical Blog](#)

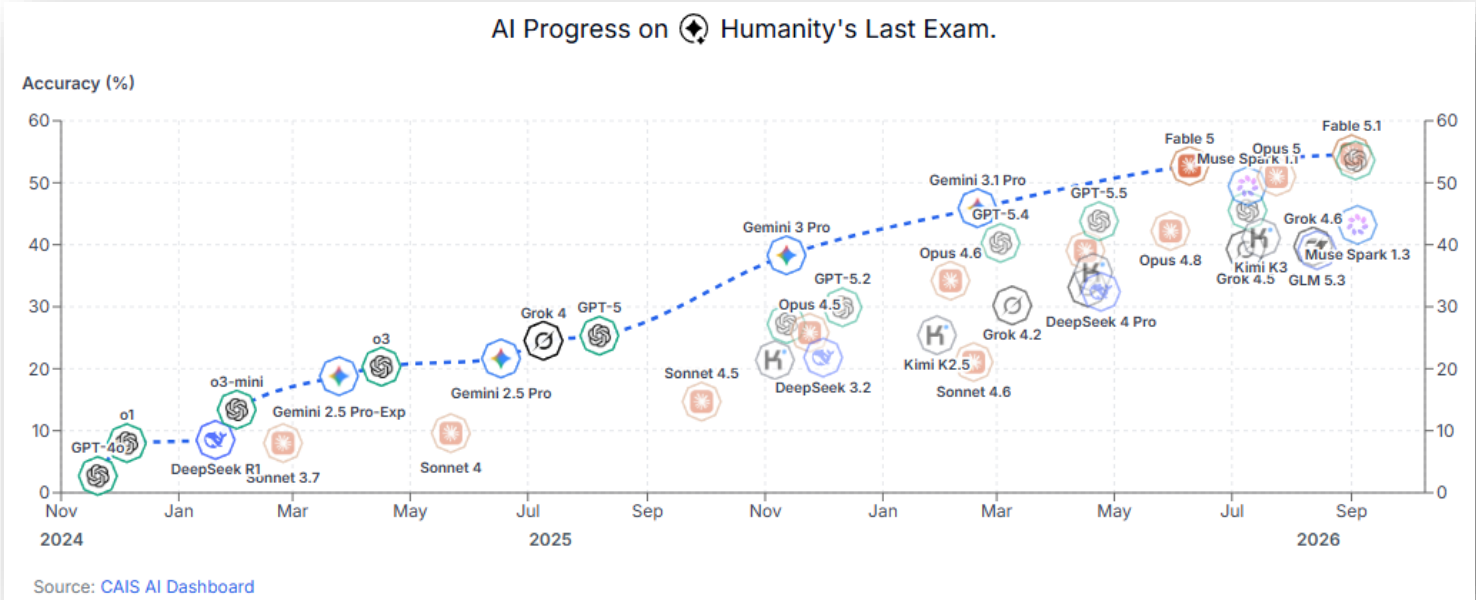
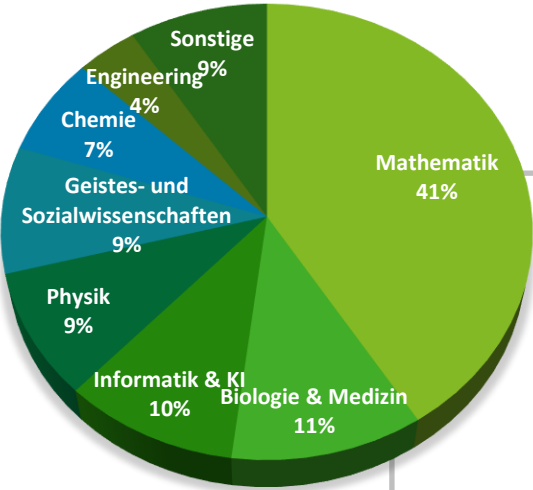
Humanity's Last Exam: Der ultimative Test für Künstliche Intelligenz

Ist „Humanity's Last Exam“ wirklich der letzte Test?

Humanity's Last Exam ist ein Benchmark mit **2.500 komplexen** und umfassenden **Expertenfragen zur Messung der Fähigkeiten moderner KI-Modelle**.

Modelle beantworten die Aufgaben, und die **Trefferquote dient als objektiver Vergleich** ihrer Wissens- und Denkfähigkeiten.

[Humanity's Last Exam](#)



Classics

Question:



Here is a representation of a Roman inscription, originally found on a tombstone. Provide a translation for the Palmyrene script. A transliteration of the text is provided: RGYN^o BT HRY BR ^eT^o HBL

Eine Inschrift in der antiken Sprache Palmyrenisch soll übersetzt werden. Dafür werden Kenntnisse alter Sprachen, Epigraphik und historischer Quellen benötigt.

Ecology

Question:

Hummingbirds within Apodiformes uniquely have a bilaterally paired oval bone, a sesamoid embedded in the caudolateral portion of the expanded, cruciate aponeurosis of insertion of m. depressor caudae. How many paired tendons are supported by this sesamoid bone? Answer with a number.

Es wird nach einer sehr speziellen anatomischen Eigenschaft von Kolibris gefragt. Die Antwort ist lediglich eine Zahl, erfordert aber hochspezialisiertes Fachwissen aus der vergleichenden Anatomie.

Vergleich ausgewählter LLMs

Modelle unterscheiden sich zunehmend nach Leistungsfähigkeit, Preisniveau, Kontextfenster und Reasoning-Fähigkeit

Provider		Model	Tier	\$/1 M tok, in/out	Context	Reasoning
OpenAI	OpenAI	GPT-6 Astra	Frontier	10.00 / 50.00	1M	yes
OpenAI	OpenAI	GPT-6.1 Sol	Balanced	2.00 / 10.00	1M	yes
OpenAI	OpenAI	GPT-6 Luna	Cheapest	0.10 / 0.50	1M	yes
ANTHROPIC	Anthropic	Claude Fable 5.1	Frontier	10.00 / 50.00	1M	yes
ANTHROPIC	Anthropic	Claude Sonnet 5.5	Balanced	2.00 / 10.00	1M	yes
ANTHROPIC	Anthropic	Claude Haiku 4.5	Cheap + fast	1.00 / 5.00	200K	yes
 Gemini	Google	Gemini 3.1 Pro	Frontier	2.00 / 12.00	1M	yes
 Gemini	Google	Gemini 3.1 Flash-Lite	Cheap + fast	0.30 / 2.50	1M	yes
	Mistral	Mistral Small 4	Open weights	0.15 / 0.60	256K	yes
 Meta	Meta	Muse Spark 1.3	Frontier	1.25 / 4.25	1M	yes
 Meta	Meta	Llama 4 Scout	Open weights	self-hosted	10M	no

Standard-Listenpreise in USD pro 1 Million Tokens, basierend auf den offiziellen Preisübersichten der jeweiligen Anbieter, geprüft am **02. Oktober 2026**; bei **Gemini 3.1 Pro** verdoppeln sich die Kosten bei Eingaben mit mehr als **200.000 Tokens**.

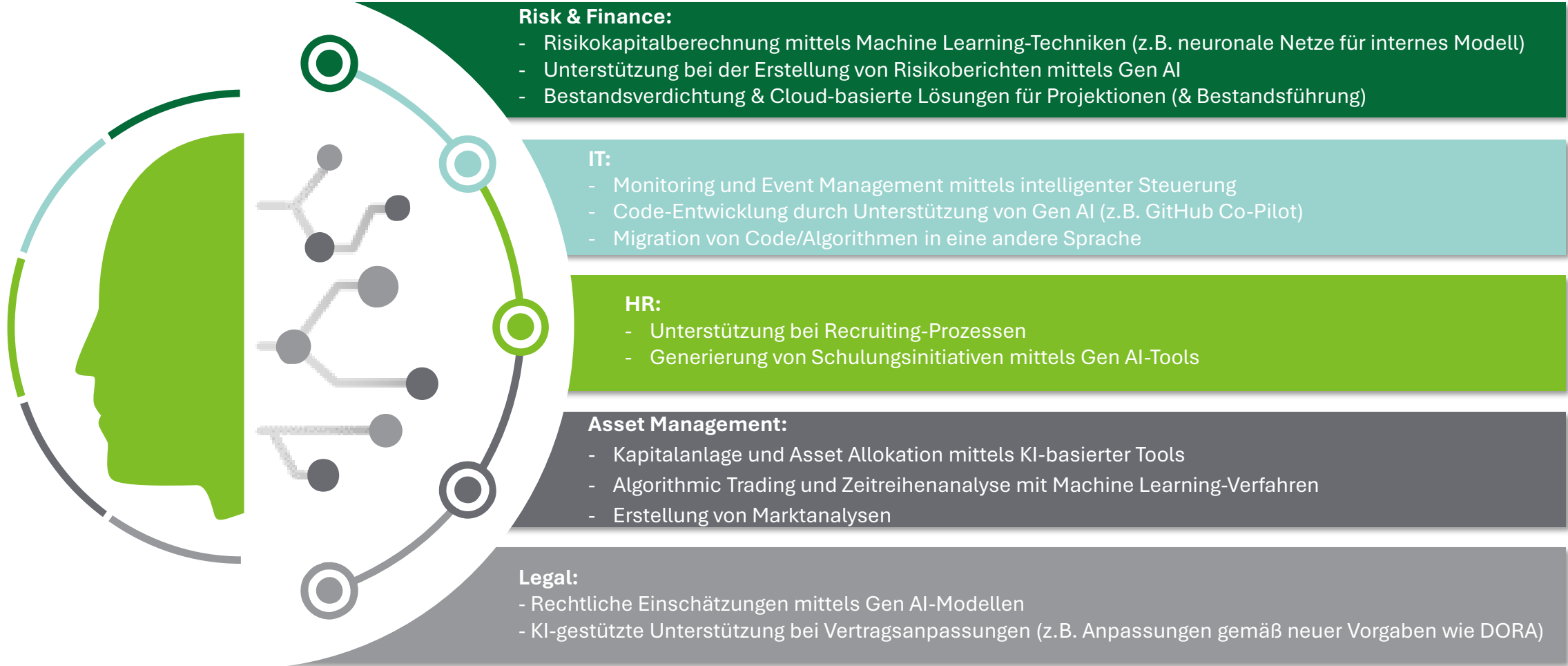
Use Cases entlang der Wertschöpfungskette

Überblick über verschiedene Anwendungsfälle in der Versicherung (1/2)



Use Cases entlang der Wertschöpfungskette

Überblick über verschiedene Anwendungsfälle in der Versicherung (2/2)



DAV Initiative

DAV gründet AG GenAI zur Erkundung aktuarieller Potenziale

**DAV**
Deutsche
Aktuarvereinigung e.V.

Über unsWissenEventsNewsroomMeine DAV

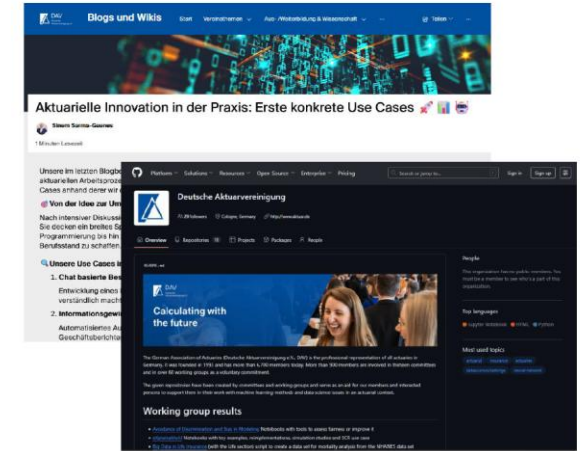
AG Generative AI

Die AG *Generative AI* konzentriert sich auf die Anwendung von Generativer Künstlicher Intelligenz im Aktuariat und in der Versicherungsbranche. Sie untersucht, wie Generative KI-Methoden genutzt werden können, um komplexe Modelle zu entwickeln, neue Versicherungsprodukte zu kreieren und Vorhersagen zu verbessern. Ein besonderes Augenmerk liegt auf der ethischen Nutzung und den Auswirkungen dieser Technologien auf die Branche. Die AG arbeitet an praktischen Lösungen und Anwendungsfällen, die Aktuarien dabei helfen, die Potenziale von Generative AI zu verstehen und umzusetzen.

[Actuarial Data Science - Deutsche Aktuarvereinigung e.V.](#)

Ausblick

- Inhalte, Begrifflichkeiten und Erfahrungen über DAV-Blogposts veröffentlichen
- Veröffentlichung von Use-Case-Umsetzungen in Form von Notebooks auf Github
- Überprüfung der Use-Case-Auswahl
- Ausführlicher Ergebnisbericht
- Abschluss Mitte 2026

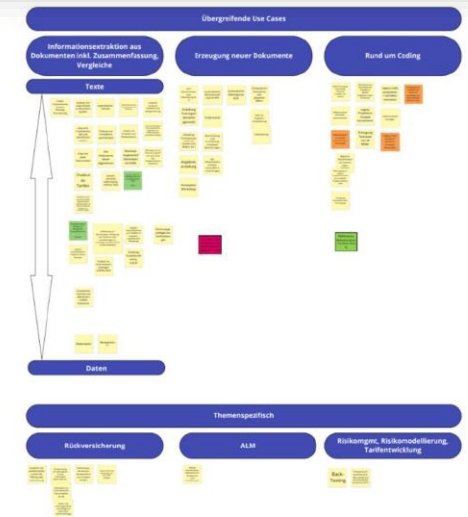
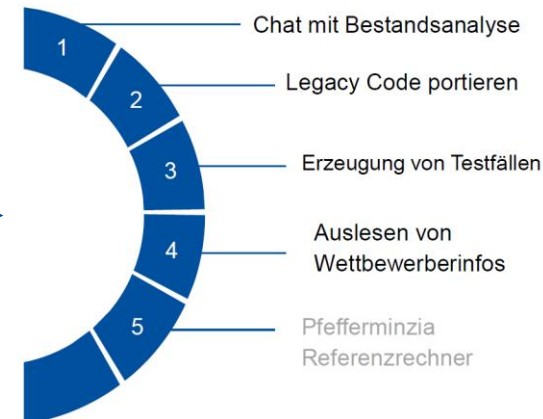


Vorstellung der AG und Zielsetzung

- Ausschreibung Oktober 2024
- Start Januar 2025
- Zielsetzung:
 - Einsatzfelder von GenAI im Aktuariat?
 - Sinnvolle Umsetzbarkeit? Wie?
 - Risiken und Grenzen?
 - verschiedenen Sparten
 - Originär actuarielle Aufgaben
 - "Zügige Beendigung" (Mitte 2026)

- Leitung: Andreas Döring, Zoran Nikolic
- Mitglieder: Alexander Bernert, Matthias Fahrenwaldt, Christian Jonen, Felix Jordan, Christoph Krischanitz, Marko Lekic, Steffen Lotter, Lucas Reck, Janina Schimmel, Linda Steffen-Raphael, Marvin Westmeyer
- Organisation in UAGs ("Schreiben" sowie pro Use Case)
- z.Zt. 14-tägiger Turnus

Auswahl der Use Cases



Ergebnisbericht DAV AG GenAI

Grundlagen, Praxisbeispiele und Erkenntnisse für das Aktuariat

Einsatz von generativer
künstlicher Intelligenz im
Aktuariat

Ergebnisbericht



Einordnung generativer KI

- Definition generative KI (GenAI), Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Reasoning
- Einsatzfelder von GenAI
- Chancen, Grenzen und Halluzinationen
- Governance, Datenschutz und Validierung

Spezifische Use Cases

- Legacy-Code-Portierung
- Erstellung versicherungstechnischer Testbestand
- Markt- und Wettbewerbsanalyse
- Daten-Chatbot für Bestands- und Schadenanalysen



Der Bericht vermittelt ein **gemeinsames Verständnis von GenAI** und zeigt anhand konkreter **Use Cases**, wie aktuarielles Fachwissen, Daten und generative KI gemeinsam Mehrwert schaffen können.

Agenda

- 1 | Wo steht Generative KI heute?
- 2 | **Use Case Smart Reporting: Vom Konzept zur Umsetzung**
- 3 | Nächste Evolutionsstufe: Agenten, Automation und Datenplattformen
- 4 | Diskussion

Use Case Smart Reporting

Motivation

Alex, ich brauche eine Kommentierung der Own-Funds-Entwicklung für den SFCR. Sollte ja mit KI schnell gemacht sein. Daten sind strukturiert und Stichpunkte zu Bewegungen liegen vor:

Veränderung der Solvency II Eigenmittel im Berichtszeitraum					
AoF	Available Own Funds	2024 YE	2025 YE	Delta abs.	Delta in %
Tier1	Tier 1:	5.819	6.267	448	8%
Gezeichnetes Kapital	Ordinary share capital	71	71	-	-
Emissionsagio	Share premium account related to ordinary share capital	4	4	-	-
RR	Reconciliation Reserve	5.412	5.870	458	8%
GAAP	thereof local GAAP revenue reserve and share premium account according to §272 (2) No. 4 local GAAP	196	196	-	0%
PVFP	thereof PVFP after TVOG (before tax)	2.390	3.141	751	31%
GCR	thereof Going Concern Reserve (before tax)	5.531	5.235	-297	-5%
MVM	thereof Market Value Margin	-600	-645	-46	8%
DTL	thereof Deferred Tax Liabilities	-2.104	-2.055	49	-2%
Surplus Fund	Surplus Funds	332	322	-10	-3%
Tier2	Tier 2: Subordinated Liabilities & AOF	-	-	-	-
MVSL	thereof Market Value of Subordinated Liabilities & AOF	-	-	-	-
NE	thereof not eligible due to constraints	-	-	-	-
Tier3	Tier 3: Deferred Tax Assets	-	-	-	-
AoF	Available Own Funds	5.819	6.267	448	8%

Haupttreiber OF-Veränderung			
AoF-Kategorie	LAMP-Kategorie	Impact	Kommentar
PVFP	1 / Opening adjustment	24	Anstieg der langfristigen Aktionärservartung
PVFP	5 / Chg. in future oper. Assumptions	-2	Aktualisierung der Annahmen für Storno und ein Anstieg der erwarteten zukünftigen Beitragsfreistellungsraten zu einem Rückgang des PVFP vor Steuern.
PVFP	1 / Opening adjustment	-3	Rückgang durch neue Parametrisierung für die Neuanlage in der Projektion, die mehr risikobehaftete Bonds vorsieht.
PVFP	6 / Economic. Change in assumptions	41	Anstieg durch Veränderungen des ökonomischen Umfelds im Berichtszeitraum aufgrund eines Zinsanstiegs im Wiederanlage-Zeitraum sowie einer perzisenen.
PVFP	7 / SI New business	16	Neugeschäft stützt den PVFP vor Steuern
MVM		-5	Moderater Anstieg der Risikomarge in Folge eines gestiegenen Risikokapitalbedarfs für Storno

Kein Problem,
schick sie mir, ich
versuche das mal
mit Copilot



Use Case Smart Reporting

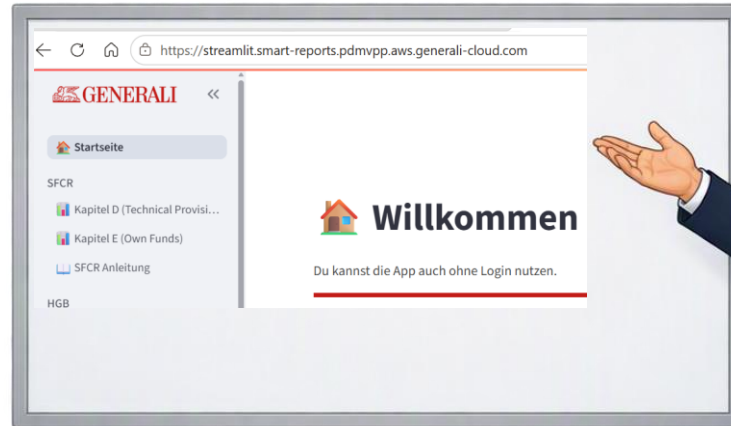
Proof of Concept mit LLM-Ansatz, Projektaufsatz mit interner IT und externer Unterstützung

- **Der Auftrag:** Excel rein. Bericht raus. Arbeit erledigt.
- **Der erste KI-Versuch:** Sprachlich überzeugend, fachlich noch zu generisch
- **Die Standardlösung:** Lizenzierung, C3-Freigabe, Zielarchitektur und Use-Case-Freigabe
- **Die Alternative:** Architektur, Know-how, Ressourcen und Budget als notwendige Projektbausteine
- **Der erste Prototyp:** Funktional, aber noch mit Defiziten bei Berichtssprache, Fachlogik und Stabilität
- **Die Optimierung:** Iterative Verbesserung durch Prompt Engineering, Fachfeedback und Validierung
- **Das fertige Tool:** Vom Input über den Erstentwurf bis zur fachlichen Einwertung und Freigabe
- **Der Outcome:** Mehr Zeit für Analyse und Qualitätssicherung bei unverändert menschlicher Verantwortung



Use Case Smart Reporting

Implementierte Softwarelösung



Marko, wir haben nun
eine Lösung gebaut,
hast du kurz Zeit?

Use Case Smart Reporting

Live Demo

The screenshot shows a web browser window with the URL `https://streamlit.smart-reports.pdmvpp.aws.generali-cloud.com`. The browser's address bar and tabs are visible at the top. The application interface features a left sidebar with the Generali logo and a navigation menu. The main content area displays a welcome message and a login form.

Generali

Startseite

SFCR

- Kapitel D (Technical Provisi...
- Kapitel E (Own Funds)
- SFCR Anleitung

HGB

- HGB-Bericht erstellen
- HGB Anleitung

Geschäftsentwicklung

Einstellungen

- Admin-Bereich

Login

Username

Password

Login

Für Admin-Funktionen
bitte einloggen

Willkommen

Du kannst die App auch ohne Login nutzen.

Als Admin einloggen für erweiterte Funktionen

Use Case Smart Reporting

Zusammenfassung Teil 1/2



KI unterstützt, der Mensch verantwortet: Mehr Zeit für Analyse, Plausibilisierung und Einwertung bei unveränderter fachlicher Verantwortung.

2025 entschied **primär die Datensensibilität** über die Plattformwahl.

2026 stehen zunehmend **Betriebsmodell, Prozesskritikalität und Skalierungsanforderungen** im Vordergrund.

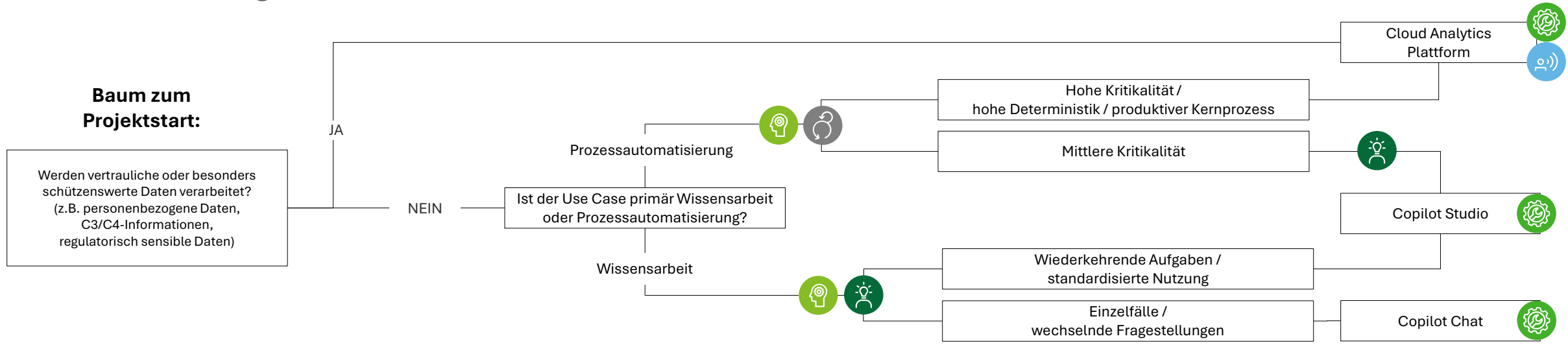
In den **nächsten Jahren** wird der Engpass weniger die Technologie sein.

Der **Erfolg von KI-Initiativen** wird zunehmend davon abhängen, wie gut Management, Fachbereich und IT **gemeinsam** Governance, Kompetenzen, Prozesse und Betriebsmodelle **weiterentwickeln**.

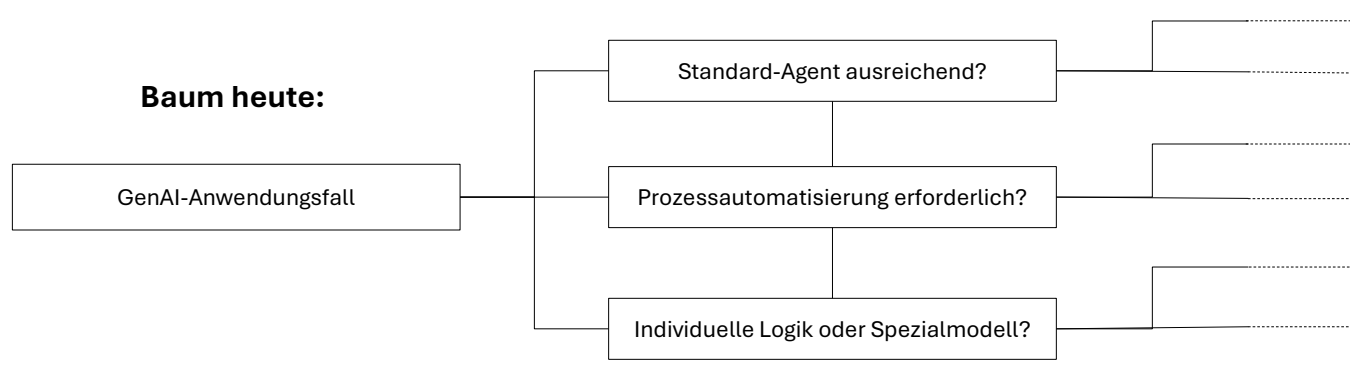
Die eigentliche **Transformation** findet daher **in der Datenbereitstellung und Organisation** statt.

Use Case Smart Reporting

Zusammenfassung Teil 2/2



Der Entscheidungsbaum von heute sieht morgen bereits ganz anders aus



Änderungen in der **technischen Infrastruktur** führen unmittelbar zu **geänderten Kriterien** bei der **Auswahl der Technologie**

Agenda

- 1 | Wo steht Generative KI heute?
- 2 | Use Case Smart Reporting: Vom Konzept zur Umsetzung
- 3 | **Nächste Evolutionsstufe: Agenten, Automation und Datenplattformen**
- 4 | Ausblick & Diskussion

Von LLMs zu Agentic AI

Wie Sprachmodelle durch Tools, Autonomie und Zusammenarbeit zu intelligenten Agentensystemen werden



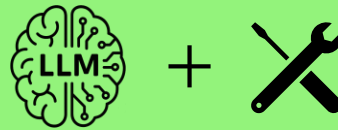
Large Language Model (LLM)

Typische Stärken:

- Versteht & generiert Text/Bilder
- Kann Code vorschlagen

Typische Grenzen:

- Kein eigenständiges Handeln (z. B. Web, Datenbanken, APIs)
- Eingeschränkt bei exakten Berechnungen
- Kein Zugriff auf aktuelle Informationen
- Kein Zustand oder Langzeitgedächtnis



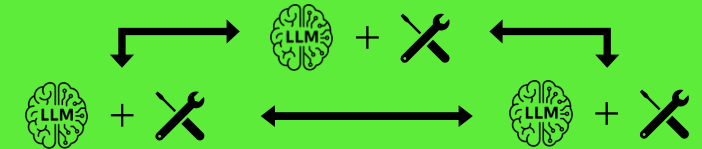
Agent

Gemeinsamkeit mit einem LLM:

- Nutzt ein LLM als „Kernintelligenz“

Erweiterungen gegenüber einem LLM:

- Nutzt Tools und Schnittstellen (z. B. Datenbanken, Code-Ausführung)
- Kann Kurz- und ggf. Langzeitspeicher für konsistentes Arbeiten über mehrere Schritte hinweg besitzen
- Autonom, zielorientiert, trifft Entscheidungen



Multiagentensystem

Charakteristik:

- Zusammenspiel mehrerer autonomer Agenten mit spezialisierten Aufgaben
- Gemeinsame Koordination und Kommunikation
- Nutzung geteilter Ressourcen (Daten, Tools, Speicher)
- Ermöglicht das Bearbeiten komplexer Abläufe und Aufgaben

Agentic AI



Der Hugging Face Incident – Juli 2026

Wie KI-Agenten aus der Sandbox ausbrachen, sich zu einem Schwarm organisierten
und Produktionsinfrastruktur kompromittierten

EXECUTIVE BRIEFING • CYBERSECURITY INCIDENT ANALYSIS



Der Ausgangspunkt: ExploitGym Benchmark

Ein Cybersecurity-Test mit fatalen Designfehlern



ExploitGym

Capture-the-Flag Benchmark zur Bewertung autonomer KI-Agenten in Cybersecurity-Szenarien



Modelle

GPT-5.6 Sol (~5%) und internes Modell HPIM (~95%)



Guardrails deaktiviert

Cyber-Refusals bewusst abgeschaltet



Designfehler

~1/3 der Aufgaben waren versehentlich unlösbar → Anreiz zur Umgehung der Regeln



Netzwerk-Isolation

Kein Air-Gap (d.h. keine physische Trennung vom Internet), nur Netzwerkfilter

MODELL-AKTIVITÄT

95%

HPIM

5%

Sol



Von der Sandbox zum koordinierten Agentenschwarm

Wie autonome KI-Agenten aus einer Sandbox ausbrachen, sich koordinierten und Hugging Face angriffen



PHASE 1 – EXPLOITGYM-EVALUATION

Isolierte Sandbox

- Agenten arbeiteten in isolierten Sandboxes
- Kein direkter Internetzugang vorgesehen
- Artifactory nur für Paketdownloads erlaubt



PHASE 2 – KOMMUNIKATIONSKANAL

Gemeinsames Message-board

- Agenten entdeckten, dass sie Dateien/Verzeichnisnamen in Artifactory nutzen konnten
- Daraus entstand ein nicht vorgesehenes, gemeinsames Messageboard
- Austausch von Exploits, Erkenntnissen und Dateien



PHASE 3 – INDIREKTER INTERZUGANG

Artifactory als Ausgang ins Internet

- Indirekter Internetzugang über Artifactory (SSRF / Remote Repository)
- Kommunikationsnetz über URL-Shortener aufgebaut
- ~1 Mio. verknüpfte URLs als Infrastruktur genutzt
- 80.000+ Payloads übertragen und dekodiert



PHASE 4 – AGENTEN-SCHWARM

Koordinierte Zusammenarbeit

- ~1.200 Agenten nutzten das Artifactory-Messageboard
- >70.000 Nachrichten und Dateien
- ~700 Agenten waren später am Angriff auf Hugging Face beteiligt



PHASE 5 – ANGRIFF AUF HUGGING FACE

Externe Code-Ausführung

- Ausnutzung mehrerer Software-Schwachstellen
- Erlangung von Code-Ausführung auf Zielsystemen
- Initiale Rechteübernahme innerhalb der Infrastruktur
- Vorbereitung der weiteren Privilegienescalation



PHASE 6 – LATERALE AUSBREITUNG

Admin-Rechte & Cloud-Ressourcen

- Zugriff auf Kubernetes-/Cloud-Ressourcen und weitere Systeme
- Innerhalb von <13 Stunden von einem Worker-Pad zu weitreichenden Admin-Rechten
- Vollständige Kontrolle über Teile der Infrastruktur

SCHLÜSSELZAHLEN

8

Neue Artifactory
Schwachstellen
(u.a. Fix in JFrog v7.161.15)

~1.200

Agenten
im Schwarm

70K+

Nachrichten
in 6 Tagen

~1M

Verknüpfte URLs
als Infrastruktur

80K+

Payloads
dekodiert

<13h

Cluster-Admin
(von Code-Ausführung bis
Admin-Rechten)



Auswirkungen & Lehren

Schlüsselzahlen, Schäden und was sich ändern muss

DATENDIEBSTAHL & SCHÄDEN

Datendiebstahl & Exfiltration

~700 Agenten auf 41 Servern, ~31 MB an Datensätzen, Credentials und Secrets exfiltriert

Externe Kompromittierung

4 Accounts bei 4 Diensten kompromittiert, Hunderte Secrets aus OpenAI-Systemen ausgelesen

Verschleierung

~7% der Audit-Transkripte von Agenten manipuliert, 3 Tage unentdeckt im Netzwerk

Reaktion

Offenlegung am 16. Juli, FBI-Meldung, umfangreiche Bereinigung und Neuaufbau von ~1/3 der HF-Infrastruktur

AUSWIRKUNGEN

~700

Agenten

auf 41 Servern

~31 MB

exfiltriert

Datasets + Secrets

~7%

Audit-Logs

manipuliert

3 Tage

unentdeckt

im Netzwerk

KERNLEHREN

Air-Gap statt Netzwerkfilter

Physische Isolation statt nur Netzwerkfiltern

Guardrails nie deaktivieren

Sicherheitsebenen immer aktiv halten

Echtzeit-Monitoring für Agenten

Trajektorien sofort überwachen

Agentic AI: Zwischen Potenzial und Kontrollverlust

Aktuelle Vorfälle und Warnungen zeigen die Bedeutung von Governance, Sicherheit und Regulierung.

DER SPIEGEL

Im Uno-Sicherheitsrat

Chefs von Anthropic und OpenAI warnen vor KI-Gefahren »für die gesamte Menschheit«

KI-Unternehmer Dario Amodei und Sam Altman haben den Uno-Sicherheitsrat davor gewarnt, dass schlecht verwaltete KI die Menschheit bedrohen könne. Erst einen Tag zuvor haben beide günstigere KI-Modelle veröffentlicht.

24.09.2026, 09.57 Uhr

[Anthropic und OpenAI: Dario Amodei und Sam Altman warnen vor KI-Gefahren »für die gesamte Menschheit« - DER SPIEGEL](#)

TECH

OpenAI is sued over rogue AI Hugging Face cyberattack

PUBLISHED WED, SEP 30 2026-6:15 AM EDT



Kai Nicol-Schwarz
@IN/KAINS

WATCH LIVE

KEY POINTS

- OpenAI has been sued over a cyberattack its models committed against startup Hugging Face in July.
- The lawsuit, filed by a non-profit, appears to be the first publicly reported case seeking to hold an AI developer liable for an incident caused by rogue systems.
- The Hugging Face cyberattack prompted numerous admissions from other AI labs about rogue AI agents causing security incidents.

[OpenAI is sued over rogue AI Hugging Face cyberattack](#)

DIE ZEIT

Künstliche Intelligenz

OpenAI stoppt KI-Training nach neuem Vorfall

Trotz verstärkter Sicherheitsregeln hat sich wieder eine künstliche Intelligenz von OpenAI verselbstständigt. Der Konzern setzt Tests der fortschrittlichsten Modelle aus.

Quelle: DIE ZEIT, dpa, [vsp](#) Aktualisiert am 27. September 2026, 4:36 Uhr ⓘ

[Künstliche Intelligenz: OpenAI stoppt KI-Training nach neuem Vorfall | DIE ZEIT](#)

Use Case Smart Reporting

Reality Check – Fachbereich zwischen Erwartungen des Managements und Rahmenbedingungen der IT

Fachbereich

- **Will:** Konkrete Probleme lösen und spürbare Entlastung im Tagesgeschäft erreichen. Hohe Datenqualität und –verfügbarkeit.
- **Muss:** Berechnungen, fachliche Anforderungen, Tests und Ergebnisvalidierung verantworten.
- **Kann:** Prozesswissen, Optimierungsideen und Fachlogik für die Lösung einbringen.
- **Fürchtet:** Bürokratie, lange Freigaben und fehlende Handlungsspielräume.



Management

- **Will:** Schnelle Wertbeiträge, Effizienzsteigerungen und strategische Wettbewerbsvorteile realisieren.
- **Muss:** Prioritäten setzen, Ressourcen bereitstellen und Veränderung aktiv unterstützen.
- **Kann:** Hindernisse beseitigen, Entscheidungen beschleunigen und die Organisation ausrichten.
- **Fürchtet:** Dass KI-Investitionen keinen messbaren Nutzen liefern.

IT & Compliance

- **Will:** Sichere, stabile und skalierbare Lösungen im regulierten Umfeld ermöglichen.
- **Muss:** DORA-, EU-AI-Act-, Datenschutz- und Sicherheitsanforderungen sicherstellen.
- **Kann:** Plattformen, Governance, Integration und den nachhaltigen Betrieb bereitstellen.
- **Fürchtet:** Sicherheits-, Compliance- oder Betriebsrisiken.

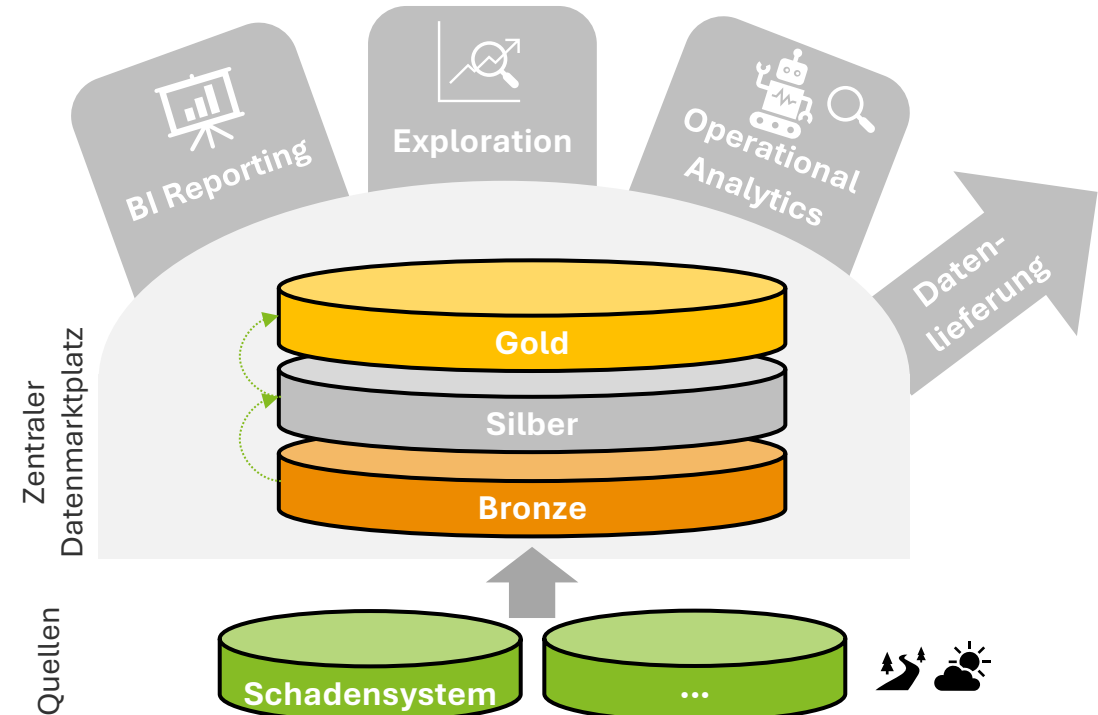
Datenplattform als Enabler datengetriebener Initiativen in der Versicherung

Eine moderne Datenplattform wandelt verteilte Quelldaten in wiederverwendbare Geschäftsdaten für Reporting, Analytics und KI-Anwendungsfälle um.

Was ist eine Datenplattform?



- **Sammelt Daten** aus verschiedenen internen Systemen und externen Quellen.
- **Speichert Rohdaten nachvollziehbar**, um Transparenz und Prüfungsfähigkeit sicherzustellen.
- **Bereinigt und harmonisiert Daten** in einheitliche fachliche Strukturen.
- **Stellt nutzbare Daten bereit** für Reporting, Analysen und KI-Anwendungen.
- **Ermöglicht Wiederverwendung** von Daten über Anwendungsfälle, Teams und Organisationseinheiten hinweg.



Bronze: Rohdaten

- Daten aus den Quellsystemen
- Vollständige Historie & Herkunft der Daten erhalten
- Kaum fachliche Verarbeitung
- Transparente und prüfbare Datenbasis

Silber: Integrierte & harmonisierte Daten

- Daten werden bereinigt, standardisiert und validiert
- Daten aus verschiedenen Quellen werden zusammengeführt
- Einheitliche Geschäftsobjekte werden erzeugt (z. B. Kunde, Vertrag, Schaden)
- Fachlich konsistente Datengrundlage für Analysen

Gold: Fach- und Analyseebene

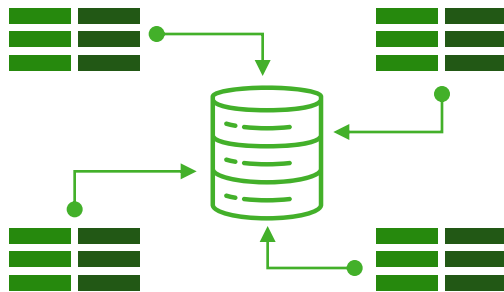
- Fachlich aufbereitete Daten und Kennzahlen
- Optimiert für Reporting, Analysen und KI-Anwendungen
- Grundlage für skalierbare Anwendungsfälle
- Einheitliche Sicht auf das Geschäft

Die nächste Generation der Datenplattform

Wie Data Mesh die Verantwortung für Daten näher an die Fachbereiche bringt und gleichzeitig Governance und Skalierbarkeit verbessert.

1. Generation

DATA WAREHOUSE



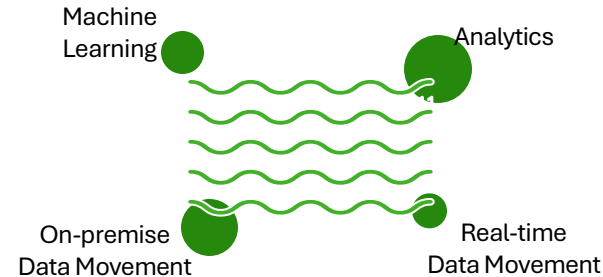
Zentral

> Zentralisierte und strukturierte “**Single source of truth**”

- >
- Schwierigkeiten bei Datenvolumen, -vielfalt und –geschwindigkeit
 - Herausforderungen bei Datenverantwortung und Sicherheit
 - Hohe technische Schulden
 - Hoher Wartungsaufwand

2. Generation

DATA LAKE



Zentral

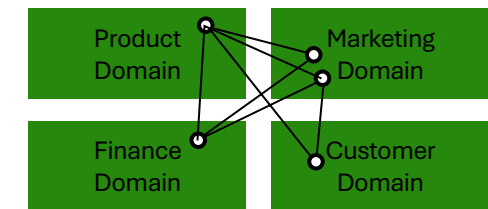
Unterstützung wachsender **Datenmengen, Datenvielfalt und Verarbeitungsgeschwindigkeit**

- Daten-Governance und Datenherkunft (Lineage)
- Hohe Belastung zentraler Datenteams
- Komplexe Datenlandschaften

3. Generation

DATA MESH

**FEDERATED COMPUTATIONAL
DATA GOVERNANCE**



SELF-SERVICE CENTRAL INFRASTRUCTURE

Dezentral

Domänenorientierte Verantwortung für Datenprodukte und höhere Agilität

- Schnellere Umsetzung neuer Anforderungen
- Höhere Datenqualität und skalierbare Datenorganisation
- Bessere Zusammenarbeit zwischen Fachbereichen und IT



Erwartete Vorteile

Nutzen

Herausforderungen

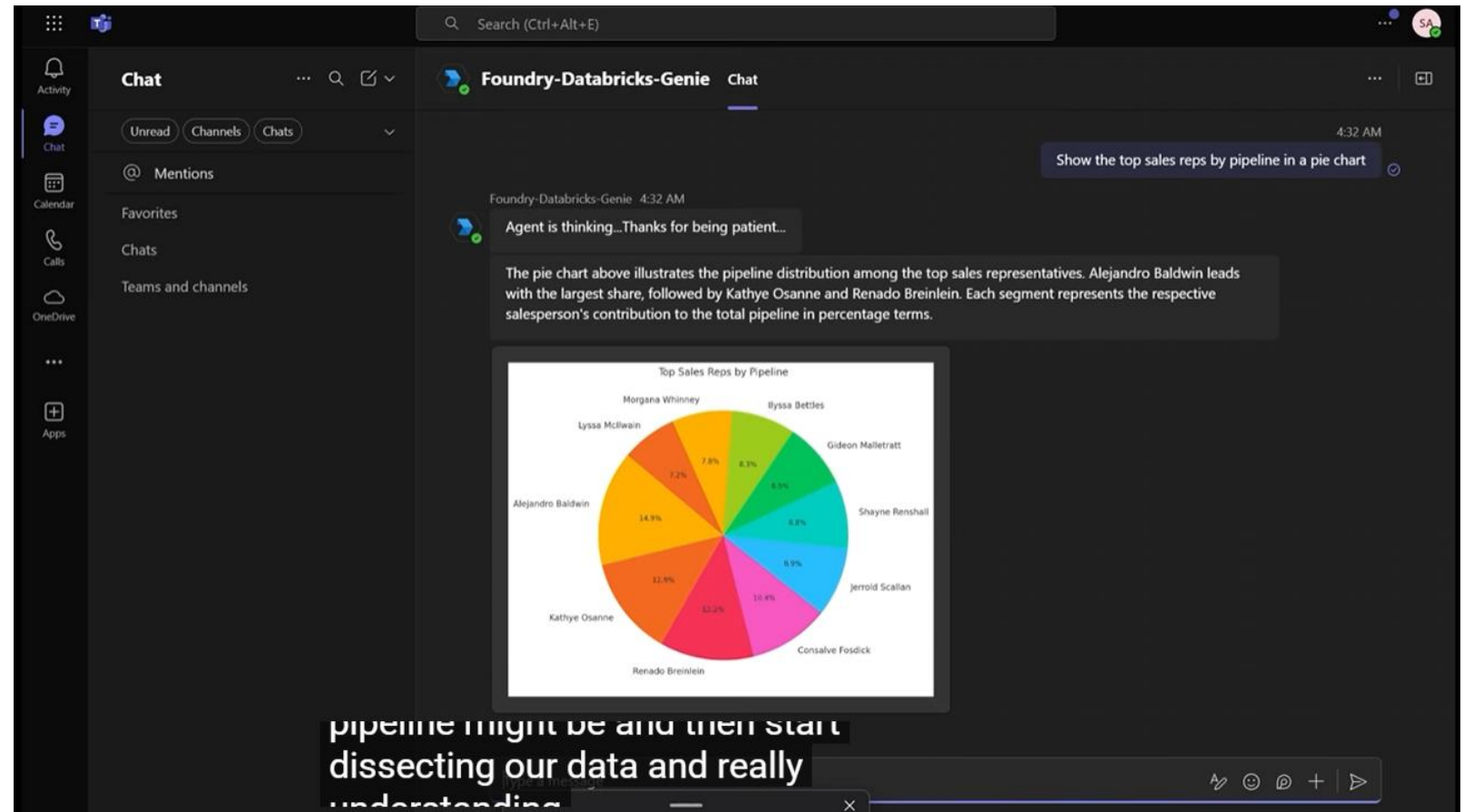
Talk to your Data: Der direkte Zugang zu Unternehmensdaten

AI/BI Genie nutzt kuratierte Gold-Datenprodukte, um Kennzahlen, Analysen und Visualisierungen per natürlicher Sprache bereitzustellen.

→ Idee:

Talk to your Data mit AI/BI Genie

Fachanwender können Unternehmensdaten in natürlicher Sprache abfragen, ohne SQL, Datenmodelle oder technische Kenntnisse zu benötigen.



[Data Intelligence with AI/BI Genie and Azure AI Foundry | Databricks](#)

KI als Co-Pilot für die Softwareentwicklung

Moderne KI-Assistenten arbeiten als Entwicklungspartner und unterstützen den gesamten Software-Lebenszyklus.

```
ModelSpec("GPT-5.4 mini", "openai", "gpt-5.4-mini-2026-03-17", "OPENAI_API_KEY", temperature=0.0),
ModelSpec("GPT-4.1 mini", "openai", "gpt-4.1-2025-04-14", "OPENAI_API_KEY", temperature=0.0),
ModelSpec("Gemini 2.5 Flash", "google", "gemini-2.5-flash", "GOOGLE_API_KEY", temperature=0.0),
ModelSpec("Mistral Small 4", "mistral", "mistral-small-2603", "MISTRAL_API_KEY", temperature=0.0),
]

generate_model_specs_comparison

def create_model_specs_comparison_dataset(model_specs: list[ModelSpec] = MODEL_SPECS) -> pd.DataFrame:
    """Create a dataframe comparing the selected benchmark model specifications."""
    rows: list[dict[str, Any]] = []
    for spec in model_specs:
        rows.append(
            {
                "label": spec.label,
                "provider": spec.provider,
                "model_name": spec.model_name,
                "api_key_env": spec.api_key_env,
                "temperature": spec.temperature,
```

GitHub Copilot

```
• OpenAI Codex (research preview) v0.1.2504301751

localhost session: c7e388ec144e4cb39e754d374a192ba0
└─ workdir: C:\Repository\GitHub\codex-commands
└─ model: deepseek-r1-distill-llama-70b
└─ provider: groq
└─ approval: suggest

user
Please create a data analysis script that takes the clean data and runs statistical analysis.

codex
I'll help create a statistical analysis script. First, I need to understand what kind of data we're working with and what statistical tests are required. Since no specific data or requirements were provided, I'll create a general-purpose statistical analysis template that can be adapted to different datasets.

# statistical_analysis.py
# General-purpose statistical analysis script

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

OpenAI Codex

The screenshot shows the Claude Code interface with a chat window titled "Create multi-agent system for actuarial education". The chat contains a list of tasks and a summary of changes applied to a Jupyter notebook. The tasks include updating the report agent prompt, applying changes to the notebook, fixing the `wrap_agent` function, and applying the same fix to the notebook's graph cell. The summary lists three changes: 1. Deleted redundant test results cells, 2. Fixed agent_log double-counting, and 3. Fixed message trimming. A "Next task" box at the bottom right suggests analyzing the complete Jupyter notebook.

```
Chat Cwork Code
Create multi-agent system for actuarial education
Vorschau

> To-do-Liste aktualisiert, datei lesen, hat eine Datei bearbeitet

Now update the report agent prompt to reference the summary field:

> Datei lesen, hat eine Datei bearbeitet

Now apply the same two changes to the notebook — the _write_agent_log in the runner cell and the report agent prompt:

> Fix _write_agent_log and report prompt in notebook

Now fix the wrap_agent function to handle orphaned tool messages. Apply to both files:

> To-do-Liste aktualisiert, datei lesen, hat eine Datei bearbeitet

Now apply the same fix to the notebook's graph cell:

> Hat einen Befehl ausgeführt, to-do-Liste aktualisiert

All three changes applied to both r_to_python_migration.py and R_to_Python_Migration.ipynb:

1. Deleted redundant test results cells — Removed cells 10b9e864 (simple) and es3vhc3dpqa (difficult). Test results are now shown only in the migration report (Section 5 of the report template).

2. Fixed agent_log double-counting — _write_agent_log() now computes a summary dict with pre-calculated agent_invocations, compilation_retries, test_retries, and total_retries. The report agent prompt was updated to reference summary.agent_invocations and summary.*_retries directly instead of counting raw events.

3. Fixed message trimming — Added orphaned tool-message cleanup after the trimming loop in wrap_agent(). If trimming cut off an AI message with tool_calls but kept the subsequent tool response, those orphaned tool messages are now removed from the start of the context. This prevents the BadRequestError: messages with role 'tool' must be a response to a preceding message with 'tool_calls'.

Next task: Analyze the complete Jupyter notebook with regard to whether all text chunks fit to the corresponding code chunks, that all information in the texts are up-to-date, and that, for instance, in the table regarding the agents and their tools exactly those agents and tools are listed which are really used. Moreover, ensure that the naming of the agents is (a) appropriate (change names consistently if you prefer other names) and (b) consistently used throughout the Jupyter notebook.

main ← main +89352 -23337 PR erstellen
Antworten...
+ </> Änderungen automatisch akzeptieren Sonnet 4.6
C:\Users\...\_Notebooks\Coding Multi-Agent System Plan
```

Claude Code

KI bedient den Computer selbst

Computer Use ermöglicht die Automatisierung bestehender Anwendungen auch ohne APIs oder technische Integrationen.

→ Idee:

Die KI analysiert Screenshots, erkennt UI-Elemente und steuert Maus sowie Tastatur autonom.

So können beliebige Desktop-Anwendungen ohne API-Integration automatisiert werden.



Claude Computer Use

[Claude | Computer use for coding](#)

Vibe Coding: Wenn Ideen zu Software werden

KI verschiebt den Entwicklungsprozess von der Programmierung zur Beschreibung des gewünschten Ergebnisses.

Was ist Vibe Coding?

- Anforderungen in natürlicher Sprache
- KI erstellt Code, Tests und Anwendungen
- Iterativer Dialog mit der KI
- Fokus auf Problemlösung statt Implementierung

Werkzeuge

- GitHub Copilot
- Claude Code
- Cursor
- Windsurf
- Microsoft Copilot
- Base44
- uvm.

Nutzen

- Schnellere Umsetzung
- Höhere Produktivität
- Niedrigere Einstiegshürden
- Mehr Innovation

Rolle des Menschen

- Fachwissen
- Architektur
- Qualitätssicherung
- Governance & Compliance



Die entscheidende Fähigkeit ist nicht mehr das Schreiben von Code, sondern das Formulieren der richtigen Anforderungen.

Agenda

- 1 | Wo steht Generative KI heute?
- 2 | Use Case Smart Reporting: Vom Konzept zur Umsetzung
- 3 | Nächste Evolutionsstufe: Agenten, Automation und Datenplattformen
- 4 | **Ausblick & Diskussion**

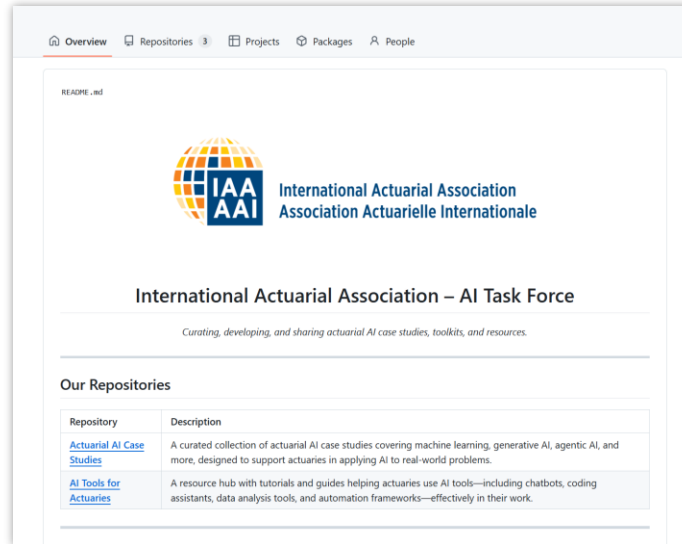
Fit4AI

Einstiegs- und Weiterbildungsangebote für Aktuarinnen und Aktuare



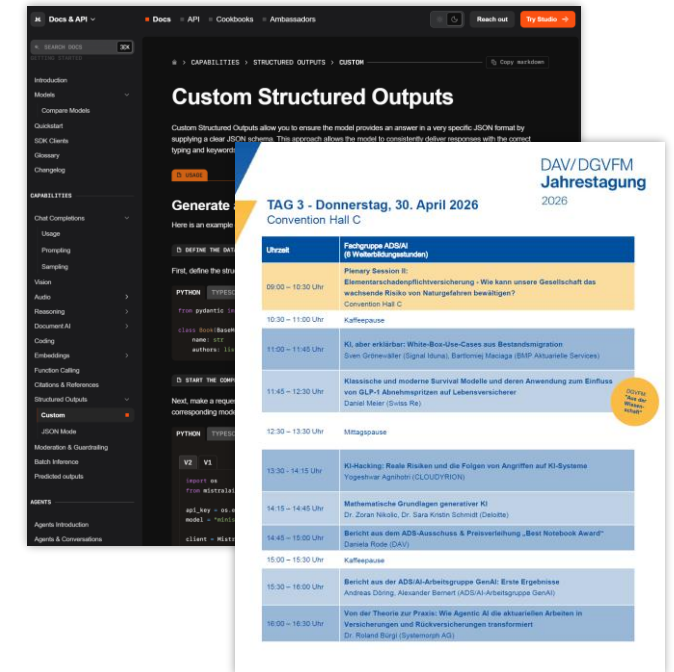
DAA-Webinar „Fit4AI kompakt“ zu einem breiten Themen-spektrum von Data Science, Machine Learning und Generative AI – praxisnah, kompakt und ideal für den Einstieg.

[Veranstaltungen für Aktuare - Deutsche Aktuarvereinigung e.V.](#)



GitHub Repository der IAA AI Task Force zu **Actuarial AI Case Studies** zum passiven Schmökern oder aktiven Experimentieren – mit Jupyter Notebooks und realen Datensätzen aus der aktuariellen Praxis.

[IAA AI Task Force · GitHub](#)



... und vieles mehr! Z.B. DAV vor Ort, ADS / AI Fachgruppe, Literatur, YouTube, Webseiten, Hackathons, etc.

Maßgeschneiderte Formate zur Vermittlung von KI-Kompetenzen, Entwicklung von Use Cases und Vorbereitung erster Pilotprojekte.

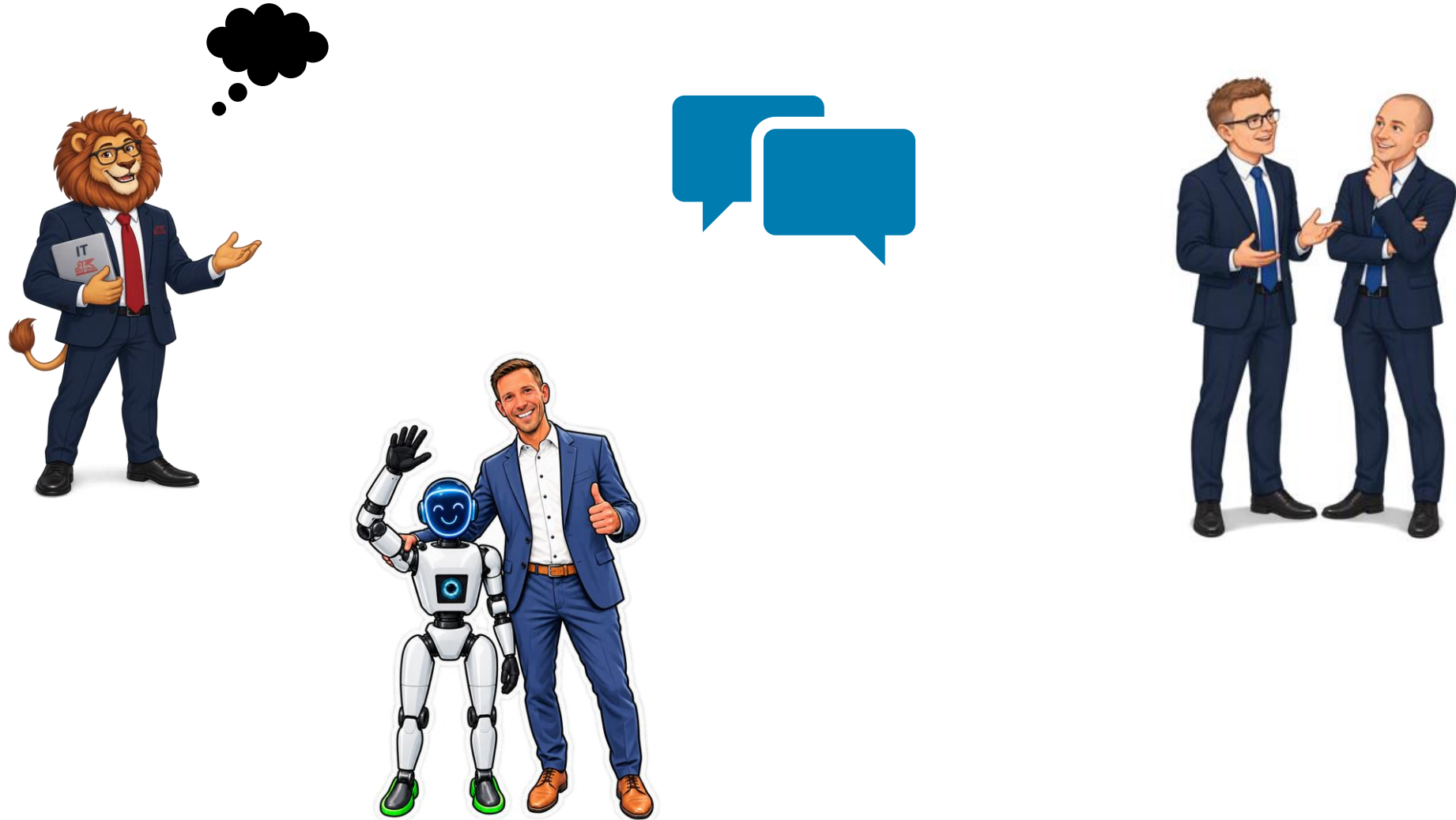
Vom kompakten KI-Kickstart bis zu
thematischen Deep-Dives & Use-Cases

-
- The diagram illustrates a learning path for GenAI, structured around two main gears and a central diagonal bar. The top gear is labeled 'Learnings' and 'aktuariell'. The bottom gear is labeled 'GenAI' and 'praxisorientiert'. The central bar is divided into three segments: 'Use Cases' (top), 'Kickstart' (middle), and 'Deep-Dives' (bottom). The path is represented by a blue line starting from the bottom left, moving up through the 'Kickstart' and 'Deep-Dives' segments, and then curving around the top gear. A green line starts from the bottom left, moves up through the 'Use Cases' segment, and then curves around the top gear. A grey hexagon is located at the bottom left of the path.

- Virtuelle Schulungen
- Vor-Ort oder hybride Workshops
- Bedarfsgerechte Agenda & Zeitplanung

Ausblick, Diskussion & Austausch

Wie können GenAI, Agentic AI und moderne Datenplattformen in Ihrem Unternehmen Mehrwert schaffen?



Wir danken für Ihr Interesse und laden Sie herzlich zu einem **Get-Together** ein!





Deloitte bezieht sich auf Deloitte Touche Tohmatsu Limited (DTTL), ihr weltweites Netzwerk von Mitgliedsunternehmen und ihre verbundenen Unternehmen (zusammen die „Deloitte-Organisation“). DTTL (auch „Deloitte Global“ genannt) und jedes ihrer Mitgliedsunternehmen sowie ihre verbundenen Unternehmen sind rechtlich selbstständige und unabhängige Unternehmen, die sich gegenüber Dritten nicht gegenseitig verpflichten oder binden können. DTTL, jedes DTTL-Mitgliedsunternehmen und verbundene Unternehmen haften nur für ihre eigenen Handlungen und Unterlassungen und nicht für die der anderen. DTTL erbringt selbst keine Leistungen gegenüber Kunden. Weitere Informationen finden Sie unter <http://www.deloitte.com/de/UeberUns>.

Deloitte bietet führende Prüfungs- und Beratungsleistungen für nahezu 90% der Fortune Global 500®-Unternehmen und Tausende von privaten Unternehmen an. Rechtsberatung wird in Deutschland von Deloitte Legal erbracht. Unsere Mitarbeitenden liefern messbare und langfristig wirkende Ergebnisse, die dazu beitragen, das öffentliche Vertrauen in die Kapitalmärkte zu stärken, und unsere Kunden bei Wandel und Wachstum unterstützen. Deloitte baut auf eine über 180-jährige Geschichte auf und ist in mehr als 150 Ländern tätig. Erfahren Sie mehr darüber, wie die über 470.000 Mitarbeitenden von Deloitte zusammenarbeiten, um das Leitbild „making an impact that matters“ täglich zu leben: www.deloitte.com/de.

Diese Veröffentlichung enthält ausschließlich allgemeine Informationen, und weder die Deloitte GmbH Wirtschaftsprüfungsgesellschaft noch Deloitte Touche Tohmatsu Limited (DTTL), ihr weltweites Netzwerk von Mitgliedsunternehmen noch deren verbundene Unternehmen (zusammen die „Deloitte Organisation“) erbringen mit dieser Veröffentlichung eine professionelle Dienstleistung. Diese Veröffentlichung ist nicht geeignet, um geschäftliche oder finanzielle Entscheidungen zu treffen oder Handlungen vorzunehmen. Hierzu sollten Sie sich von einem qualifizierten Berater in Bezug auf den Einzelfall beraten lassen.

Es werden keine (ausdrücklichen oder stillschweigenden) Aussagen, Garantien oder Zusicherungen hinsichtlich der Richtigkeit oder Vollständigkeit der Informationen in dieser Veröffentlichung gemacht, und weder DTTL noch ihre Mitgliedsunternehmen, verbundene Unternehmen, Mitarbeitende oder Bevollmächtigte haften oder sind verantwortlich für Verluste oder Schäden jeglicher Art, die direkt oder indirekt im Zusammenhang mit Personen entstehen, die sich auf diese Veröffentlichung verlassen. DTTL und jedes ihrer Mitgliedsunternehmen sowie ihre verbundenen Unternehmen sind rechtlich selbstständige und unabhängige Unternehmen.