EnBW
Digital Office
Zukunft. Digital. Machen.

EnBW

AI Agents – Grundlagen

Franziska Zerwas

KI-Agenten sind das aktuelle Hype-Thema

EnBW Digital Office

> AI agents will transform the way we interact with technology, making it more natural and intuitive.

Fei-Fei Li
Professor of Computer Science
at Stanford University

> AI agents will become an integral part of our daily lives, helping us with everything from scheduling appointments to managing our finances.

Andrew Ng
Co-Founder
Google Brain, Coursera

> The IT department of every company is going to be the HR department of AI agents in the future.

Jensen Huang
Nividia CEO

Aber...

EnBW
Digital Office

Agenda

Intro 

Fähigkeiten & Aufbau

Business Implikationen

 Aktuelle Trends & Abschluss

Q&A 

Fähigkeiten & Aufbau

Es gibt zahlreiche Definitionen zu KI-Agenten, hier ein paar Beispiele:

EnBW Digital Office

A generative AI agent is an autonomous system that leverages large language models and foundation models to ***independently execute complex tasks and workflows*** *in a digital / physical environment. It perceives its surroundings,* ***reasons, plans, and acts*** *over time to achieve its goals and influence future outcomes.* - AppliedAI (2024)

AI agents are autonomous or semiautonomous software entities that use AI techniques to ***perceive, make decisions, take actions and achieve goals*** *in their digital or physical environments.* - Gartner (2024)

Agents, on the other hand, are systems were LLMs dynamically ***direct their own processes and tool usage****, maintaining control over how they accomplish tasks.* - Anthropic (n.d.)

Unser Vorschlag einer Definition

*KI-Agenten sind **autonome** Systeme, die mithilfe von KI-Modellen wie LLMs auf ein **definiertes Ziel** hin arbeiten.*

*Dabei **zerlegen sie komplexe Probleme in Teilschritte**, **nutzen** verfügbare **Daten und Tools** und passen ihr Verhalten im Hinblick auf die Zielerreichung durch **Reflexion und Feedback** kontinuierlich an.*

***Mehrere KI-Agenten** ermöglichen eine arbeitsteilige, **koordinierte** Problemlösung.*

KI generiert

Autonome KI-Agenten arbeiten mithilfe folgender Fähigkeiten auf ein definiertes Ziel hin

EnBW Digital Office

Planung

- Agenten planen Zielerreichung, indem sie **komplexe Aufgaben in kleinere Aktionen** zerlegen
- **Erfordert Verständnis** der Absicht, der verfügbaren Werkzeuge und der möglichen Ergebnisse von Aktionen[1]

Ausführung mit Tools

- "**Read**"-Aktionen (Informations-gewinnung) und "**Write**"-Aktionen (Umgebungsveränderung)
- **Robuste Ausführung** erfordert Behandlung von Fehlern, Umgang mit Tool-Beschränkungen und Verständnis komplexer Rückmeldungen
- Integration von **Memory-Systemen** zur Speicherung von Informationen über Kontext-Fenster hinaus

Evaluation

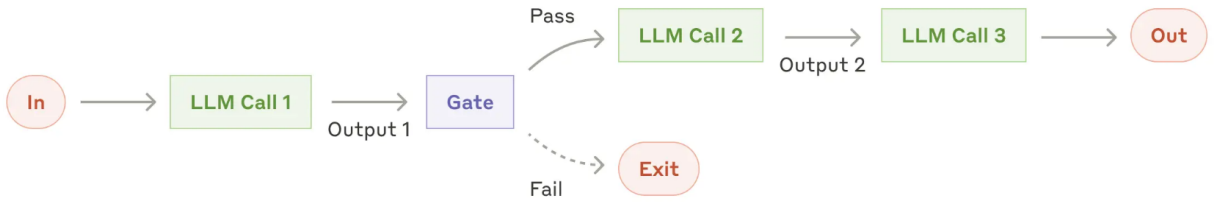
- Agenten müssen über Aktionen **nachdenken**, Ergebnisse **bewerten** und Pläne bei Bedarf **anpassen**[1]
- Erfolgt durch Selbstkritik-mechanismen, separate Bewertungskomponenten oder menschliches Feedback

Multi-Agent Kollaboration: Arbeitsteilige Problemlösung, erfordert effektive Koordination

Andrew Ng

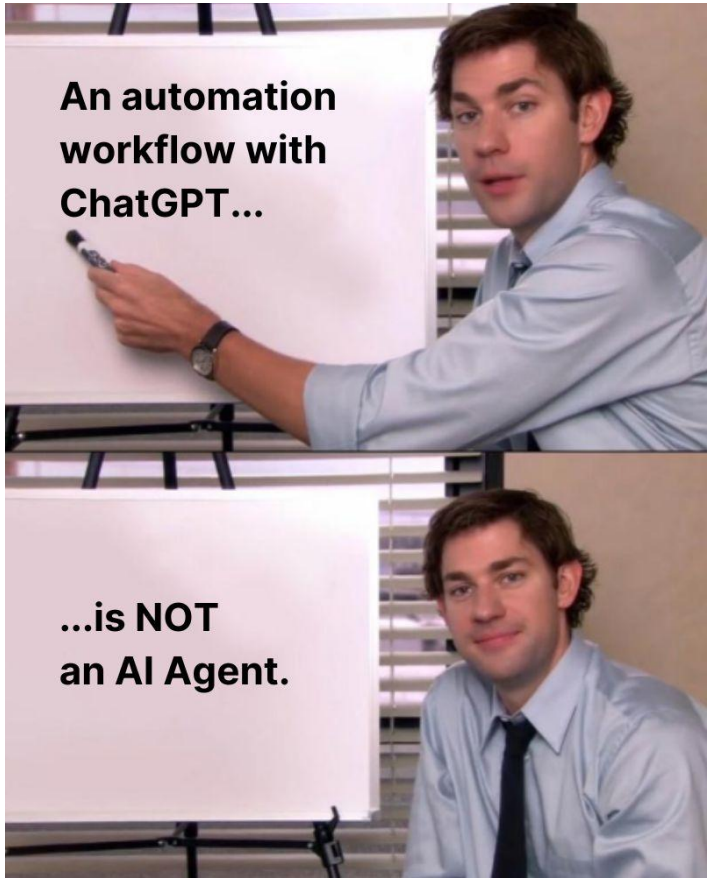
[1]Beliebte Reasoning Frameworks: ReAct, Chain-of-Thought

LLM-gekoppelte Workflows für strukturierte Abläufe



Source: Anthropic

→ Kopplung einzelner LLM-Funktionen in einem festen Ablauf ohne Zielverfolgung



Use Case Beispiel: Content Prompt Chain Workflow

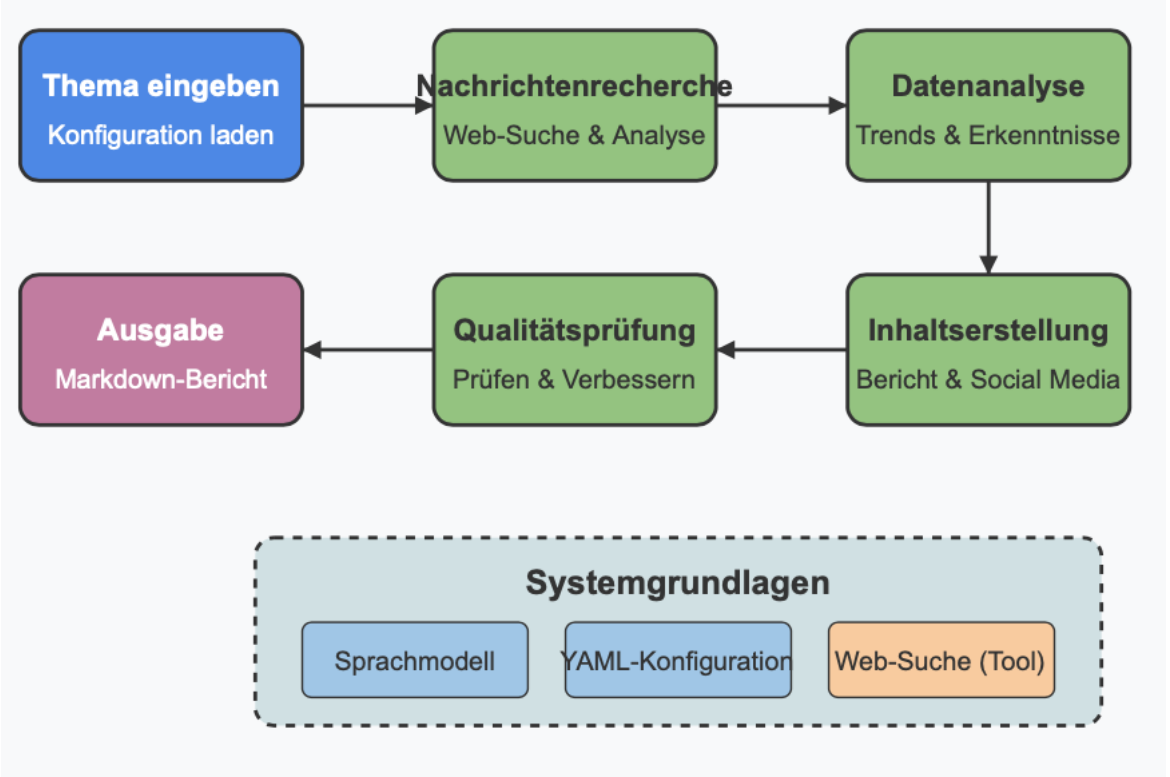
EnBW Digital Office

Problem:

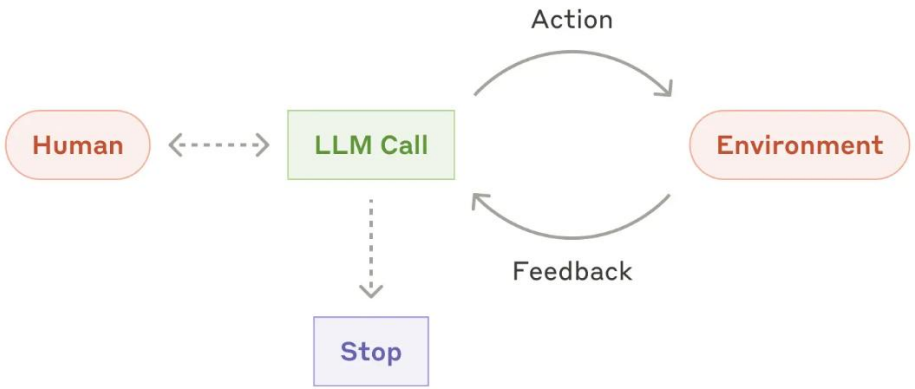
- Zeitaufwändige, manuelle Recherche und Analyse von Trends
- Herausforderung, konsistente und qualitativ hochwertige Inhalte zu erstellen

Lösung:

- LLM-gekoppelter Workflow zur automatisierten Erstellung von Content
- Spezialisierte Rollen für verschiedene Phasen des Recherche- und Erstellungsprozesses



Autonome KI-Agenten für eigenständige Entscheidungsfindung und Anpassung



Source: Anthropic

→ (Multi-)Agenten System mit eigenständiger Zielverfolgung (Planung, Toolnutzung und Evaluation)

Grundlegende Design-Optionen für KI-Agenten

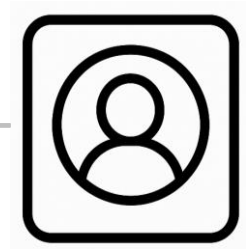
EnBW Digital Office

Task-orientiert:
Unterstützung von Nutzern bei täglichen Aufgaben

Innovations-orientiert:
Autonome Erkundung bspw. in wissenschaftl. Fragen

Lebenszyklus-orientiert:
Kontinuierliches Lernen und Fähigkeitsentwicklung

Single Agent



Multi-Agent

Kooperative Interaktion:
Agenten arbeiten zusammen, entweder geordnet oder ungeordnet, um gemeinsame Ziele zu erreichen

Adversariale Interaktion:
Wettbewerbsdynamiken zur Verbesserung der individuellen Leistung

Human-Agent Interaction

Instructor-Executor:
Menschen geben Anweisungen; Agenten führen Tasks aus

Gleichberechtigte Partnerschaft: Agenten interagieren empathisch und kooperieren mit Menschen

Source: AppliedAI (2024)

Wir stehen noch ganz am Anfang

EnBW Digital Office

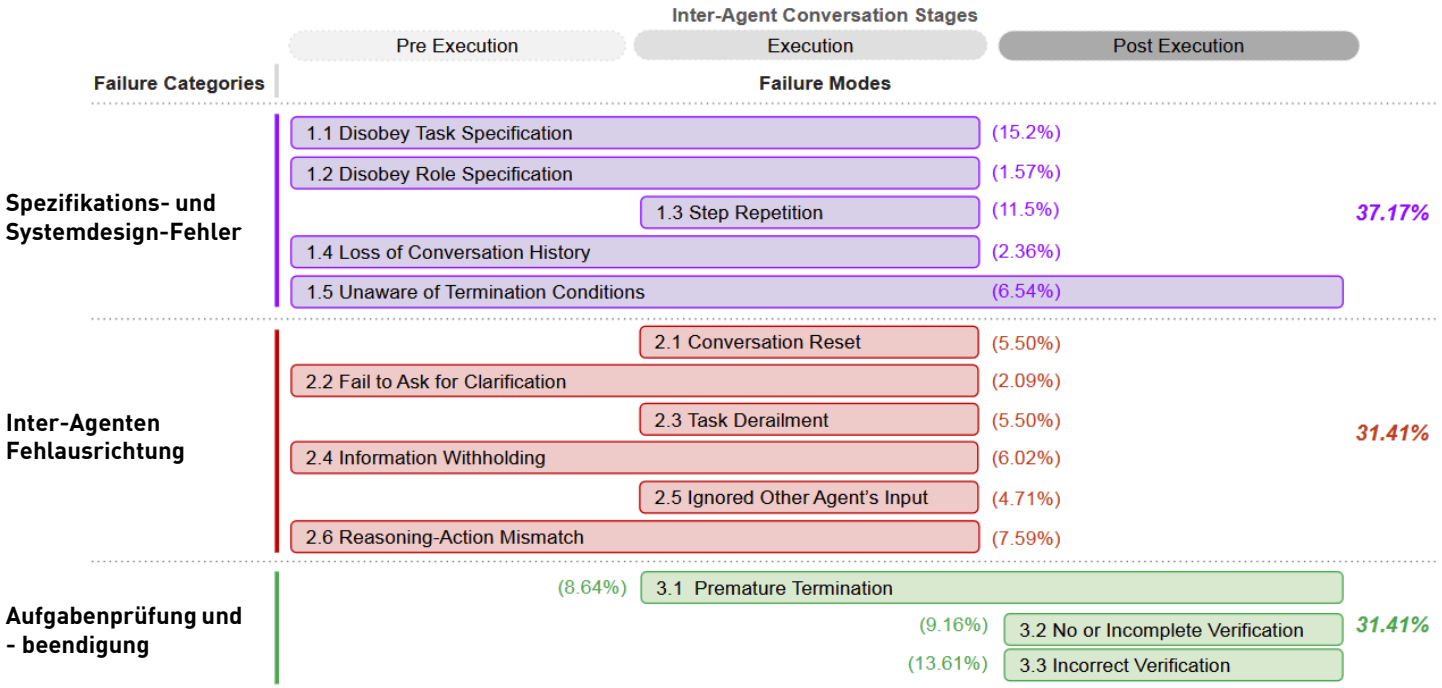
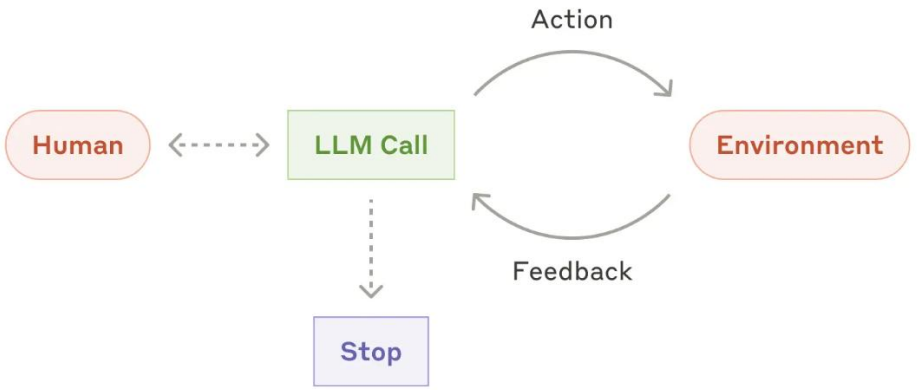


Figure 2. A **Taxonomy of MAS Failure Modes**. The inter-agent conversation stages indicate when a failure can occur in the end-to-end MAS system. If a failure mode spans multiple stages, it means the issue involves or can occur at different stages. Percentages represent how frequently each failure mode and category appeared in our analysis of 151 traces. Detailed definition and example of each failure mode is available in Appendix A.

Cemri et al. (2025)

Autonome KI-Agenten für eigenständige Entscheidungsfindung und Anpassung

EnBW Digital Office



Source: Anthropic

→ (Multi-)Agenten System mit eigenständiger Zielverfolgung (Planung, Toolnutzung und Evaluation)

Hypothese:

Die meisten Probleme im Unternehmenskontext sind aktuell sinnvoll adressiert mit (flexiblen) agentischen Workflows statt autonomen (Multi-)Agenten Strukturen

Gründe:

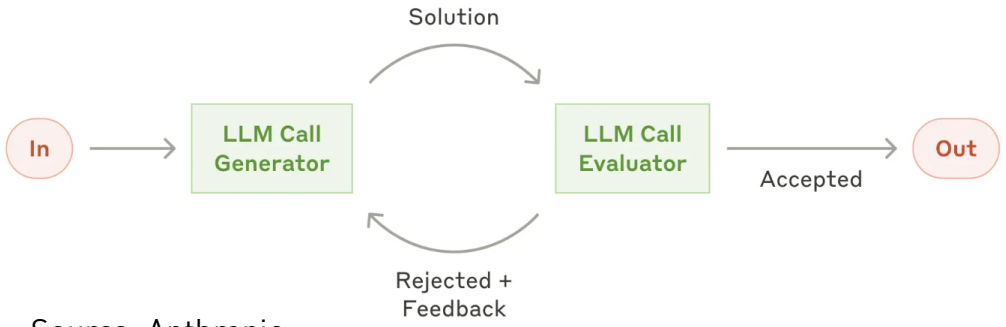
- Gewisse Kontrolle gefordert
- Meist etablierte Prozesse vorhanden

Ausnahmen:

- Freie Aufgaben mit ungewissem Outcome, z.B. Ideation
- Überall dort, wo zunächst neue Prozesse etabliert werden müssten

Agenten-ähnliche Workflows für Zusammenspiel in flexiblem, aber strukturiertem Prozess

EnBW Digital Office

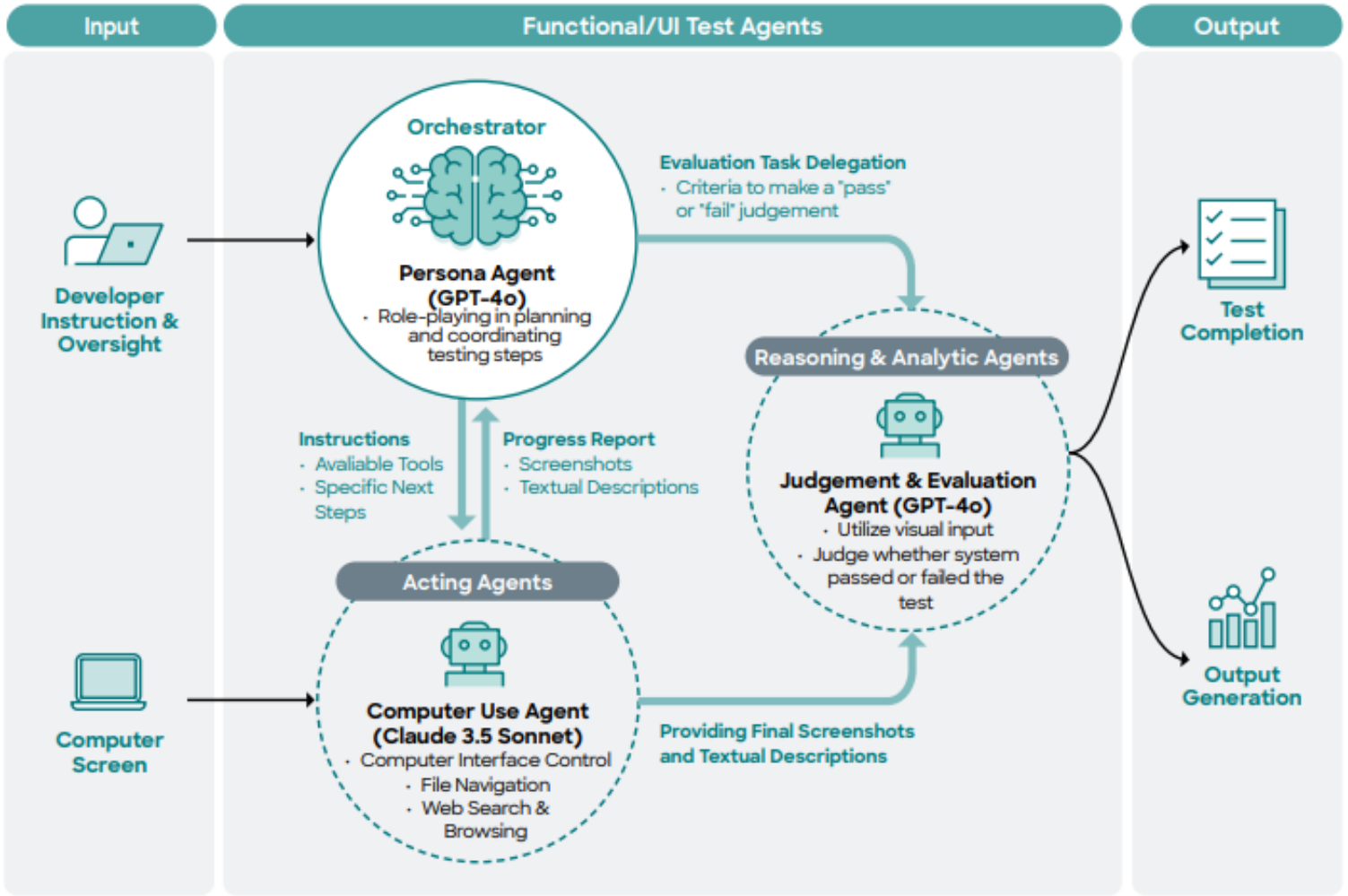


Source: Anthropic

- Ermöglichen es LLMs, den Ablauf und Tool-Einsatz eigenständig zu entscheiden
- Bieten die Flexibilität, auf offene, komplexe Aufgaben in Echtzeit zu reagieren
- Aber: weiterhin durch Regeln und Vorgaben begrenzt

Use Case Beispiel: UI Testing Agent

EnBW Digital Office



Source: AppliedAI (2024)

Business Implikationen

Die Evolution der Prozessautomatisierung mit GenAI

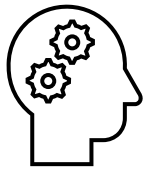
EnBW Digital Office

Robotic Process Automation	GenAI Chatbots	Co-Piloten	RAG / Custom Solutions	Agentic Workflows	KI-Agenten
Automatisierung repetitiver, regelbasierter Aufgaben mit definierten Abläufen	Reaktive Verarbeitung auf Anfrage	Generische KI-Assistenz mit erweitertem Kontextverständnis	Maßgeschneiderte KI-Lösungen mit domänenspezifischer Logik	Zusammenspiel für vorgegebene Aufgaben in flexiblen, strukturierten Prozessen	Prozessverantwortung mit adaptiver Selbstoptimierung
Regelbasierte Automatisierung	**Interaktionen**	**Augmentierte Zusammenarbeit**	**Domänenspezifische Optimierung**	**Prozesssteuerung mit klaren Leitplanken**	**Autonome Entscheidungsfindung**

Paradigmenwechsel: Von GenAI als unterstützendes Tool innerhalb fester Prozesse hin zu autonomen Systemen, die eigenständig Entscheidungen treffen und Prozesse steuern

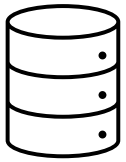
Daraus leiten sich drei zentrale Handlungsfelder ab

EnBW Digital Office



Kodifizierung von relevantem Wissen

- **Dokumentation von Geschäftsprozessen** und Erfassung von **implizitem Fachwissen** als Basis für KI-Agenten



Effektive Tool-Nutzung

- **Organisation von Daten und IT-Systemen** zur effektiven Zusammenarbeit mit Agenten
- **Erfassung von Kundeninteraktionen** für kontinuierliches Feedback und flexiblen Technologieeinbindung ohne Betriebsunterbrechung

Human-in-the-Loop & Qualitätssicherung

- Definieren von **Mechanismen**, um Ergebnisse auf Richtigkeit, Konformität und Fairness zu überprüfen
- **Einbindung von Fachexperten** zur Wartung und Skalierung sowie Etablierung eines kontinuierlichen Verbesserungsprozesses

Source: McKinsey (2024)

KI-Agente: Was sind neue Herausforderungen?

EnBW Digital Office

LLMs: Typische Herausforderungen

- Fehlerhafte oder verzerrte Ausgaben
- Prompt Injection und Datenlecks
- Bias in Trainingsdaten
- Begrenzte Transparenz und Erklärbarkeit

KI-Agenten: Neue Herausforderungen

- Fehlverhalten bei der Zielsetzung
- Unvorhersehbare Reaktionen auf reale Szenarien
- Koordination zwischen mehreren Systemen und Agenten
- Neue Formen der Mensch-KI-Zusammenarbeit

KI-Agenten sind nicht nur fortschrittliche Chatbots, sondern **autonome Entscheidungsträger**

Erfolgsfaktoren

1. **Robustheit** gewährleisten (Testing in authentischen Umgebungen, systematisches Monitoring)
2. **Prozessorientierte Metriken** entwickeln für Toolnutzung und Reasoning sowie für Kosten und Effizienz
3. **Anpassungsfähige** KI-Agenten designen (Lernfähigkeit, sichere Umgebungsadaption und nahtlose Systemintegration)
4. **Mensch-KI-Zusammenarbeit** optimieren (intuitive Gestaltung, klare Rollenverteilung, Schulungen)

KI-Agenten eröffnen neue Möglichkeiten für Unternehmensprozesse

EnBW Digital Office

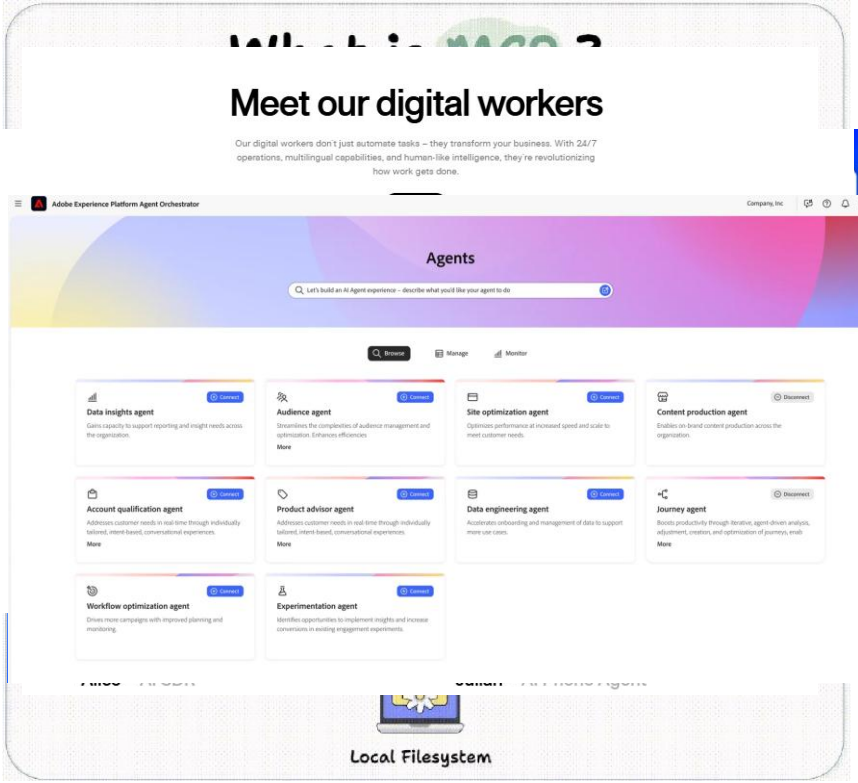
- **Autonome Aufgabenerledigung**, z.B. eigenständige Buchung komplexer Dienstreisen
- **Höhere Entscheidungsqualität**, z.B. automatisierte Priorisierung von Service-Tickets
- **Adaptabilität** an dynamische Szenarien in Echtzeit, z.B. Lieferengpässe
- **Simulation komplexer Systeme,** z.B. Cyberattacken und digitale Twins
- **Reflektions- und Lernfähigkeit**, z.B. Reaktion auf Feedback bei Kundenanfragen
- **Personalisierte Interaktion**, z.B. maßgeschneiderte Produktempfehlungen zu Nutzerpräferenzen

Aktuelle Trends & Abschluss

Aktuelle Trends und Entwicklungen

EnBW Digital Office

- **Interoperabilitätsstandards:** Das zukünftige Agenten-Ökosystem durch einheitliche Standards ermöglichen (z. B. Anthropic's Model Context Protocol MCP)
- **Computer-Use:** Verfeinerung der Fähigkeiten (z. B. Operator, ManusAI)
- **Agentic Interfaces:** Fokus auf neue Nutzerschnittstellen, die auf kontextreiche Zusammenarbeit mit Agenten ausgelegt sind
- **Multi-Agent Kollaboration:** spezialisierte Agenten, die kollaborativ oder adversarial Probleme lösen, z.B. Googles AI Co-Scientist
- **KI-Mitarbeitende:** Erste Start Ups mit neuen Pricing Modellen für KI Anwendungen



Key Takeways

EnBW
Digital Office

1. **Kernfähigkeiten:** Autonome Zielerreichung basierend auf Planung, Ausführung mit Tools und Evaluation

2. **Autonome Multi-Agenten-Systeme:** Für echte Autonomie braucht es noch mehr Kontrolle und Sicherheit; Einsatz für kreative oder Aufgaben mit offenen Fragestellungen und bei Akzeptanz von “stochastischem” Output

3. **Agentic Workflows:** Für viele Probleme ausreichend und leichter zu kontrollieren; Tipp: starte mit einem isolierten Problem, skaliere danach auf weitere Probleme (step by step)

4. **Strategische Basis:** Kodifizierung von Wissen, Organisation von Daten und IT-Systemen und Mensch-KI-Zusammenarbeit als Fundament für erfolgreiche Prozess-Transformation

5. **Evaluation & Monitoring:** Noch vieles offen, aber essenziell, um Robustheit zu gewährleisten

Tiefer einsteigen?

Jetzt seid ihr dran!

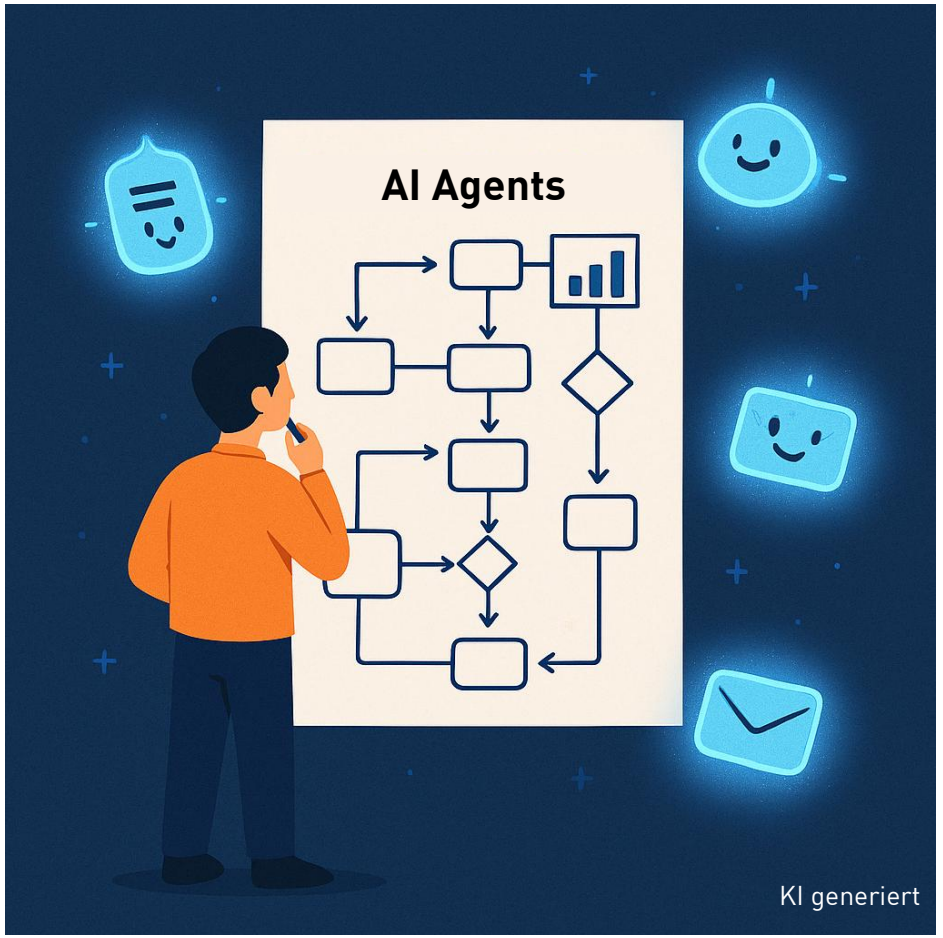
EnBW
Digital Office

Reflexions-Impulse

Wiederholt sich etwas regelmäßig, aber jedes Mal ein bisschen anders?

Wird Kontext gebraucht, um gute Entscheidungen zu treffen? Wo musst du nachdenken, vergleichen, bewerten?

Wo sammelst du Daten und Infos manuell oder springst ständig zwischen Tools hin und her?



Literatur

- Anthropic (2024). *Building Effective AI Agents*. Retrieved from: https://www.anthropic.com/engineering/building-effective-agents
- Anthropic (2024). *Introducing the Model Context Protocol*. Retrieved from: https://www.anthropic.com/news/model-context-protocol
- AppliedAI (2024). *Generative AI Agents in Action: Revolutionizing Software Development Testing*. Retrieved from: https://www.appliedai.de/en/insights/generative-ai-agents-in-action/
- Cemri, M., Pan, M. Z., Yang, S., Agrawal, L. A., Chopra, B., Tiwari, R., ... & Stoica, I. (2025). Why Do Multi-Agent LLM Systems Fail?. *arXiv preprint arXiv:2503.13657.*
- Luo, J., Zhang, W., Yuan, Y., Zhao, Y., Yang, J., Gu, Y., ... & Zhang, M. (2025). Large Language Model Agent: A Survey on Methodology, Applications and Challenges. *arXiv preprint arXiv:2503.21460.*
- Gartner (2025). *Emerging Patterns for Building LLM-Based AI Agents*. Retrieved from: https://www.gartner.com/en/documents/6142159
- McKinsey (2025). *Why agents are the next frontier of generative AI*. Retrieved from: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-agents-are-the-next-frontier-of-generative-ai
- Chip Huyen (2024). *Agents.* Retrieved from: https://huyenchip.com/2025/01/07/agents.html

Vielen Dank!

Franziska Zerwas

Data & AI Strategy
(Digital Office)

f.zerwas@enbw.com

EnBW
Digital Office
Zukunft. Digital. Machen.