

E06

AGENTEN BRAUCHEN ARCHITEKTUR

Warum Handlungsfähigkeit eine stärkere Steuerungsebene verlangt.

Ein KI-Agent kann Informationen abrufen, Schritte planen und über Werkzeuge Aktionen auslösen. Je mehr ein System selbstständig handeln darf, desto klarer müssen Auftrag, Berechtigungen, Zustand, Kontrolle und Rückfallwege geregelt sein.

MERKSATZ

Agenten handeln. Architektur führt.

1
Agent ist mehr als Chat

Er steuert Abläufe und kann externe Systeme oder Werkzeuge nutzen.

2
Autonomie braucht Grenzen

Berechtigungen, Freigaben und reversible Schritte müssen zum Risiko passen.

3
Klein starten ist professionell

Ein enger Agent mit wenigen Tools ist oft belastbarer als ein zu komplexes Netzwerk.

VON DER ANTWORT ZUR HANDLUNG

Stufe	Agent darf	Kontrolle
0 Auskunft	Strukturieren und vorschlagen.	Mensch handelt vollständig selbst.
1 Entwurf	E-Mail oder Bericht vorbereiten.	Prüfung vor externer Wirkung.
2 Vorbereitung	Daten abrufen, reversible Schritte vorbereiten.	Freigabe vor Versand oder Änderung.
3 Begrenzte Aktion	Niedrigrisiko-Aktionen in klaren Regeln.	Monitoring und definierte Stopps.
4 Hochrisiko	Finanzielle, rechtliche, irreversible Aktionen.	Keine autonome Entscheidung.

PRAXISBEISPIEL: AGENT FÜR NACHFASSUNGEN

Sicherer Start: freigegebene CRM-Daten lesen, offene Vorgänge erkennen und einen Entwurf erstellen. Keine Preisänderung, Vertragszusage oder Aussendung ohne Freigabe.

KURZCHECK

- ☐ Der Auftrag ist eng und eindeutig.
- ☐ Jedes Tool besitzt nur minimale Rechte.
- ☐ Irreversible Aktionen brauchen Freigabe.
- ☐ Stopps, Fehlergrenzen und Fallbacks sind definiert.

SIEBEN ARCHITEKTURBAUSTEINE

- 1
Rolle: Auftrag, Ziel, Zuständigkeit und ausgeschlossene Handlungen.
- 2
State: Was ist erledigt, offen und wann endet der Auftrag?
- 3
Memory: Was darf gespeichert, wiederverwendet oder gelöscht werden?
- 4
Tool-Rechte: Lesen, Schreiben, Senden und Bezahlen getrennt behandeln.
- 5
Guardrails: Eingaben, Ausgaben und Aktionen auf Grenzen prüfen.
- 6
Evaluation: Erfolg, Fehler und Abbrüche über reale Fälle messen.
- 7
Fallback: Bei Unsicherheit stoppen und nachvollziehbar an Menschen übergeben.

PRAXIS-TRANSFER: IN DREI SCHRITTEN STARTEN

1
Workflow begrenzen

Mit einer Aufgabe, wenigen Tools und einem klaren Endzustand beginnen.

2
Entwurf vor Aktion

Der Agent bereitet vor; externe oder irreversible Wirkung bleibt freigabepflichtig.

3
Kontrolliert erweitern

Fehler dokumentieren, Rechte staffeln und Autonomie nur bei belegter Zuverlässigkeit erhöhen.

TRANSFERFRAGE FÜR IHREN BETRIEB

Welche Agenten-Aktion wäre nützlich, aber noch sicher genug, um zunächst nur als Entwurf zu starten?

A practical guide to building agents
OpenAI for Business

Quelle

Building Effective AI Agents
Anthropic Research

Quelle