

Agenten sind keine Blackbox

Drei Ringe zeigen, was wirklich drinsteckt: Kern, Ausstattung,
Rahmen

Die **Demo** funktioniert immer

Ob dahinter ein Agent steckt oder ein abgesicherter Chatbot, siehst du erst, wenn du weißt, wonach du fragen musst.

Ringmodell

Drei konzentrische Ringe. Der **Kern** denkt, die Ausstattung macht das Modell zum eigenen Agenten, der Rahmen sichert ihn ab.

Seit Juli 2026 die verbindliche Antwort auf die Frage, was ein Agent ist.

Was das **Modell** selbst mitbringt

01 Reasoning

Mehrschrittig planen, sich selbst korrigieren. Der Sprung von Antworten zu Arbeiten.

02 Tool-Calling

Strukturierte Werkzeugaufrufe. So holt sich das Modell fehlende Fähigkeiten.

03 Instruction-Following

Sich zuverlässig an Vorgaben halten. Ohne das greift der Rahmen später ins Leere.

04 Loop*

Schleifen denken, reflektieren. Kommt nicht aus dem Modell, sondern über die Schnittstelle.

Was den Agenten zum eigenen Agenten macht

01 Tools

APIs, Browser, Datenbanken. Die Werkzeuge selbst.

02 Wissen

Was der Agent nachschlägt, nicht was er ist.

03 Skills

Gebündeltes Prozesswissen. Die Anleitung, die Werkzeuge benutzt, kein Werkzeug selbst.

04 Memory

Was über Schritte und Sessions hinweg bleibt. Das Gedächtnis, nicht seine Überwachung.

Was den Agenten absichert

01 Identität

Persona, Werte, Rolle im Systemprompt. Wer der Agent ist.

02 Guardrails

Verbotene Handlungen, harte Constraints.
Verbindlich erst durch technische Durchsetzung.

03 Prüfstand

Observability, Tracing, Evals. Der Blick von außen.

04 Interface

Was beim Nutzer ankommt. Die Bühne des Agenten.

Nachgezählt

12

Bestandteile, verteilt auf drei Ringe: vier im **Kern**, vier in der Ausstattung, vier im Rahmen.

kb:AI-Agents/Anatomie-Ringmodell.md

Was fehlt, wenn ein Ring leer bleibt

Nur Ausstattung

Mächtig und unkontrolliert. Tools, Wissen, Skills und Memory sind da, aber nichts hält sie im Zaum.

Ausstattung plus Rahmen

Steuerbar und geprüft. Derselbe Agent, aber mit Identität, Guardrails und einem Prüfstand, der zusieht.

Die **Lesekarte** in drei Sätzen

Kern, Ausstattung, Rahmen: Ein Agent ist immer alle drei zusammen, nie nur einer.

Instruction-Following ist die Brücke. Ohne sie greifen Identität und Guardrails ins Leere.

Der Prüfstand fehlt am häufigsten. Genau da lohnt die nächste Frage an jeden Anbieter.

Nimm die Lesekarte mit

Frag beim nächsten Pitch nach Kern, Ausstattung und Rahmen. Und danach, wer den Prüfstand macht.

unlearn.how