

PERSPEKTIVE · NOVEMBER 2026 · 9 MIN. LESEZEIT

# Agenten sind Mitarbeitende ohne Arbeitsvertrag

## Wie Unternehmen Aufgabe, Berechtigungen und Aufsicht für AI-Agenten festlegen



**Philipp Pöschke**
Founder, Paice

AUF EINEN BLICK

- Ein Agent verfolgt ein Ziel über mehrere Schritte und verändert dabei etwas in Ihren Systemen; deshalb muss vor dem Start feststehen, welche Schritte er ohne Freigabe gehen darf.
- Die Autonomie wird für jeden Schritt einzeln festgelegt und wächst stufenweise, nach einer Probezeit mit vorher vereinbarten Prüfgrößen.
- Ein Agenten-Steckbrief hält auf einer Seite fest, was der Agent tut, was er darf, wer ihn verantwortet und wie seine Arbeit nachvollziehbar bleibt.

Agenten übernehmen in Unternehmen Arbeitsschritte, die bisher Menschen erledigt haben. Sie legen Vorgänge an, ändern Datensätze, schreiben Nachrichten und stoßen Bestellungen an. Die Diskussion dreht sich meist um Modelle und Plattformen. Für den Einsatz im Betrieb zählen drei andere Fragen: Welche Schritte darf ein Agent ohne Freigabe gehen? Wer ist für ihn verantwortlich? Wie bleibt nachvollziehbar, was er getan hat?

Wir schlagen vor, Agenten einzuführen wie neue Mitarbeitende, mit einer Aufgabenbeschreibung, begrenzten Berechtigungen, einer verantwortlichen Person und einer Probezeit, nach der die Autonomie stufenweise wächst. In diesem Beitrag beschreiben wir vier Autonomiestufen und die Kriterien für die Einstufung, die Prüfgrößen der Probezeit, die typischen Risiken und einen Steckbrief, der auf eine Seite passt.

## Was einen Agenten ausmacht

Die alte Unterscheidung, ein Assistent antworte und ein Agent handle, trägt nicht mehr, seit verbreitete Büro- und Entwicklungswerkzeuge selbst Agentenfunktionen haben. Präziser ist eine Definition über die Wirkung. Ein Agent verfolgt ein Ziel über mehrere Schritte und nutzt dabei Werkzeuge, die etwas verändern. Er schreibt in Systeme, versendet, bestellt oder bucht. Für die Steuerung zählt deshalb nicht das Produkt, sondern ob ein Schritt Wirkung nach außen hat.

Im Modell des AI-Deltas gehören Agenten vor allem in das Feld Bauen. Ein Agent, der in der Vorführung überzeugt, muss im Betrieb mit echten Berechtigungen, Ausnahmen und Kosten zurechtkommen. Das Analystenhaus Gartner erwartet, dass bis Ende 2027 mehr als 40 Prozent der Projekte mit agentischer AI eingestellt werden, wegen steigender Kosten, unklaren Nutzens oder unzureichender Risikokontrollen. Ein Ordnungsrahmen, der vor dem Start steht, macht Nutzen, Kosten und Risiken prüfbar, bevor viel Geld gebunden ist.

## Vier Autonomiestufen

Autonomie lässt sich in Stufen festlegen. Die Stufe gilt dabei für einen einzelnen Schritt. Derselbe Agent kann eingehende Rechnungen selbst prüfen und trotzdem jede Zahlung zur Freigabe vorlegen.

ABBILDUNG 1

## Vier Autonomiestufen für Agenten




Darstellung: Paice

Welche Stufe ein Schritt bekommt, hängt von fünf Kriterien ab:

- **Umkehrbarkeit:** Lässt sich der Schritt ohne Schaden rückgängig machen, wie eine interne Weiterleitung, oder ist er endgültig, wie eine Zahlung oder eine Nachricht an Kunden?
- **Schadenshöhe:** Was kostet ein einzelner Fehler, finanziell und rechtlich?
- **Häufigkeit:** Ein Schritt, der tausendmal im Monat vorkommt, macht die Freigabe durch Menschen teuer und rechtfertigt mehr Aufwand für Stichproben und Überwachung.

- **Erkennbarkeit von Fehlern:** Fällt ein Fehler sofort auf, etwa durch eine Plausibilitätsprüfung, oder erst Wochen später beim Kunden?
- **Rechtliche Einordnung:** Berührt der Schritt einen Hochrisiko-Bereich des AI Act, personenbezogene Daten oder die Mitbestimmung?

Ein vereinfachtes Beispiel aus dem Einkauf zeigt, wie das aussieht. Ein Agent prüft Bestellanforderungen gegen Rahmenverträge. Bestellungen bis zu einem festen Betrag bei gelisteten Lieferanten löst er selbst aus, das ist Stufe 3. Alles darüber und jede Bestellung bei einem neuen Lieferanten legt er zur Freigabe vor, das ist Stufe 2. Den Betrag legt der Einkauf fest.

## Die Probezeit

Ein Agent startet auf der niedrigsten Stufe, die für seine Aufgabe sinnvoll ist. Ob er eine Stufe aufsteigt, entscheidet eine Probezeit mit Prüfgrößen, die vor dem Start feststehen:

- **Abweichungen in der Stichprobe:** Anteil der geprüften Vorgänge, in denen ein Mensch anders entschieden hätte.
- **Eingriffe:** Wie oft Menschen einen Schritt korrigieren, stoppen oder zurücknehmen mussten.
- **Schwere der Fehler:** Fehler nach Klassen gezählt, von folgenlos über korrigierbar bis zu Folgen außerhalb des Unternehmens.
- **Vorlagen:** Legt der Agent die Fälle vor, die er vorlegen soll? Ein übersehener Ausnahmefall wiegt schwerer als ein unnötig vorgelegter.
- **Kosten je Vorgang:** Modellnutzung, Rechenzeit und die Prüfzeit der Menschen zusammen.
- **Nachvollziehbarkeit:** Anteil der Vorgänge mit vollständigem Protokoll aus Eingang, Schritten, Begründung und Ergebnis.

Die Schwellen für jede Größe legt der verantwortliche Fachbereich fest, bevor der Agent startet, zusammen mit Dauer und Umfang der Stichprobe. Eine Stufe steigt erst, wenn alle Größen über eine vorher vereinbarte Mindestzahl von Vorgängen eingehalten sind. Umgekehrt führt ein schwerer Fehler sofort zurück auf die vorherige Stufe. Nach jedem Wechsel des Modells oder einer größeren Änderung der Anweisungen beginnt die Probezeit neu, denn dann arbeitet im Grunde ein anderer Agent. Das gilt auch, wenn der Anbieter das Modell hinter einer Plattform austauscht; Verträge sollten deshalb regeln, dass solche Wechsel angekündigt werden.

Über die höhere Stufe entscheidet die verantwortliche Person im Fachbereich, nachdem sie die Ergebnisse der Probezeit mit dem Team besprochen hat, das den Agenten betreibt. Die Entscheidung und ihre Grundlage gehören in den Steckbrief. So lässt sich auch Monate später zeigen, warum ein Agent einen Schritt selbst gehen durfte.

## Typische Risiken

Agenten erben die Risiken von Sprachmodellen und bringen eigene mit. Das OWASP-Projekt für die Sicherheit generativer AI hat sie im Dezember 2025 in einer eigenen Liste für agentische Anwendungen zusammengefasst. Vier davon sind für den Einsatz in Unternehmen besonders wichtig:

- **Manipulation über gelesene Inhalte.** Ein Agent liest E-Mails, Dokumente oder Webseiten. Enthalten sie versteckte Anweisungen, kann er sein Ziel wechseln, ohne dass es jemand bemerkt. Fachleute sprechen von indirekter Prompt Injection, OWASP von „Agent Goal Hijack“.
- **Zu weit gefasste Berechtigungen.** Die Versuchung ist groß, einem Agenten mehr Funktionen, Rechte oder Freiheit zu geben, als seine Aufgabe verlangt. OWASP nennt das „Excessive Agency“. Ein Agent, der Post nur sortieren soll, braucht kein Recht zum Versenden.
- **Fehlende eigene Identität.** Arbeitet ein Agent mit dem Konto einer Person oder einem gemeinsamen Dienstkonto, lässt sich später nicht trennen, wer was getan hat, und Rechte lassen sich nicht gezielt entziehen.
- **Fehlerketten.** Ein falscher Zwischenschritt wird zur Grundlage der nächsten Schritte, bei mehreren zusammenarbeitenden Agenten auch über Systemgrenzen hinweg. Dazu kommen Schleifen, in denen ein Agent immer wieder dieselbe Aktion versucht und dabei Kosten erzeugt.

Die Gegenmittel sind bekannt. Berechtigungen gehören in die Zielsysteme, weil sich ein Modell durch Inhalte umstimmen lässt. Jeder Agent bekommt eine eigene Identität mit genau den Rechten, die seine Aufgabe braucht. Folgenreiche Aktionen brauchen eine Freigabe, Budget- und Zeitgrenzen beenden Schleifen, und jedes Protokoll lässt sich einem Agenten und einer Aufgabe zuordnen.

## Was rechtlich dazukommt

Stand: 17. November 2026. Wird ein Agent in einem Hochrisiko-Bereich nach Anhang III des AI Act eingesetzt, etwa bei der Auswahl von Bewerberinnen und Bewerbern, gelten die Pflichten für Betreiber nach Art. 26. Sie verlangen unter anderem eine menschliche Aufsicht durch Personen mit der nötigen Kompetenz und Befugnis, die Überwachung des Betriebs und die Aufbewahrung der automatisch erzeugten Protokolle für mindestens sechs Monate. Arbeitgeber müssen Arbeitnehmervertretung und betroffene Beschäftigte vor dem Einsatz informieren. Nach dem Omnibus (Verordnung (EU) 2026/1744) gelten diese Pflichten für Systeme nach Anhang III ab 2. Dezember 2027.

Im Betrieb kommt das Betriebsverfassungsgesetz hinzu. Nach § 90 Abs. 1 Nr. 3 BetrVG ist der Betriebsrat rechtzeitig über die Planung von Arbeitsverfahren und Arbeitsabläufen zu unterrichten, ausdrücklich einschließlich des Einsatzes von Künstlicher Intelligenz. § 87 Abs. 1 Nr. 6 BetrVG gibt ihm ein Mitbestimmungsrecht bei technischen Einrichtungen, die dazu bestimmt sind, Verhalten oder Leistung zu überwachen. Nach ständiger Rechtsprechung des Bundesarbeitsgerichts genügt dafür, dass eine Einrichtung objektiv zur Überwachung geeignet ist. Ein Agent, der festhält, wer welchen Vorgang wie bearbeitet hat, kann das erfüllen. Muss der Betriebsrat den Einsatz von AI beurteilen, gilt die Hinzuziehung eines Sachverständigen als erforderlich (§ 80 Abs. 3 BetrVG).

Für Unternehmen, die Agenten in eigene Produkte einbauen, wird die neue Produkthaftungsrichtlinie (EU) 2024/2853 wichtig. Sie bezieht Software ausdrücklich ein und ist bis 9. Dezember 2026 umzusetzen. Der Bundestag hat das neue Produkthaftungsgesetz am 8. Oktober 2026 beschlossen; es soll am 9. Dezember 2026 in Kraft treten.

Dieser Abschnitt ist eine allgemeine Einordnung und keine Rechtsberatung.

## Der Agenten-Steckbrief

Was vor dem Start feststehen muss, passt auf eine Seite. Der Steckbrief dient als Freigabeunterlage, als Grundlage für die Probezeit und als Eintrag in Ihr Verzeichnis der AI-Systeme. Er wird gepflegt wie eine Stellenbeschreibung und bei jeder Stufenänderung aktualisiert.

ABBILDUNG 2

### Der Agenten-Steckbrief: was vor dem Start feststeht

01

**Aufgabe**

Was der Agent erledigt, für wen und wo seine Aufgabe endet. Dazu die Fälle, die er ausdrücklich nicht bearbeitet.

02

**Berechtigungen**

Systeme und Daten, auf die er zugreift, jeweils lesend oder schreibend, unter einer eigenen Identität. Budget- und Zeitgrenzen je Vorgang.

03

**Autonomiestufe**

Die Stufe für jeden Schritt, begründet mit Umkehrbarkeit, Schadenshöhe, Häufigkeit, Erkennbarkeit von Fehlern und rechtlicher Einordnung.

04

**Verantwortung**

Eine verantwortliche Person im Fachbereich, eine Stelle für den technischen Betrieb und die Personen, die freigeben und eingreifen dürfen.

05

**Probezeit**

Prüfgrößen, Schwellen, Dauer und Stichprobe. Wer über eine höhere Stufe entscheidet und was zur Rückstufung führt.

06

**Nachweis und Notfall**

Was protokolliert wird und wie lange. Wie der Agent angehalten wird und wer das darf.

Darstellung: Paice

## Was sich für die Teams ändert

Mit Agenten verschiebt sich die Arbeit der Menschen, die mit ihnen arbeiten. Sie bearbeiten weniger Standardfälle selbst und mehr Ausnahmen, sie prüfen Stichproben und entscheiden über die Fälle, die der Agent vorlegt. Diese Aufgaben brauchen eine eigene Beschreibung und Zeit im Arbeitsalltag, sonst bleibt die Aufsicht ein Häkchen im Formular. Eine Lösung, die technisch funktioniert und im Alltag umgangen wird, ist nicht verankert.

> Behandeln Sie die Einführung eines Agenten wie eine Neubesetzung, mit klarer Aufgabe, begrenzten Berechtigungen, einer verantwortlichen Person und einer Probezeit.

## Quellen


- OWASP Top 10 for Agentic Applications for 2026, OWASP GenAI Security Project, 09.12.2025
- LLM06:2025 Excessive Agency, OWASP GenAI Security Project, 2025
- Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027, Gartner, 25.06.2025
- Art. 26 AI Act, Pflichten der Betreiber von Hochrisiko-Systemen, AI Act Service Desk der EU-Kommission, Fassung vom 27.07.2026
- Verordnung (EU) 2026/1744 (Digital Omnibus on AI), Amtsblatt der EU, 24.07.2026
- § 90 BetrVG, Bundesministerium der Justiz, gesetze-im-internet.de
- § 87 BetrVG, Bundesministerium der Justiz, gesetze-im-internet.de
- § 80 BetrVG, Bundesministerium der Justiz, gesetze-im-internet.de
- Richtlinie (EU) 2024/2853 über die Haftung für fehlerhafte Produkte, Amtsblatt der EU, 18.11.2024
- Bundestag beschließt Modernisierung des Produkthaftungsrechts, Deutscher Bundestag, 08.10.2026





AUTOR

### Philipp Pöschke

Founder, Paice

Philipp Pöschke hat Paice 2026 in Bremen gegründet. Er verantwortet jedes Vorhaben persönlich und arbeitet selbst an Analysen, Prototypen und Lösungen mit. Die Beiträge unter Insights verantwortet er als Autor.