



HARBOUR

WHITEPAPER

Von Chatbots zu Agenten

Wo sie wirklich arbeiten, und was im Mittelstand funktioniert.

DER AGENTEN-REALITÄTSCHECK 2026

Von Chatbots zu Agenten

Agenten sollen die nächste Stufe der KI sein. In Benchmarks erreichen sie Spitzenwerte, **im Betrieb scheitern die meisten.**

Dieser Bericht zeigt, wo Agenten heute wirklich arbeiten, warum der Hype größer wirkt als die Praxis, und was ein Setup braucht, das im deutschen Mittelstand funktioniert.

40%

der agentischen KI-Projekte werden bis Ende 2027 abgebrochen

GARTNER, 2025

~17%

der Mittelständler setzen bereits KI-Agenten ein

KI-INDEX MITTELSTAND

95%

der KI-Pilotprojekte ohne messbaren Ergebnisbeitrag

MIT · STATE OF AI 2025

Was dieser Bericht ordnet

IN KÜRZE

Vieles, was als Agent verkauft wird, ist regelbasierte Automatisierung mit einem KI-Schritt. Entscheidend ist die richtige Zuordnung.

Die produktivsten Agenten haben kein Chat-Interface. Sie laufen unsichtbar im Hintergrund.

Der Sprung von 95 auf 100 % Genauigkeit ist der teuerste. Bei Geld und Recht sind 5 % Fehler ein Haftungsrisiko.

Erfolgreiche Setups arbeiten nach dem Human-in-the-Loop-Prinzip: Der Agent macht die Vorarbeit, der Mensch gibt frei.

Gartner erwartet, dass über 40 % der Agenten-Projekte bis Ende 2027 abgebrochen werden – meist fehlt der Geschäftsnutzen.

KAPITEL

01	Vom Chatbot zum Agenten	04
02	Wo das im Arbeitsalltag vorkommt	06
03	Die Kluft zwischen Benchmark und Büro	07
04	Warum der Hype größer wirkt als die Praxis	08
05	Der deutsche Kontext	10
06	Was die erfolgreichen Setups anders machen	11
07	So arbeitet Harbour	12
—	Anhang: Agentenreife-Check und Datenlage	13

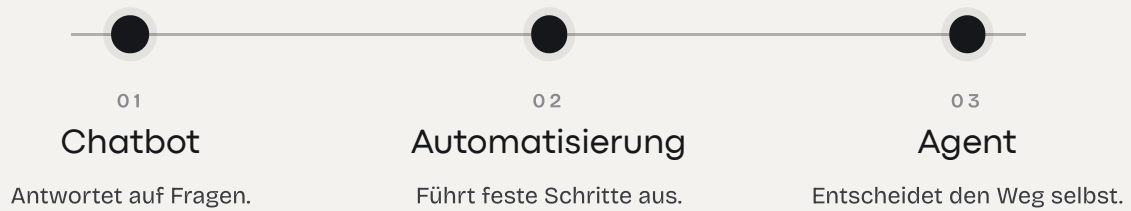
EIN BERICHT MIT FOKUS DEUTSCHER MITTELSTAND: WO AGENTEN HEUTE ARBEIT ÜBERNEHMEN, WORAN DIE MEISTEN SCHEITERN, UND WAS EIN SETUP BRAUCHT, DAS IM BETRIEB LÄUFT STATT IN DER DEMO.

01

DREI STUFEN DER EIGENSTÄNDIGKEIT

Vom Chatbot zum Agenten

ZUNEHMENDE EIGENSTÄNDIGKEIT



DER MENSCH ARBEITET

DAS SYSTEM ARBEITET

Die Entwicklung lässt sich in drei Stufen erzählen. Der **Chatbot** antwortet: Man stellt eine Frage, er gibt Text zurück, die Arbeit bleibt beim Menschen. Die **Automatisierung** führt feste Schritte aus: Ein Ablauf steht fest und läuft immer gleich, KI erledigt darin einzelne Handgriffe wie das Auslesen eines Dokuments. Der **Agent** bekommt ein Ziel und entscheidet selbst, welche Schritte und Werkzeuge er dafür nutzt.

Wer legt den Weg fest?

Der Unterschied zwischen Automatisierung und Agent klingt akademisch, entscheidet aber über Kosten und Risiko. Bei einer Automatisierung legt der Entwickler den Weg im Voraus fest – gleiche Eingabe, gleicher Pfad. Bei einem Agenten entscheidet das System den Weg zur Laufzeit: Es bekommt ein Ziel und wählt seine Schritte selbst.



Fachlicher Bezug: entspricht der Linie aus Anthropics „Building Effective Agents“ – vordefinierte Code-Pfade gegen ein Modell, das seinen Ablauf selbst steuert.

Wichtig ist, was diese Entwicklung **nicht** ist: Keine Stufe löst die vorige ab. Nicht jeder Prozess braucht einen Agenten – an einer Stelle, an der eine feste Regel genügt, ist er teurer, langsamer und schwerer zu kontrollieren. Die Kunst liegt in der richtigen Zuordnung.

02

VIER FELDER AUS DER PRAXIS

Wo das im Arbeitsalltag vorkommt

In der Praxis mischen sich feste Abläufe und selbstständige Entscheidungen. Die nützliche Frage ist deshalb weniger, ob etwas ein Agent ist, als an welcher Stelle im Prozess ein eigenes Urteil nötig wird.

01 Beleg- und Rechnungsverarbeitung

Rechnungen werden ausgelesen, mit Bestellung und Lieferschein im ERP abgeglichen und bei Übereinstimmung eingebucht. Abweichungen gehen an einen Menschen.

02 Kundenservice und IT-Ticketing

Routine-Tickets wie Stornierungen oder Rechtevergaben werden selbstständig gelöst. Komplexe Anfragen werden angereichert und vorausgefüllt übergeben.

03 Recruiting und HR

Lebensläufe werden gegen Anforderungsprofile geprüft, Termine über mehrere Kalender abgestimmt, das Bewerbermanagement gepflegt.

04 Code-Wartung

Kleinere Refactorings entstehen als fertiger Pull-Request, Fehlerprotokolle werden ausgewertet, fehlende Tests ergänzt.

~17%

der Mittelständler setzen bereits KI-Agenten ein, fast doppelt so viele wie im Vorjahr.

KI-INDEX MITTELSTAND (SALESFORCE / DMB)

03

DEMO GEGEN BETRIEB

Die Kluft zwischen Benchmark und Büro

Agenten lösen in Demos beeindruckende Aufgaben. Der Alltag im Unternehmen sieht anders aus.

Auf isolierten, sauber vorbereiteten Pfaden erreichen Agenten Spitzenwerte. Im Betrieb scheitern sie selten am Modell. Sie scheitern an Berechtigungen, an verstreuten Daten und an einer Null-Toleranz für Fehler in Prozessen, die Geld oder Recht berühren.

**95 % der KI-Pilotprojekte liefern keinen
messbaren Beitrag zum Geschäft.**

Es ist dieselbe Lücke, die KI insgesamt bremst. Bei Agenten wird sie nur greifbarer, weil ein Agent handelt und nicht nur antwortet.

04

DREI GRÜNDE FÜR DEN ABSTAND

Warum der Hype größer wirkt als die Praxis

01 Unsichtbarer Hintergrund-Betrieb

Die produktivsten Agenten sind eng umgrenzte Dienste in Automatisierungsplattformen. Ohne Chat-Oberfläche nimmt sie niemand als Agent wahr. Der sichtbare Chatbot nimmt selten wirklich Arbeit ab.

02 Die 5-Prozent-Hürde

Ein Agent, der zu 95 % korrekt arbeitet, erreicht im Benchmark Spitzenwerte. Bei Verträgen oder Überweisungen sind 5 % Fehler ein Haftungsrisiko. Der Weg von 95 auf 100 % ist der teuerste Teil eines Projekts.

03 Vorbereitung statt Autonomie

Die erfolgreichen Setups arbeiten nach dem Human-in-the-Loop-Prinzip. Der Agent erledigt den Großteil der Vorarbeit, die Ausführung bleibt am Klick eines Menschen.

//
95 % richtig klingt gut.
Bei 1.000 Überweisungen
sind das 50 Fehler.

Der Sprung von 95 auf 100 Prozent Genauigkeit ist der teuerste. Bei Rechnungen, Verträgen oder Überweisungen ist eine Fehlerquote von 5 Prozent ein Haftungsrisiko, kein Benchmark-Detail.

05

HAFTUNG, DATENSCHUTZ, GEWACHSENE SYSTEME

Der deutsche Kontext

Im deutschen Mittelstand ist die 5-Prozent-Hürde schärfer: weniger Toleranz für Fehler, strengere Anforderungen an den Datenschutz und mit dem EU AI Act eine wachsende Nachweispflicht. Dazu kommen gewachsene Systeme, in die ein Agent sich einfügen muss, statt auf der grünen Wiese zu starten.

Wer im deutschen Kontext startet, ohne Haftung, Datenschutz und Anbindung von Anfang an mitzudenken, landet in der Abbruchquote.

DAS WICHTIGSTE

Weniger Fehlertoleranz, mehr Nachweispflicht (EU AI Act), gewachsene Systeme. Erfolg entscheidet sich an Haftung, Datenschutz und Anbindung – von Anfang an.

40%

der agentischen KI-Projekte werden bis Ende 2027 abgebrochen – wegen unklarem Geschäftsnutzen, steigender Kosten oder unzureichender Kontrolle.

GARTNER, 2025

06

DREI PRINZIPIEN

Was die erfolgreichen Setups anders machen

01 Eng beginnen

Ein klar umrissener Prozess mit hohem Volumen, einer klaren Datenquelle und der Möglichkeit, Fehler abzufangen, bevor sie teuer werden.

02 Den Menschen in der Kontrolle halten

Genau an den Stellen, an denen ein Fehler teuer wird, gibt ein Mensch frei. Human-in-the-Loop ist kein Kompromiss, sondern das Setup.

03 Den Betrieb sichtbar machen

Welche Aufgaben laufen, was sie kosten, wo jemand freigegeben muss. Transparenz ist die Voraussetzung dafür, dass ein Team Agenten überhaupt vertraut.

Agenten, die im Betrieb laufen statt in der Demo.

Genau solche Agenten bauen wir: unsichtbar, eng umgrenzt, an echte Prozesse gekoppelt, mit dem Menschen in der Kontrolle und einem Dashboard, das zeigt, was läuft und was es kostet.

01 Wir fangen dort an, wo der Aufwand ist

Bei dem Prozess, der bei Ihnen den größten manuellen Aufwand bindet.

02 Wir liefern Betrieb, keine Demo

Eng umgrenzt, an echte Systeme gekoppelt, mit Freigabe an den kritischen Stellen und einem Dashboard über Aufgaben und Kosten.

03 Wir bleiben, bis es Ihres ist

Bis Ihr Team eigenständig damit arbeitet. Kein Konzept für die Schublade. Etwas, das läuft.

SELBSTCHECK

Welche Prozesse sind agentenreif?

Je mehr Punkte ein Prozess erfüllt, desto eher funktioniert ein Agent.

-
- 01 Er wiederholt sich oft und folgt klaren Regeln.
-
- 02 Das Volumen ist hoch genug, dass sich Automatisierung lohnt.
-
- 03 Die nötigen Daten sind auffindbar und in nutzbarem Zustand.
-
- 04 Ein Fehler lässt sich auffangen, bevor er teuer wird.
-
- 05 Es gibt eine Stelle, an der ein Mensch freigibt.
-

Zur Datenlage

95 % der Pilotprojekte ohne Ergebnis – MIT, State of AI in Business 2025.

Abbruchquote agentischer Projekte, über 40 % bis Ende 2027 – Gartner, 2025.

Verbreitung von KI-Agenten im Mittelstand – KI-Index Mittelstand (Salesforce / DMB).

Die vier Anwendungsfälle illustrieren gängige Praxis. Anbieter-Blogs dienen als Ideengeber, nicht als Beleg.

HINWEIS

Wirkungszahlen aus Anbieter-Blogs sind Marketing-Angaben und stehen daher bewusst nicht in diesem Bericht. Aussagekräftig ist das Muster, nicht der beworbene Bestwert.