WHITEPAPER

KI-Agenten im Mittelstand einführen

Leitfaden von der ersten Idee bis zum laufenden Betrieb

Stand: 8. November 2026

intell▲agens

Inhalt

Das Wichtigste in Kürze 3

Den Einsatzfall wählen 3

Vorher messen 5

Abnahmekriterien schriftlich festlegen 8

Modell, Hosting, Bau und Einbindung 11

Betreiben 14

Kosten trennen 16

So erkennen Sie eine gute KI-Agentur 19

Quellen 22

Das Wichtigste in Kürze

- Wählen Sie einen Ablauf, nicht „KI im Haus“. Ein guter erster Einsatzfall ist häufig, folgt klaren Regeln, hat einen eindeutigen Anfang und ein eindeutiges Ende und wird heute von Hand erledigt.
- Messen Sie vorher. Ohne Zahlen zu Minuten je Vorgang, Menge je Woche und Fehlern können Sie später nicht sagen, ob der Agent geholfen hat. Der Messbogen in Kapitel 3 reicht dafür.
- Legen Sie vor dem Bau schriftlich fest, wann der Agent als fertig gilt: Prüffälle aus echten Vorgängen, geforderte Trefferquote, Umgang mit Ausnahmen, Freigaben durch Menschen, Frist zur Nachbesserung. Die Vorlage steht in Kapitel 4.
- Wählen Sie Modell und Hosting nach Ihren Daten, nicht nach dem Namen des Modells, und bauen Sie so, dass sich das Modell später tauschen lässt.
- Planen Sie den Betrieb ein. Formate, Kunden, Software und Modelle ändern sich. Jemand muss den Agenten beobachten und anpassen.
- Trennen Sie die Kosten in Bau, Betrieb sowie Modell und Hosting. Nach Bitkom liegt der Aufwand bei Anwendern weniger in Lizenzen als in Infrastruktur, Daten und Anbindung.
- Prüfen Sie Agenturen an ihren schriftlichen Antworten, nicht an Vorführungen. Die Prüfliste in Kapitel 8 enthält zehn Fragen.

Wer diese sieben Punkte beherzigt, hat nach dem Lesen alles in der Hand, um einen ersten Einsatzfall auszuwählen, den heutigen Stand zu messen und Abnahmekriterien zu formulieren. Das gelingt auch ohne Agentur. Wenn Sie danach bauen lassen, sprechen Sie mit der Agentur über konkrete Zahlen statt über Absichten.

Dieser Leitfaden stammt von Intellagens, einer KI-Agentur in Hamburg. Wir beschreiben darin die Methode, nach der wir selbst arbeiten. Wo wir etwas aus eigener Erfahrung sagen und nicht aus einer Studie, steht das dabei. Zahlen aus Studien haben eine Quelle im selben Absatz. Rechtliche Hinweise sind Stand 8. November 2026 und keine Rechtsberatung.

Den Einsatzfall wählen

Was ein Agent ist und was nicht

Ein Chatprogramm antwortet, wenn Sie etwas fragen. Ein Agent handelt selbstständig innerhalb eines Ablaufs. Bitkom beschreibt ihn so: Anders als ein Chatbot, der auf Fragen antwortet, handelt ein Agent selbstständig, indem er zum Beispiel Preisvergleiche durchführt, Angebote einholt und Entscheidungsoptionen ausarbeitet (Bitkom). Für Ihren Alltag heißt das: Ein Agent liest die Mail, schaut ins Warenwirtschaftssystem, legt etwas an oder meldet sich, wenn er nicht weiterkommt. Er ist kein weiteres Fenster, in das jemand Fragen tippt, sondern Teil eines Ablaufs.

Wo der Mittelstand steht

Bitkom hat 603 Unternehmen ab 20 Beschäftigten befragt. Danach nutzen 57 Prozent der Unternehmen KI. Bei Agenten gilt: 11 Prozent der Unternehmen, die KI nutzen oder deren Einsatz planen oder diskutieren,

setzen sie schon ein, 29 Prozent planen den Einsatz und 31 Prozent diskutieren ihn. Für 24 Prozent ist das Thema derzeit keines (Bitkom). Wichtig ist die Bezugsgröße: Die 11 Prozent sind kein Anteil aller Unternehmen. Agenten stehen am Anfang. Das ist Ihre Chance und Ihr Risiko zugleich: Es gibt noch wenig erprobte Muster, deshalb lohnt es sich, den ersten Fall klein und messbar zu wählen.

Fünf Merkmale eines guten ersten Einsatzfalls

Wir wählen Fälle nach diesen fünf Merkmalen aus; sie sind keine Studienergebnisse. Je mehr zutrifft, desto eher lohnt sich der Fall.

1. **Häufig.** Der Ablauf kommt mehrmals pro Woche oder täglich vor. Bei seltenen Abläufen frisst die Einrichtung den Nutzen.
2. **Regelhaft.** Ein Mensch kann die Entscheidungen in wenigen Sätzen erklären. Wenn jedes Mal anders entschieden wird, ist der Ablauf (noch) kein Fall für einen Agenten.
3. **Eindeutiger Anfang und Ende.** Es gibt einen Auslöser (eine Mail, eine Datei, ein Zeitpunkt) und ein Ergebnis, das man prüfen kann (ein angelegter Auftrag, eine versendete Erinnerung, eine Meldung).
4. **Prüfbar.** Ein Mensch kann in kurzer Zeit sehen, ob das Ergebnis stimmt. Das ist die Voraussetzung dafür, Ausnahmen abzufangen.
5. **Gepflegte Daten.** Die Systeme, die der Agent lesen oder beschreiben soll, sind aufgeräumt: Artikelnummern, Kundenstämme, Preise. Nach Bitkom nennen KI-Anwender als großen Kostenpunkt die Aufbereitung von Daten (50 Prozent) und die Integration in bestehende Systeme (41 Prozent), Abos und Lizenzen nur 21 Prozent (Bitkom). Wer seine Daten nicht kennt, kennt den Aufwand nicht.

Wo ein Agent nichts bringt

Ein Agent lohnt sich nicht, wenn ein Ablauf selten vorkommt, wenn jedes Mal anders entschieden wird oder wenn die Daten dahinter nicht gepflegt sind. Dann ist die bessere Investition zuerst die Ordnung der Daten oder eine einfache Regel im vorhandenen Programm. Das Mittelstand-Digital Zentrum Berlin formuliert es nüchtern: KI sei ein möglicher Ansatz, aber kein universelles Allheilmittel, und eine Bedarfsanalyse solle zunächst klären, ob der Einsatz von KI sinnvoll und erforderlich ist (Mittelstand-Digital Zentrum Berlin). Wer das ehrlich prüft, spart sich ein Projekt, das nach drei Monaten wieder verschwindet.

Arbeitshilfe: Bewertungsbogen für Einsatzfälle

Sammeln Sie zuerst fünf bis zehn Abläufe, die Ihr Team viel Zeit kosten. Fragen Sie in den Fachabteilungen: „Welche Arbeit machen Sie jede Woche, die sich ähnelt und die Sie ungern tun?“ Tragen Sie die Abläufe in den Bogen ein und vergeben Sie je Merkmal null bis zwei Punkte (0 = trifft nicht zu, 1 = teilweise, 2 = trifft zu).

Ablauf	Häufig	Regelhaft	Klarer Anfang und Ende	Prüfbar	Daten gepflegt	Summe
1. _____	_	_	_	_	_	_
2. _____	_	_	_	_	_	_
3. _____	_	_	_	_	_	_
4. _____	_	_	_	_	_	_
5. _____	_	_	_	_	_	_

So werten Sie aus: Die höchste Summe nennt den Kandidaten. Zehn Punkte sind das Maximum. Wählen Sie nicht blind den Ablauf mit der höchsten Summe, sondern prüfen Sie ihn auf zwei Fragen: Wer in Ihrem Haus ist fachlich verantwortlich und hat Zeit, mitzuarbeiten? Und: Was passiert, wenn der Agent einen Fehler macht? Ein Ablauf, bei dem ein Fehler Kunden verärgert oder Geld bewegt, braucht strengere Kriterien (siehe Kapitel 4) und gehört deshalb eher an den zweiten Platz.

Ein Beispiel für gut gewählte Fälle aus unserer Praxis: Beim Großhändler Wheyster.de laufen Bestellungen, Zahlungserinnerungen und ein Lager-Alarm als Agenten. Sie erfüllen die Merkmale oben in unterschiedlichem Maß, haben aber alle einen klaren Auslöser und ein prüfbares Ergebnis. Der Alarm vergleicht täglich um 12 Uhr 428 Artikel mit ihrem Mindestbestand; die Zahlungserinnerungen gehen nur an Rechnungen, die wirklich offen sind. Stand der Zahlen: 29.10.2026, Minuten und Häufigkeit nennt der Kunde.

Vorher messen

Warum messen, bevor etwas gebaut wird

Wer vorher nichts misst, kann hinterher nichts belegen. „Es fühlt sich schneller an“ überzeugt weder die Geschäftsführung noch den Betriebsrat noch Sie selbst nach einem Jahr. Mit einer Messung vor dem Bau erreichen Sie drei Dinge:

- Sie wissen, ob sich der Fall überhaupt lohnt, bevor Geld fließt.
- Sie haben die Grundlage für die Abnahmekriterien (Kapitel 4). Die Trefferquote eines Agenten bewerten Sie am besten gegen die heutige Fehlerquote der Handarbeit.
- Sie können nach drei Monaten Betrieb denselben Bogen noch einmal ausfüllen und die Spalten vergleichen.

Das Mittelstand-Digital Zentrum Berlin empfiehlt dasselbe aus Sicht der Anbieterwahl: Statt allgemeiner Effizienzgewinne sollen Sie Kriterien festlegen, an denen sich Erfolg messen lässt, zum Beispiel eine schnellere Bearbeitung von Anfragen (Mittelstand-Digital Zentrum Berlin).

Was Sie messen

Für jeden Ablauf erfassen Sie fünf Dinge:

1. **Menge:** Wie viele Vorgänge pro Woche? Zählen Sie vier typische Wochen, nicht eine. Schwankungen (Monatsende, Saison) gehören dazu.
2. **Zeit je Vorgang:** Wie viele Minuten braucht ein Vorgang von Hand? Stoppen Sie mindestens zehn Vorgänge, getrennt nach einfach und schwierig. Fragen Sie nicht aus dem Gedächtnis, Schätzungen fallen fast immer zu niedrig oder zu hoch aus.
3. **Fehler:** Wie viele Vorgänge enthalten heute einen Fehler, und welche Folgen hat er? Auch Handarbeit macht Fehler. Wenn Sie die heutige Quote nicht kennen, können Sie die des Agenten nicht einordnen.
4. **Durchlaufzeit:** Wie lange dauert es vom Auslöser bis zum Ergebnis? Manchmal ist nicht die Arbeitszeit das Problem, sondern das Warten.
5. **Ausnahmen:** Wie viele Vorgänge weichen vom Normalfall ab, und woran erkennen Sie das? Das sind die späteren Fälle, die der Agent an einen Menschen abgeben soll.

Arbeitshilfe: Messbogen

Kopieren Sie diese Tabelle je Ablauf. Füllen Sie die Spalte „Heute“ vor dem Bau aus und die Spalte „Nach drei Monaten Betrieb“ mit denselben Methoden erneut.

Ablauf: _____ **Verantwortlich im Haus:** _____ **Messzeitraum von/bis:** _____

Messgröße	So wird gemessen	Heute	Nach drei Monaten Betrieb
A: Vorgänge pro Woche, Woche 1	zählen	___	___
Vorgänge pro Woche, Woche 2	zählen	___	___
Vorgänge pro Woche, Woche 3	zählen	___	___
Vorgänge pro Woche, Woche 4	zählen	___	___
Mittelwert A (Summe ÷ 4)	rechnen	___	___
B: Minuten je einfachem Vorgang	mindestens 10 Vorgänge stoppen	___	___
C: Minuten je schwierigem Vorgang	mindestens 10 Vorgänge stoppen	___	___
D: Anteil schwieriger Vorgänge	schwierige ÷ alle Vorgänge	___ %	___ %
E: Vorgänge mit Fehler (Anzahl von 100 geprüften)	100 Vorgänge nachprüfen	___	___
F: Folge eines Fehlers	beschreiben: Kosten, Zeit, Ärger	___	___
G: Durchlaufzeit vom Auslöser zum Ergebnis	Zeitstempel vergleichen	___	___
H: Beteiligte Personen	zählen	___	___

Rechnung (Zeit je Woche heute):

Zeit je Woche in Minuten = A × [(B × (1 − D)) + (C × D)]

Dabei ist A der Mittelwert der Vorgänge pro Woche, B die Minuten je einfachem Vorgang, C die Minuten je schwierigem Vorgang und D der Anteil schwieriger Vorgänge als Dezimalzahl (zehn Prozent = 0,1). Teilen Sie das Ergebnis durch 60 für Stunden.

Zeit je Jahr: Stunden je Woche × W, wobei W die Zahl der Wochen im Jahr ist, in denen der Ablauf läuft. Ziehen Sie Wochen ab, in denen er ruht.

Prüfen Sie das Ergebnis auf Plausibilität: Stimmt die Summe über alle Abläufe grob mit dem überein, was die Fachabteilung insgesamt an Arbeitszeit für diese Tätigkeiten rechnet? Weicht sie stark ab, haben Sie sich verzählt oder eine Tätigkeit vergessen.

Wie Sie einen Messwert ehrlich angeben

Vermerken Sie bei jeder Zahl, wer sie ermittelt hat und wie. „Gestoppt von zwei Personen an zehn Vorgängen“ ist belastbarer als „geschätzt“. Rechnen Sie vorsichtig: Im Zweifel nehmen Sie den niedrigeren Nutzen. Wir empfehlen dieses Vorgehen; so sind auch die Wheyster-Zahlen entstanden: Minuten und Häufigkeit nennt der

Kunde, es ist keine unabhängige Messung, gezählt ist jeweils der vorsichtige Wert. Dort sind es vorher mindestens 8 Stunden Büroarbeit pro Woche von Hand, insgesamt sind seit dem Start 272 Stunden gespart, vorsichtig gerechnet. Nicht eingerechnet sind von uns behobene eigene Fehler und die Wirkung der Mahnungen auf Zahlungseingänge. Stand der Zahlen: 29.10.2026.

Was die Messung entscheidet

Nach der Messung entscheiden Sie anhand dreier Fragen, ob der Fall weiterverfolgt wird:

- Wie viele Stunden je Jahr sind betroffen? Vergleichen Sie das Ergebnis mit dem, was Bau und Betrieb kosten (Kapitel 7). Eine Schwelle für „lohnt sich“ legen Sie selbst vorab fest, etwa als Stunden je Woche, ab denen Sie weitermachen. Tun Sie das vor der Messung, damit das Ergebnis die Schwelle nicht verschiebt.
- Wie hoch ist der Schaden eines Fehlers? Davon hängt die geforderte Trefferquote ab.
- Gibt es Hindernisse bei Daten oder Schnittstellen? Eine Messung, die zeigt, dass die Stammdaten nicht gepflegt sind, ist ein gutes Ergebnis: Sie sparen sich einen Bau, der daran gescheitert wäre.

Abnahmekriterien schriftlich festlegen

Fertig heißt: besteht vereinbarte Prüffälle

Ein KI-Agent ist fertig, wenn er vorab vereinbarte Prüffälle besteht, nicht wenn er in einer Vorführung gut aussieht. Klassische Software tut bei gleicher Eingabe immer das Gleiche. Ein Agent mit Sprachmodell antwortet in Grenzen unterschiedlich und macht Fehler, die man in der Vorführung nicht sieht. Ob er verlässlich arbeitet, zeigt sich erst an vielen echten Fällen. Das beschäftigt auch die Anwender: Laut Bitkom nennen 26 Prozent der Unternehmen mit KI im Einsatz schlechte Qualität der Ergebnisse als Hindernis, 25 Prozent mangelnde Nachvollziehbarkeit (Bitkom). Messbare Kriterien machen beides prüfbar.

Das Mittelstand-Digital Zentrum Berlin rät ebenfalls, mit den Fachabteilungen Akzeptanzkriterien festzulegen, etwa die maximal akzeptable Fehlerrate und den Grad menschlicher Unterstützung (Mittelstand-Digital Zentrum Berlin). Wir gehen weiter und halten fünf Bestandteile schriftlich fest.

Die fünf Bestandteile

1. Prüffälle aus echten Vorgängen

Die Prüffälle stammen aus Ihrem Alltag: echte Aufträge, Anfragen oder Rechnungen aus den letzten Monaten, mit dem richtigen Ergebnis, das ein Mitarbeiter von Hand bestätigt hat. Erfundene Beispiele sind zu glatt. Wichtig ist die Mischung. Neben den Normalfällen gehören die schwierigen hinein: unvollständige Angaben, Tippfehler, ungewöhnliche Formate.

Wie viele Fälle? Das hängt vom Ablauf ab. Wir empfehlen als Ausgangspunkt 50 bis 100 Fälle; bei 50 Fällen verschiebt ein einzelner Fehler die Quote um zwei Prozentpunkte, bei weniger Fällen sagt sie noch weniger aus. Nehmen Sie lieber mehr schwierige Fälle als mehr leichte.

Teilen Sie die Fälle in zwei Gruppen. Die erste kennt der Bauende und kann daran üben. Die zweite bleibt ihm bis zur Abnahme unbekannt. Sonst baut er auf bekannte Fälle hin, und der Agent versagt im Betrieb bei Fällen, die er nie gesehen hat.

2. Trefferquote

Legen Sie für jede Aufgabe des Agenten fest, wie viele Prüffälle er richtig bearbeiten muss. Die Zahl hängt vom Schaden eines Fehlers ab (Messbogen, Zeile F). Bei einer Vorsortierung, die ein Mensch ohnehin prüft, darf sie niedriger liegen als bei einer Aufgabe, die direkt in ein Buchungssystem schreibt. Als Ausgangspunkt dient die heutige Fehlerquote aus Zeile E: Der Agent sollte sie nicht überschreiten. Wichtig ist auch, was als „richtig" zählt. Ist ein Fall richtig, wenn alle Felder stimmen, oder reicht es, wenn die wichtigsten stimmen? Das gehört in die Vereinbarung, mit Beispielen.

3. Umgang mit Ausnahmen

Kein Agent beherrscht jeden Fall. Entscheidend ist, was er tut, wenn er unsicher ist. Ein guter Agent gibt solche Fälle an einen Menschen ab und rät nicht. Vereinbaren Sie deshalb eine zweite Messzahl: Wie viele Fälle gibt der Agent zurück, und wie viele davon hätte er selbst lösen können? Ein Agent, der alles zurückgibt, ist sicher, aber nutzlos. Einer, der nie zurückgibt, macht Fehler im Verborgenen. Legen Sie zusätzlich eine Obergrenze für Fälle fest, die er falsch und ohne Rückgabe bearbeiten darf. Bei Aufgaben mit hohem Schaden setzen Sie diese Grenze bei null.

4. Freigabe durch Menschen

Legen Sie fest, welche Schritte ein Mensch freigibt, bevor sie wirken: zum Beispiel jede Mail an Kunden, jede Buchung über einer Grenze, jede Änderung von Stammdaten. Diese Freigaben gehören in den Test. Prüfen Sie auch, ob die Person, die freigibt, die nötigen Angaben vor sich hat, um wirklich zu entscheiden, und nicht nur abnickt. Eine Freigabe, die in drei Sekunden erteilt wird, ist selten eine Prüfung.

5. Nachbesserung

Kein Agent ist nach dem ersten Tag im Betrieb fehlerfrei. Vereinbaren Sie deshalb, bis wann Abweichungen von den Kriterien kostenlos behoben werden. Die Frist beginnt mit der Abnahme, damit sie Fehler aus dem echten Betrieb abdeckt und nicht nur aus dem Test. Bei uns sind es 60 Tage ab Abnahme ohne Aufpreis, unbeschadet Ihrer gesetzlichen Rechte. Fragen Sie jede Agentur nach ihrer Frist und danach, ob sie schriftlich zugesagt wird.

Der rechtliche Rahmen: nur zur Orientierung

Das Bürgerliche Gesetzbuch kennt zwei Grundformen. Beim Dienstvertrag schuldet der eine Teil die versprochenen Dienste. Beim Werkvertrag schuldet der Unternehmer die Herstellung des versprochenen Werkes, der Besteller die vereinbarte Vergütung; Gegenstand kann auch ein durch Arbeit herbeizuführender Erfolg sein. Wer einen fertigen Agenten bestellt, will einen Erfolg und nicht bloß Arbeitsstunden. Dann ist die Abnahme der Punkt, an dem geprüft wird, ob das Werk vertragsgemäß ist. Nach § 640 BGB muss der Besteller ein vertragsmäßig hergestelltes Werk abnehmen, wegen unwesentlicher Mängel kann er die Abnahme nicht verweigern. Als abgenommen kann ein Werk auch gelten, wenn der Unternehmer eine angemessene Frist zur Abnahme gesetzt hat und der Besteller sie nicht unter Angabe mindestens eines Mangels verweigert.

Stand dieser Hinweise ist der 8. November 2026. Es ist nur der gesetzliche Rahmen, keine Rechtsberatung. Wie Ihr Vertrag einzuordnen ist und was er im Einzelnen regelt, klären Sie mit Ihrer Rechtsberatung. Ohne messbare Kriterien bleibt aber auch der beste Vertragstyp unscharf: „vertragsmäßig" lässt sich nur prüfen, wenn der Vertrag festlegt, was gemeint ist.

Arbeitshilfe: Vorlage für die Abnahmevereinbarung

Kopieren Sie die Vorlage und füllen Sie die Felder aus, bevor gebaut wird. Beide Seiten unterschreiben.

Abnahmevereinbarung für den Agenten „_____“

1. Aufgaben

- Der Agent übernimmt: _____
- Der Agent übernimmt ausdrücklich nicht: _____

2. Prüffälle

- Herkunft: echte Vorgänge aus dem Zeitraum _____ bis _____
- Anzahl bekannte Fälle: ____ · Anzahl zurückgehaltene Fälle: ____
- Anteil schwieriger Fälle (unvollständig, Tippfehler, ungewöhnliches Format): ____ %
- Das richtige Ergebnis je Fall hat bestätigt: (Name, Funktion) _____
- Aufbewahrung der Fälle: _____

3. Was als „richtig“ zählt

- Ein Fall ist richtig, wenn: _____
- Ein Fall ist falsch, wenn: _____
- Beispiele (je zwei richtige und zwei falsche Fälle): _____

4. Trefferquote

Aufgabe	Mindest-Trefferquote auf zurückgehaltenen Fällen	Höchstzahl Fehler ohne Rückgabe an Menschen
_____	____ %	____
_____	____ %	____
_____	____ %	____

5. Ausnahmen

- Der Agent gibt einen Fall an einen Menschen ab, wenn: _____
- Abgabe an (Name, Funktion, Frist): _____
- Zulässiger Anteil von Rückgaben, die er hätte lösen können: höchstens ____ %

6. Freigaben

Schritt	Freigabe durch	Angaben, die der Freigebende sieht
_____	_____	_____
_____	_____	_____

7. Abnahme

- Den Test führt durch: _____ am _____
- Die Abnahme erklärt: _____ bis spätestens _____

- Bei nicht bestandenem Test: Nachbesserung und neuer Test bis _____

8. Nachbesserung nach Abnahme

- Abweichungen von diesen Kriterien werden ohne Aufpreis behoben für ___ Tage ab Abnahme.
- Die gesetzlichen Rechte der Parteien bleiben unberührt.

9. Wiederholbarkeit

- Ein Dritter kann den Test wiederholen, weil folgende Unterlagen vorliegen: _____

Der letzte Punkt ist der Prüfstein. Wenn Sie nach dem Ausfüllen nicht sagen können, wie ein Dritter den Test wiederholen würde, ist die Vereinbarung zu vage. „Der Agent soll zuverlässig arbeiten" ist kein Kriterium. „Der Agent bearbeitet von den vereinbarten Prüffällen mindestens den vereinbarten Anteil richtig und gibt den Rest an einen Menschen ab" ist eines.

Wie wir das selbst handhaben

Bei unseren Einführungen legen wir die Abnahmekriterien gemeinsam mit Ihnen in den ersten zwei Wochen fest, noch vor dem Bau, auf Prüffällen aus Ihren echten Vorgängen. Der Agent geht erst in Betrieb, wenn er diese Prüffälle besteht. Danach gilt die Nachbesserung von 60 Tagen ab Abnahme, unbeschadet Ihrer gesetzlichen Rechte.

Modell, Hosting, Bau und Einbindung

Modell und Hosting nach Ihren Daten wählen

Ein Agent liest Ihre Mails, Aufträge und Dokumente. Deshalb gehört die Frage, wo das Modell läuft und wer die Daten sieht, zu jeder Einführung. Datenschutzanforderungen nennen 66 Prozent der KI-Anwender als Hürde, rechtliche Verunsicherung 56 Prozent und die Sorge, dass Unternehmensdaten in falsche Hände gelangen, 45 Prozent (Bitkom). Wählen Sie Modell und Hosting deshalb nach Ihren Daten, nicht nach dem Namen des Modells.

Es gibt drei Stufen:

- **Cloud in Europa.** Ein Sprachmodell läuft bei einem großen Cloud-Anbieter in einer europäischen Region. Anthropic führt Europa für Claude auf Amazon Bedrock und Google Vertex als verfügbare Region auf und beschreibt regionale Endpunkte, mit denen sich der Ort von Speicherung und Verarbeitung festlegen lässt (Anthropic, Regional Compliance, Anbieterangabe). Das passt, wenn Sie ohnehin einen Vertrag mit einem dieser Anbieter haben. Dann laufen Einkauf, Abrechnung und Sicherheitsprüfung über Verträge, die es schon gibt.

- **Offene Modelle bei einem Anbieter, der Modelle als Dienst betreibt.** Offene Modelle wie Mistral oder gpt-oss laufen bei Anbietern, die Modelle als Dienst betreiben. STACKIT beschreibt seinen Modelldienst als Umgebung für große Sprachmodelle wie Llama und Mistral (STACKIT Docs, Anbieterangabe) und betreibt seine Regionen nach eigener Angabe ausschließlich in Deutschland (eu01) und Österreich (eu02) (STACKIT Docs, Regionen, Anbieterangabe). IONOS listet in seinem Modelldienst unter anderem Mistral Small und gpt-oss-120b (IONOS, Anbieterangabe); einen Standort für diesen Dienst haben wir nicht belegt, prüfen Sie ihn im Vertrag. Diese Stufe wählen Sie bei strengen Datenvorgaben, etwa aus Kundenverträgen oder internen Regeln. Prüfen Sie Standort und Vertragspartner jeweils im Vertrag, nicht auf der Werbeseite, und welcher Dienst in welcher Region läuft. Achten Sie außerdem auf die Lizenz des einzelnen Modells.
- **Auf dem eigenen Server.** Das Modell läuft auf Hardware in Ihrem Haus, die Daten verlassen es nicht. Das ist der Weg, wenn Daten das Haus gar nicht verlassen dürfen. Der Aufwand für Betrieb und Hardware ist höher. Aus unseren Projekten können wir diese Stufe noch nicht an einem Kundenfall belegen.

Was Sie nicht erwarten sollten

Zwei Aussagen sollten Sie von keinem Anbieter als Zusage akzeptieren, auch von uns nicht:

- **Vollständige Souveränität.** Auch mit europäischer Region unterliegen US-Cloudanbieter weiter US-Recht. Das ist unsere Einschätzung und keine Rechtsberatung.
- **Gleiche Qualität auf jedem Modell.** Offene Modelle liegen bei Agentenaufgaben vermutlich zurück. Das ist eine Schlussfolgerung ohne Primärquelle. Belegen lässt es sich nur an Ihren Daten, mit dem Modellwechsel-Test.

Auch ein Gütesiegel wie „datenschutzkonform" sollten Sie misstrauisch betrachten. Ob ein Einsatz rechtmäßig ist, hängt von Ihrem Fall ab. Das Mittelstand-Digital Zentrum Berlin rät, sich die Einhaltung der Vorgaben schriftlich bestätigen zu lassen und zu fragen, wo die Modelle betrieben werden, welche Daten ein Dritter erhält und wo die Daten liegen (Mittelstand-Digital Zentrum Berlin).

Arbeitshilfe: Datenklassen und Hosting-Stufe

Ordnen Sie die Daten, die der Agent sieht, in Klassen ein. Die Zeile mit der strengsten Klasse bestimmt die Mindeststufe.

Datenklasse	Beispiel aus Ihrem Ablauf	Darf die Daten sehen: Cloud Europa	Offenes Modell bei Dienstanbieter	Eigener Server
Öffentlich	Produktkatalog, Preislisten für Kunden	ja / nein	ja / nein	ja / nein
Intern	Bestelllisten, Lagerbestände	ja / nein	ja / nein	ja / nein
Personenbezogen	Namen und Adressen von Ansprechpartnern, Bewerberdaten	ja / nein	ja / nein	ja / nein
Vertraulich	Konditionen, Verträge, Geschäftszahlen	ja / nein	ja / nein	ja / nein
Besondere Kategorien (Art. 9 DSGVO)	z. B. Gesundheitsdaten, Gewerkschaftszugehörigkeit	ja / nein	ja / nein	ja / nein

Tragen Sie je Zeile ein, ob Ihre interne Vorgabe die Stufe erlaubt. Streichen Sie, was Ihr Datenschutzbeauftragter ausschließt. Die niedrigste Stufe, die in jeder Zeile mit „ja“ endet, ist Ihre Wahl. Gibt es keine, ändern Sie den Zuschnitt des Falls: Was der Agent nicht sehen muss, soll er nicht sehen.

Das Modell später tauschen können

Die Entscheidung muss nicht endgültig sein. Dahinter steht eine Schicht, die Agenten und Modelle entkoppelt, eine Vermittlungsschicht: Der Agent spricht mit der Schicht, die Schicht mit dem Modell. Mit einer Vermittlungsschicht können Sie das Modell wechseln, ohne den Agenten neu zu bauen. Ob die Qualität nach einem Wechsel reicht, zeigt ein Test mit denselben Prüffällen auf mindestens zwei Modellen. Deshalb gehören die Prüffälle aus Kapitel 4 nicht in die Schublade, sondern in die Dokumentation. Sie sind Ihr Maßstab bei jedem Modellwechsel. Das Mittelstand-Digital Zentrum Berlin nennt als Risiko fremder Lösungen wirtschaftliche Abhängigkeiten: Änderungen der Preisstruktur, der Lizenzbedingungen oder gar die Einstellung des Produkts können unmittelbare Auswirkungen auf den laufenden Betrieb haben (Mittelstand-Digital Zentrum Berlin). Eine Vermittlungsschicht verringert diese Abhängigkeit beim Modell.

Bauen und einbinden

Ein Agent hilft erst, wenn er in Ihre Programme eingebaut ist. Ein Agent neben Ihren Systemen, in den jemand Daten kopiert, spart wenig Zeit. Bitkom nennt als großen Kostenpunkt die Integration in bestehende Systeme (41 Prozent der Anwender), und die Aufbereitung von Daten (50 Prozent) (Bitkom). Planen Sie hier den Löwenanteil der Arbeit. Rechnen Sie deshalb damit, dass die Anbindung mehr Arbeit macht als das Sprachmodell selbst.

Klären Sie vor dem Bau diese Punkte:

1. **Auslöser und Eingang.** Wie kommt der Vorgang zum Agenten: Postfach, Ordner, Schnittstelle, Zeitplan? Welche Formate kommen vor?
2. **Zugriff auf Ihre Systeme.** Welche Programme muss der Agent lesen, welche beschreiben? Welche Rechte braucht er genau? Geben Sie ihm so wenig Rechte wie möglich: Lesen, wo Lesen genügt.
3. **Stammdaten.** Wer pflegt die Daten, auf die der Agent zugreift? Welche Fehler enthalten sie heute?
4. **Prüfoberfläche.** Wo sieht Ihr Team jeden Vorschlag, bevor er wirkt, und wo erscheinen Ausnahmen? Eine Freigabe braucht einen Ort, an dem man sie sinnvoll erteilt.
5. **Protokoll.** Was hat der Agent wann mit welchem Ergebnis getan? Ohne Protokoll lassen sich Fehler nicht aufklären und Kriterien nicht belegen.
6. **Warnungen.** Fällt ein Schritt aus, muss das jemand erfahren, statt dass still etwas fehlt. Bei Wheyster.de kommt in diesem Fall eine Warnung.
7. **Ihr Team.** Wer bedient die Prüfoberfläche, wer vertritt ihn? Bitkom nennt fehlende Akzeptanz der Beschäftigten als Hemmnis für 42 Prozent der Anwender (Bitkom). Beziehen Sie die Menschen ein, die den Ablauf heute machen. Sie wissen, wo die Ausnahmen liegen.

Zwei Beispiele aus unserer Praxis zeigen, was „eingebunden“ heißt. Bei Wheyster.de schickt ein Großkunde jede Woche eine Bestellliste für viele Standorte. Daraus werden automatisch Aufträge im Shopsystem, mit Artikeln, Mengen und Lieferadresse. Am 5. Oktober waren es 64 Bestellungen aus einer Liste, alle korrekt angelegt. Bei der Intumind GmbH läuft ein WhatsApp-Support mit Wissensbasis: Ein Bot beantwortet wiederkehrende Fragen aus einer gepflegten Wissensbasis, mit Quelle, und was er nicht weiß, geht an das Team. Dazu laufen Check-ins automatisch, und zwei bestehende Coaching-Bots sind auf eine neue Plattform

umgezogen; insgesamt sind drei Systeme im Betrieb. Für Intumind haben wir keine gemessenen Zeitersparnisse, nur den Stand der Systeme (Stand: 8. November 2026). Beide Fälle zeigen: Wert entsteht dort, wo der Agent in die vorhandene Infrastruktur eingebaut ist und Ausnahmen an Menschen gehen.

Wann Kaufen, wann Bauen lassen

Nicht jeder Fall braucht einen individuellen Agenten. Gibt es ein fertiges Werkzeug, das Ihren Ablauf abdeckt, spricht vieles für den Kauf, vor allem bei Standardabläufen. Ein Bau lohnt, wenn Ihr Ablauf Besonderheiten hat oder mehrere Systeme verbunden werden müssen. Das Mittelstand-Digital Zentrum Berlin beschreibt für zugekaufte Lösungen geringe Einstiegshürden, aber neue Abhängigkeiten (Mittelstand-Digital Zentrum Berlin). Die Abnahmekriterien aus Kapitel 4 gelten für beide Wege: Auch ein gekauftes Werkzeug testen Sie an Ihren Prüffällen, bevor Sie es einführen.

Betreiben

Warum der Betrieb ein eigener Teil ist

Ein Agent ist nach der Abnahme nicht fertig, er läuft nur. Formate ändern sich, Kunden schicken neue Vorlagen, Programme werden aktualisiert, Stammdaten wachsen, Modelle werden ersetzt. Wer das nicht einplant, hat nach einigen Monaten einen Agenten, der langsam schlechter wird, ohne dass es jemand bemerkt. Das ist unsere Einschätzung, keine Studienaussage. Bei einem Ablauf, in dem ein Mensch prüft und freigibt, fällt es vermutlich schneller auf als bei einem, der still Daten schreibt.

Im Betrieb geht es um drei Aufgaben: überwachen, ändern, Modelle wechseln.

Überwachen

Legen Sie fest, woran Sie erkennen, dass der Agent gesund ist. Sinnvolle Messgrößen sind die gleichen wie in Kapitel 3 und 4:

- **Laufzeit und Menge.** Läuft der Agent zur vereinbarten Zeit, und stimmt die Zahl der bearbeiteten Vorgänge ungefähr mit der erwarteten überein? Ein Tag ohne Vorgänge ist verdächtig.
- **Rückgabequote.** Wie viele Fälle gibt der Agent an Menschen ab? Steigt der Anteil plötzlich, hat sich vermutlich etwas an den Eingängen geändert.
- **Fehler in der Stichprobe.** Prüfen Sie regelmäßig eine Stichprobe bearbeiteter Vorgänge von Hand, zum Beispiel jede Woche zwanzig Fälle. Vergleichen Sie das Ergebnis mit der vereinbarten Trefferquote.
- **Warnungen.** Jeder Ausfall eines Schritts löst eine Meldung an eine benannte Person aus, nicht an ein Postfach, das niemand liest.
- **Rückmeldungen des Teams.** Das Team, das die Prüfoberfläche nutzt, merkt Veränderungen zuerst. Geben Sie ihm einen einfachen Weg, Auffälligkeiten zu melden.

Ändern

Jede Änderung am Agenten (neuer Kunde, neues Format, neue Regel) ist ein kleiner Bau. Gehen Sie so vor, wie beim ersten Mal, nur kürzer:

1. Beschreiben Sie die Änderung und das erwartete Ergebnis in zwei Sätzen.
2. Ergänzen Sie die Prüffälle um Fälle, die die Änderung betreffen.
3. Testen Sie die Änderung und die alten Prüffälle. Das Ziel: Die Änderung hat nichts kaputt gemacht, was vorher lief.
4. Geben Sie die Änderung frei und halten Sie Datum und Verantwortlichen fest.

Ohne Schritt 3 verbessert ein Agent oft an der einen Stelle und verschlechtert an der anderen.

Modelle wechseln

Modelle werden weiterentwickelt, Preise und Bedingungen ändern sich. Nach Bitkom nennen 54 Prozent der KI-Anwender schwer zu kalkulierende Kosten als Hürde (Bitkom). Deshalb sollten Sie die Möglichkeit haben, das Modell zu wechseln, ohne den Agenten neu zu bauen (siehe Kapitel 5).

Gehen Sie bei einem Wechsel so vor:

1. Lassen Sie dieselben Prüffälle auf dem alten und dem neuen Modell laufen.
2. Vergleichen Sie Trefferquote und Rückgaben je Aufgabe. Zählt der neue Wert ebenfalls vor den Kriterien aus der Abnahmevereinbarung?
3. Prüfen Sie Standort und Vertragsbedingungen des neuen Modells, bevor Daten dorthin fließen. Wechselt ein Unterauftragnehmer oder ein Modell, sollten Sie vorab schriftlich informiert werden.
4. Schalten Sie um, wenn die Kriterien erfüllt sind, und beobachten Sie die ersten Wochen enger.

Wer betreibt und wie lange sind Sie gebunden

Klären Sie vor dem Vertrag, wer den Agenten überwacht, was der Betrieb pro Monat umfasst und ob Sie ihn kündigen können. Das Mittelstand-Digital Zentrum Berlin rät, sich neben der Anwendung auch den zugrunde liegenden Programmcode und Informationen zu den Trainingsdaten vertraglich zusichern zu lassen. Ohne das riskieren Sie spätere Einschränkungen bei Anpassung und Prüfung (Mittelstand-Digital Zentrum Berlin). Wir halten den Betrieb in unseren Verträgen getrennt vom Bau und monatlich kündbar.

Arbeitshilfe: Betriebsplan auf einer Seite

Frage	Antwort
Wer ist im Haus verantwortlich für den Agenten (Name, Vertretung)?	_____
Wer betreut ihn technisch (Name, Erreichbarkeit)?	_____
Woran sehen wir täglich, dass er läuft?	_____
Welche Stichprobe prüfen wir, wie oft, wer?	_____
Ab welcher Rückgabequote untersuchen wir?	___ %
Ab welcher Trefferquote in der Stichprobe handeln wir?	___ %
Wer erhält Warnungen?	_____
Wie melden Mitarbeiter Auffälligkeiten?	_____
Wie lange dauert eine Änderung höchstens bis zur Freigabe?	_____
Wann prüfen wir, ob ein anderes Modell besser oder günstiger ist?	_____
Wann messen wir mit dem Messbogen nach?	nach 3 Monaten, dann jährlich

Kosten trennen

Drei Posten statt einer Zahl

Rechnen Sie bei einem KI-Agenten mit drei getrennten Posten: dem einmaligen Bau, dem laufenden Betrieb und den Kosten für Modell und Hosting. Ein Angebot, das nur eine Gesamtsumme nennt, lässt Sie nicht erkennen, was Arbeit ist und was Verbrauch, und macht Anbieter schwer vergleichbar. Bitkom zeigt, dass Anwender den Aufwand woanders sehen, als viele vermuten: Als großen Kostenpunkt nennen 51 Prozent der Anwender die Infrastruktur wie Cloud und Rechenleistung, 50 Prozent die Aufbereitung von Daten und 41 Prozent die Integration in bestehende Systeme, Abos und Lizenzen für KI-Software nur 21 Prozent (Bitkom). Schwer zu kalkulierende Kosten nennen 54 Prozent der Unternehmen mit KI als Hürde, bei Unternehmen ohne KI sind es 63 Prozent (ebenfalls Bitkom).

Was in welchen Posten gehört

Bau (einmalig).

- Aufnahme des Ablaufs, Messung, Abnahmekriterien
- Entwicklung des Agenten und Anbindung an Ihre Systeme
- Prüfoberfläche, Protokoll, Warnungen
- Test gegen die Prüffälle, Abnahme

- Einweisung Ihres Teams
- Dokumentation

Betrieb (laufend).

- Überwachung, Fehlerbehebung
- Anpassung an neue Formate, Kunden und Software
- Verbesserung der Trefferquote
- Berichte, Vertretung bei Ausfall
- Personal bei Ihnen, das Ausnahmen prüft (häufig der größte versteckte Posten)

Modell und Hosting (nach Verbrauch).

- Nutzung der Sprachmodelle, meist nach Menge abgerechnet
- Hosting der Anwendung
- Gegebenenfalls Hardware beim eigenen Server

Ein seriöses Angebot weist alle drei getrennt aus. Das gilt auch für unsere Angebote. Bei Modell- und Hostingkosten sollten Sie die Weitergabe zum Selbstkostenpreis prüfen oder zumindest die Aufschläge kennen. Die Höhe hängt von Menge und Modell ab und lässt sich vorab nur schätzen. Erst die ersten Monate liefern Messwerte.

Arbeitshilfe: Rechenblatt Nutzen und Kosten

Das Blatt arbeitet mit Platzhaltern. Setzen Sie Ihre Werte aus dem Messbogen und aus den Angeboten ein.

Zeile	Rechnung	Ihr Wert
1	Zeit heute je Woche in Stunden (Messbogen) = Z	____
2	Anteil, den der Agent übernimmt, laut Abnahmekriterien (Dezimalzahl) = P	____
3	Prüfzeit Ihres Teams für die Fälle, die der Agent bearbeitet (Freigaben, Stichproben), je Woche in Stunden = R. Die Fälle, die der Agent nicht übernimmt, stecken schon im Rest (1 – P) × Z und zählen hier nicht noch einmal	____
4	Gesparte Stunden je Woche = Z × P – R	____
5	Gesparte Stunden je Jahr = Zeile 4 × W (W = Wochen mit diesem Ablauf)	____
6	Interner Stundensatz (Vollkosten) = S	____
7	Nutzen je Jahr = Zeile 5 × S	____
8	Bau einmalig = K1	____
9	Betrieb je Monat = K2, je Jahr = K2 × 12	____
10	Modell und Hosting je Monat = K3, je Jahr = K3 × 12	____
11	Kosten im ersten Jahr = K1 + K2 × 12 + K3 × 12	____
12	Kosten ab dem zweiten Jahr = K2 × 12 + K3 × 12	____
13	Ergebnis erstes Jahr = Zeile 7 – Zeile 11	____
14	Ergebnis ab zweitem Jahr = Zeile 7 – Zeile 12	____
15	Amortisation in Monaten = K1 ÷ (Zeile 7 ÷ 12 – K2 – K3)	____

Hinweise zum Ausfüllen: Rechnen Sie P vorsichtig, also mit der Mindest-Trefferquote aus der Abnahmevereinbarung, nicht mit dem Wunschwert. Vergessen Sie Zeile 3 nicht: Menschen prüfen weiter, zählen Sie aber nur die Prüfzeit für Fälle, die der Agent bearbeitet. Ist Zeile 15 negativ oder nicht berechenbar, trägt der Nutzen die laufenden Kosten nicht. Dann lohnt sich der Fall in dieser Form nicht, und das ist ein brauchbares Ergebnis. Nicht in das Blatt gehören schwer messbare Nutzen wie weniger Ärger oder schnellere Antworten. Halten Sie sie getrennt fest, rechnen Sie sie aber nicht in die Wirtschaftlichkeit ein.

Festpreis und Bedingungen

Beim Bau schulden Sie ein Ergebnis nach Kriterien, beim Betrieb sorgfältige Betreuung. Deshalb ist der Festpreis für den Bau mit schriftlichen Abnahmekriterien sinnvoll, und der Betrieb sollte monatlich kündbar

sein. Ein Preis, der erst im Projekt entsteht, ist keine Planungsgrundlage. Das Mittelstand-Digital Zentrum Berlin warnt vor versteckten Kosten für Implementierung, Lizenzen oder Wartung (Mittelstand-Digital Zentrum Berlin). Prüfen Sie im Angebot:

- Festpreis und Umfang stehen vor dem Start schriftlich fest.
- Es gibt messbare Abnahmekriterien auf vereinbarten Prüffällen.
- Eine vorausgehende Bestandsaufnahme wird bei einem Folgeauftrag angerechnet, und die Frist dafür ist genannt.
- Modell- und Hostingkosten stehen getrennt, und der Betrieb ist kündbar.
- Ihre Mitwirkung (Zugänge, Beispieldaten, Ansprechpartner, Zeit) ist beschrieben.

So erkennen Sie eine gute KI-Agentur

Eine gute KI-Agentur erkennen Sie weniger an ihren Vorführungen als an Antworten auf Fragen, die vor dem Start geklärt sein müssen: Was kostet es, wann gilt das Ergebnis als abgenommen, wer betreibt es danach, und was passiert mit Ihren Daten? Das Mittelstand-Digital Zentrum Berlin warnt vor einem häufigen Fehler: Unternehmen engagieren direkt einen Dienstleister, ohne Vorarbeit zu leisten, was oft zu Missverständnissen, ineffizienten Projekten und unnötigen Kosten führe (Mittelstand-Digital Zentrum Berlin). Mit den Kapiteln 2 bis 4 haben Sie diese Vorarbeit geleistet. Jetzt können Sie die Fragen stellen.

Arbeitshilfe: Prüfliste mit zehn Fragen

Schicken Sie die Fragen vorab per Mail und bitten Sie um schriftliche Antworten. Wer auf eine Frage nicht schriftlich antworten will, hat Ihnen damit schon etwas gesagt.

Nr.	Frage	Das ist eine gute Antwort	Antwort der Agentur	Bewertung (0–2)
1	Gibt es einen klar beschriebenen Ablauf und ein messbares Ziel?	Die Agentur beginnt mit einem konkreten Einsatzfall und misst den heutigen Stand, bevor sie baut.	____	__
2	Kenne ich den Preis, bevor etwas beginnt?	Festpreis schriftlich, mit Angabe, was nicht darin steckt.	____	__
3	Gibt es eine schriftliche Abnahme und Nachbesserung?	Prüffälle aus Ihren Vorgängen, Trefferquote, Frist für kostenlose Nachbesserung, gesetzliche Rechte bleiben unberührt.	____	__
4	Zeigt die Agentur Referenzen aus dem laufenden Betrieb?	Fälle, die im Alltag laufen; die Agentur sagt, wer die Zahlen genannt hat (Kunde, Agentur, unabhängige Messung) und wann.	____	__
5	Setzt die Agentur KI im eigenen Haus ein?	Sie nennt einen Ablauf, den sie täglich nutzt, und Fehler, die sie dabei gemacht hat.	____	__
6	Wer klärt Daten und Schnittstellen?	Die Agentur fragt im ersten Gespräch nach Ihren Systemen und der Pflege Ihrer Stammdaten.	____	__
7	Sind Datenschutz und KI-Verordnung vor dem Start geklärt?	Datenfluss-Diagramm, Liste der Unterauftragnehmer und Auftragsverarbeitungsvertrag liegen vor den ersten Daten bei Ihnen. Kein Siegel, das alles pauschal für erledigt erklärt.	____	__
8	Lernt mein Team selbst mit?	Einweisung und Schulung sind Teil des Angebots, auch ohne Programmierkenntnisse.	____	__
9	Wer betreibt es danach, und wie lange bin ich gebunden?	Betrieb getrennt vom Bau, monatlich kündbar, Modell und Hosting wählbar, Rechte an Programmcode und Dokumentation geklärt.	____	__
10	Sagt die Agentur auch, wann KI nicht hilft?	Sie rät auch einmal ab und begründet es mit Ihren Zahlen.	____	__

Auswertung: Zählen Sie die Bewertungen zusammen (0 = keine Antwort oder ausweichend, 1 = Antwort mündlich oder allgemein, 2 = Antwort schriftlich und konkret). Höchstens 20 Punkte. Vergleichen Sie danach

nicht den Preis, sondern die Antworten auf Frage 3, 7 und 9. Dort trennen sich Anbieter, die ein Ergebnis liefern, von denen, die Verantwortung übernehmen.

Zu den Fragen im Einzelnen

Zu Frage 1. Das Mittelstand-Digital Zentrum Berlin nennt einen konkreten Anwendungsfall als Grundlage für alles Weitere. Eine Agentur, die ohne messbares Ziel anfängt, kann später nicht zeigen, ob sie etwas erreicht hat.

Zu Frage 3. Der Leitfaden rät, Berichte und Abstimmungen nicht nur informell, sondern vertraglich zu regeln (Mittelstand-Digital Zentrum Berlin). Eine Zusage, die Ihre gesetzlichen Rechte unberührt lässt, ist ein gutes Zeichen. Eine Zusage, die sie einschränken will, sollten Sie von Ihrer Rechtsberatung prüfen lassen.

Zu Frage 4. Der Leitfaden rät, nach Erfahrung mit vergleichbaren Unternehmen und nach Referenzprojekten zu fragen (Mittelstand-Digital Zentrum Berlin). Entscheidend ist, ob das Gezeigte im Alltag läuft oder nur vorgeführt wurde. Auch wir nennen zu jedem Fall den Stand und wer die Zahlen genannt hat.

Zu Frage 5. Dass eine Agentur selbst mit Agenten arbeiten sollte, ist unsere Überzeugung und lässt sich nicht aus einer Studie belegen. Wer Agenten verkauft, sollte sagen können, was dabei schiefgeht.

Zu Frage 7. Misstrauen Sie Siegeln, die alles pauschal für erledigt erklären. Rechtlich zulässig ist ein Ablauf erst, wenn er für Ihren Fall geprüft ist. Fragen Sie nach Zertifikaten, aber nur als Zusatz: Das Mittelstand-Digital Zentrum Berlin nennt ISO 27001 als Hinweis darauf, dass ein Anbieter systematisch Maßnahmen zum Schutz sensibler Daten etabliert hat. Das ersetzt keine Antwort auf Ihre Fragen zu Ihrem Ablauf (Mittelstand-Digital Zentrum Berlin). Bei uns ist ein solches Zertifikat noch nicht vorhanden. Die Sicherheitsunterlagen bekommen Sie vor dem Start.

Zu Frage 9. Fragen Sie auch, ob Sie wählen dürfen, welches Modell läuft und wo, und ob sich die Agentur zum Modellwechsel-Test auf Ihren Prüffällen verpflichtet.

Wenn Sie keine Agentur brauchen

Mit diesem Leitfaden können Sie die ersten drei Schritte allein gehen: den Einsatzfall wählen (Bewertungsbogen), den Stand messen (Messbogen) und die Abnahmekriterien formulieren (Vorlage). Das lohnt sich auch dann, wenn Sie am Ende kein Projekt vergeben oder ein fertiges Werkzeug kaufen. Sie wissen danach, was Sie wollen, und können es in jedem Gespräch belegen.

Ihr nächster Schritt

1. Füllen Sie diese Woche den Bewertungsbogen aus Kapitel 2 mit Ihren Fachabteilungen aus.
2. Wählen Sie einen Kandidaten und beginnen Sie den Messbogen aus Kapitel 3. Vier Wochen Zählung brauchen Zeit, fangen Sie früh an.
3. Sammeln Sie parallel als Ausgangspunkt 50 bis 100 echte Vorgänge als Prüffälle, mit dem bestätigten richtigen Ergebnis.
4. Entwerfen Sie die Abnahmevereinbarung aus Kapitel 4 und lassen Sie Fachabteilung, Datenschutz und Einkauf darüberlesen.
5. Erst dann holen Sie Angebote ein und stellen die zehn Fragen aus Kapitel 8.

Quellen

1. Bitkom, Presseinformation vom 14.09.2026: Erstmals nutzt die Mehrheit der Unternehmen KI
   https://www.bitkom.org/Presse/Presseinformation/Erstmals-nutzt-Mehrheit-Unternehmen-KI
2. Mittelstand-Digital Zentrum Berlin: Leitfaden, So finden Sie den richtigen KI-Anbieter in Deutschland
   https://digitalzentrum-berlin.de/leitfaden-so-finden-sie-den-richtigen-ki-anbieter-in-deutschland
3. Mittelstand-Digital Zentrum Berlin: Eigenbau oder Standardlösung?
   https://digitalzentrum-berlin.de/eigenbau-oder-standardloesung-ein-leitfaden-fuer-den-ki-einstieg-im-mittelstand
4. Bürgerliches Gesetzbuch § 611 Vertragstypische Pflichten beim Dienstvertrag
   https://www.gesetze-im-internet.de/bgb/__611.html
5. Bürgerliches Gesetzbuch § 631 Vertragstypische Pflichten beim Werkvertrag
   https://www.gesetze-im-internet.de/bgb/__631.html
6. Bürgerliches Gesetzbuch § 640 Abnahme
   https://www.gesetze-im-internet.de/bgb/__640.html
7. STACKIT Docs: Regions
   https://docs.stackit.cloud/platform/regions/
8. Anthropic: Regional Compliance
   https://claude.com/regional-compliance
9. STACKIT Docs: AI Model Serving
   https://docs.stackit.cloud/products/data-and-ai/ai-model-serving/
10. IONOS: AI Model Hub
    https://cloud.ionos.com/managed/ai-model-hub

Autor: Christian Marvin Müller · Stand: 8. November 2026

Wie Sie weitermachen

KI-Potenzialanalyse: Wir messen zwei bis drei Ihrer Abläufe, rechnen den Nutzen und führen den Agenten an zehn echten Vorgängen aus Ihrem Haus vor.

https://intellagens.de/leistungen/ki-potenzialanalyse

Impressum

Apex Automate, Inhaber: Joshua-Komi Müller
Achtern Born 85, 22549 Hamburg, Deutschland

E-Mail: kontakt@intellagens.de · Telefon: +49 174 5470224
Umsatzsteuer-Identifikationsnummer: DE450174793

Intellagens ist eine Marke von Apex Automate.