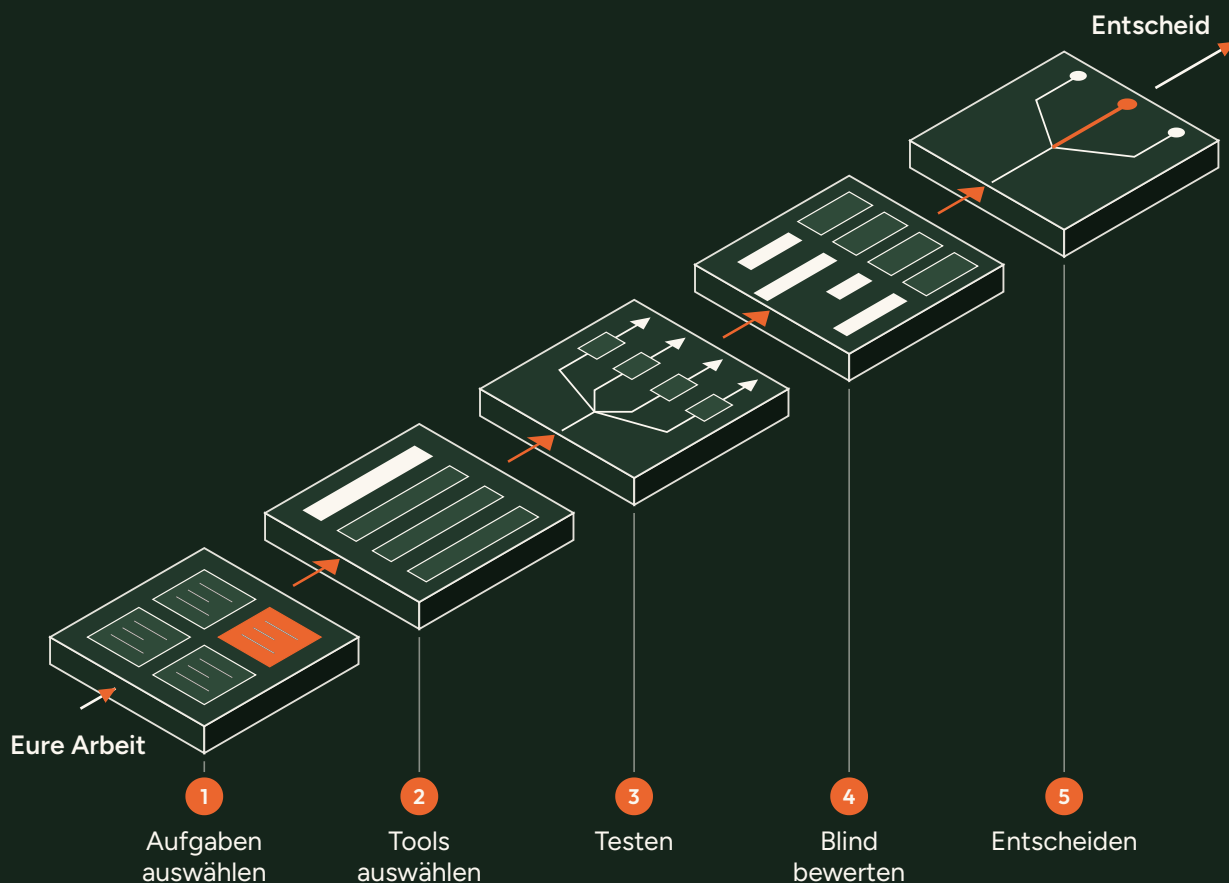


Betriebsanleitung

Das richtige KI-Tool finden

In fünf Schritten herausfinden, welches KI-Tool zu eurer Arbeit passt.



Auf einen Blick

Testet KI-Tools mit euren eigenen Aufgaben. Bewertet die Antworten, ohne zu wissen, von welchem Tool sie stammen. Entscheidet erst dann.

1	Aufgaben auswählen 4 bis 6 typische Aufgaben aus eurem Alltag	Woche 1
2	Tools auswählen euer heutiges Tool und bis zu drei neue	
3	Testen alle Tools bekommen dieselben Aufgaben	Woche 2
4	Blind bewerten Noten von 1 bis 5, ohne zu wissen, welches Tool geantwortet hat	Woche 3
5	Entscheiden wechseln, aufteilen oder bleiben	

Was ihr am Ende habt

Einen Entscheid, der auf euren eigenen Aufgaben beruht. Testaufgaben, die ihr wiederverwenden könnt. Und einen Ablauf, den ihr bei der nächsten grossen Neuerung schnell wiederholt.

Mit «Tool» meinen wir ein KI-Werkzeug wie Copilot oder ChatGPT, ein bestimmtes Modell darin oder auch nur eine Einstellung. Der Ablauf ist immer derselbe.

Was ihr braucht

Drei bis fünf Fachleute, eine neutrale Testleitung, erfundene Beispieldaten und rund drei Wochen. Technisches Wissen ist nicht nötig.

Euer Massstab ist eure eigene Arbeit.

Warum ein eigener Test?

Fast jede Woche erscheint ein neues KI-Modell, angeblich besser als alle anderen. Ob es eure Arbeit besser macht, zeigt nur ein Test mit euren eigenen Aufgaben.

Was die Hersteller zeigen

Ranglisten und Testergebnisse

Demos, die in den sozialen Medien gut aussehen

erste Eindrücke nach wenigen Tagen

Sagt wenig darüber, ob ein Tool eure Offerten, Protokolle und Mails gut schreibt.

Was euer Test zeigt

eure eigenen Aufgaben

bewertet von euren Fachleuten

ohne Namen, also ohne Vorurteile

Zeigt, welches Tool zu eurer Arbeit passt.

Viele Testaufgaben der Hersteller kennen die Modelle schon aus dem Training. Das ist, als würde man eine Prüfung schreiben, deren Fragen man vorher gesehen hat. Und bei Texten ist vieles Geschmackssache: Das Modell auf Platz 1 kann bei eurer Aufgabe schlechter abschneiden als eines weiter hinten.

Zwei Einwände, die ihr hören werdet

«Die Tools sind doch alle gut genug.»

Oft stimmt das. Aber manchmal kann ein neues Tool etwas, das bisher nicht klappte. Das merkt ihr nur, wenn ihr testet.

«Wir dürfen nur ein Tool nutzen.»

Dann vergleicht ihr die Einstellungen darin, etwa den schnellen und den gründlichen Modus. Und ein Test liefert Argumente für eine weitere Lizenz.

Eine Rangliste sagt, was ein Modell kann. Euer Test sagt, ob es zu euch passt.

Bevor ihr startet

Ihr braucht die richtigen Leute und vier klare Regeln.

Wer macht mit?

Koordination

1 Person

Plant den Test und hält alles zusammen.

Fachleute

3 bis 5 Personen

Beschreiben Aufgaben aus ihrem Alltag und bewerten später die Antworten.

Testleitung

1 neutrale Person

Gibt die Aufgaben in die Tools ein und verdeckt die Namen.

Geschäftsleitung

mit IT und Datenschutz

Klärt vorher, was erlaubt ist, und entscheidet am Schluss.

Vier Regeln

Achtung

Keine echten Daten

Keine Kunden- oder Personaldaten im Test. Nehmt erfundene Beispiele.

Achtung

Vorher klären, was erlaubt ist

Welche Tools und Daten erlaubt sind, entscheiden Geschäftsleitung, IT und Datenschutz.

Achtung

Namen verdecken

Wer weiss, von welchem Tool eine Antwort stammt, urteilt nicht mehr neutral.

Achtung

Höchstens vier Tools

Mehr lässt sich kaum sauber vergleichen.

Testet mit erfundenen Daten und verdeckten Namen.



Schritt 1 von 5

Aufgaben auswählen

Wählt 4 bis 6 Aufgaben aus eurem Alltag. Nehmt nur solche, bei denen ihr wirklich etwas ändern würdet, wenn ein Tool klar besser ist.

häufig Sitzungsprotokoll schreiben kommt jede Woche vor	heikel Reklamation beantworten ein Fehler wäre teuer	neu Offerte erstellen beginnt bei null	überarbeiten Mail verbessern baut auf Bestehendem auf	Wunschliste Wochenplan fürs Team klappte mit KI bisher nicht
---	--	--	---	--

Ein guter Mix enthält häufige und heikle Aufgaben, solche, die bei null beginnen, und solche, die etwas Bestehendes verbessern. Die Aufgabe von der Wunschliste zeigt am schnellsten, ob ein neues Tool einen echten Sprung macht.

So beschreibt ihr eine Aufgabe

Aufgabe	Offerte für einen Stammkunden
Auftrag an die KI, für alle Tools Wort für Wort gleich	«Schreibe eine Offerte für die jährliche Wartung von drei Anlagen. Nutze die Preisliste im Anhang. Höchstens eine Seite, mit Leistungen, Preis und Gültigkeit.»
Gut ist es, wenn	die Preise stimmen, alle Leistungen drin sind und der Ton passt.
Auf keinen Fall	falsche oder erfundene Preise

vage

«Schreib eine gute Offerte.»

genau

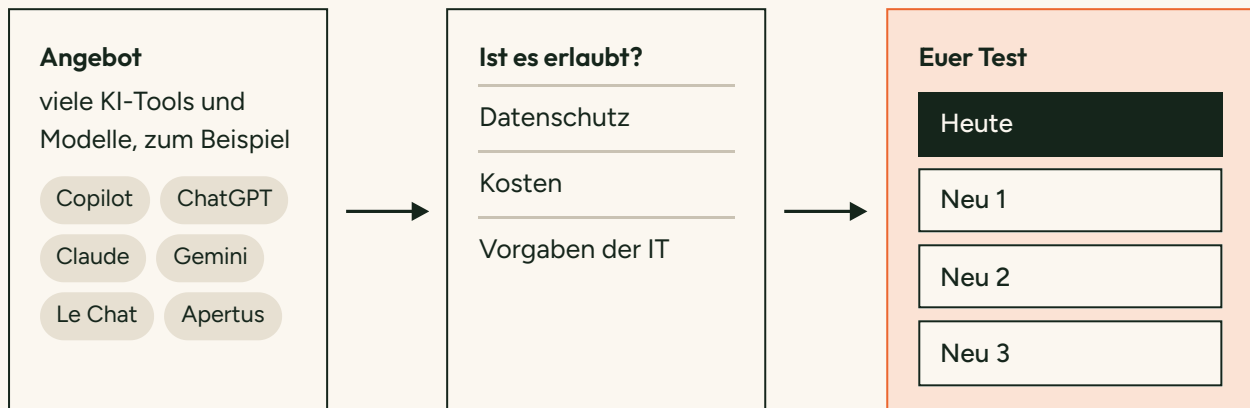
«Schreib eine Offerte für drei Anlagen, höchstens eine Seite, mit Preis und Gültigkeit.»

Je genauer der Auftrag, desto fairer der Vergleich.

2 Schritt 2 von 5

Tools auswählen

Vergleicht euer heutiges Tool mit höchstens drei neuen. Prüft vorher, was in eurer Firma erlaubt ist.



Euer heutiges Tool gehört immer dazu. Nur so seht ihr, ob ein Wechsel etwas bringt. Arbeitet ihr heute ohne KI, ist das euer Vergleichspunkt.

Fragt vorher bei IT und Datenschutz nach: Wo werden die Daten gespeichert? Werden eure Eingaben zum Training der KI verwendet? Was kostet es pro Person?

Ein Tool kann auch eine Einstellung sein, etwa der schnelle oder der gründliche Modus im selben Tool.

Beispiele, Stand Oktober 2026, keine Empfehlung. Apertus ist ein offenes Modell aus der Schweiz, das sich auch selbst betreiben lässt.

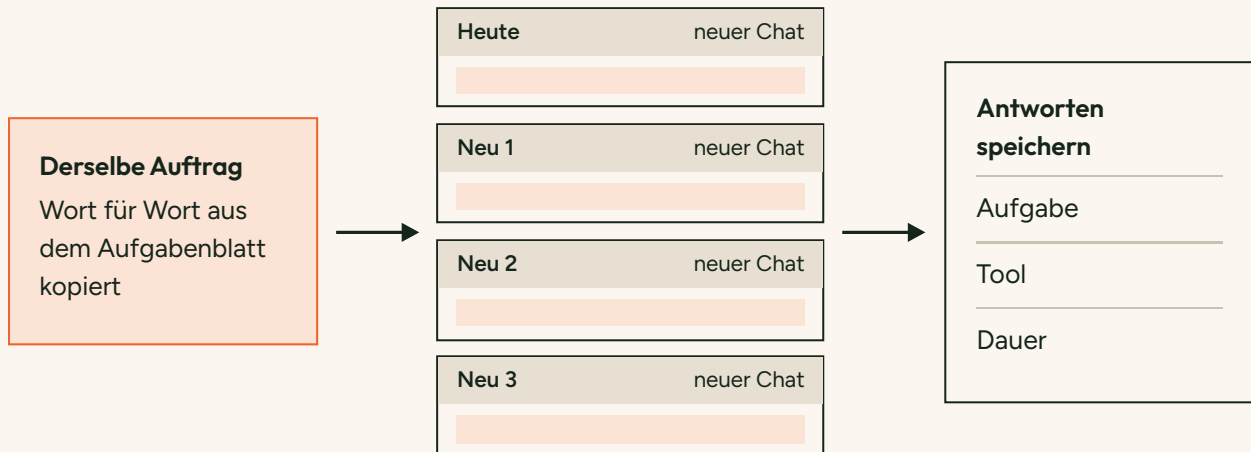
Testet nur, was ihr am Ende auch einsetzen dürft.

3

Schritt 3 von 5

Testen

Alle Tools bekommen dieselbe Aufgabe. Nur so ist der Vergleich fair.



Auftrag nicht verändern. Kopiert ihn Wort für Wort und helfst nicht nach, auch wenn eine Antwort danebenliegt.

Neuer Chat für jede Aufgabe. Schaltet auch die Gedächtnis-Funktion der Tools aus, damit frühere Gespräche nicht mitspielen.

Gleiche Einstellungen. Wählt überall den gleichen Modus, zum Beispiel überall den gründlichen.

Wichtiges mehrmals testen. Dieselbe Frage ergibt jedes Mal eine etwas andere Antwort. Lasst wichtige Aufgaben zwei- bis dreimal laufen.

Für später

Für den ersten Test reicht es, die Aufträge von Hand in die Tools zu kopieren. Wer regelmäßig testet, kann das Verteilen automatisieren lassen. Das lohnt sich erst später.

Fair ist ein Test, wenn sich nur eines unterscheidet: das Tool.

4

Schritt 4 von 5

Blind bewerten

Wie bei einer Blindverkostung: Ihr bewertet die Antworten, ohne zu wissen, von welchem Tool sie stammen.

Zuerst	Dann	Am Schluss
Namen verdecken	Bewerten	Auflösen
A Name verdeckt	A ☐ ☐ ☐ ☐ ☐ ☐ zu lang	A = Neu 1
B Name verdeckt	B ☐ ☐ ☐ ☐ ☐ fast fertig	B = Heute
C Name verdeckt	C ☐ ☐ ☐ ☐ ☐ Preis falsch	C = Neu 3
D Name verdeckt	D ☐ ☐ ☐ ☐ ☐ sehr gut	D = Neu 2

Die Testleitung mischt die Antworten und gibt ihnen Buchstaben. Welcher Buchstabe zu welchem Tool gehört, verrät sie erst am Schluss.

Die Fachleute bewerten alle Antworten zu einer Aufgabe nebeneinander: eine Note von 1 bis 5 und ein kurzer Satz dazu.

Auch das Bauchgefühl zählt. Gefällt euch eine Antwort besser, ohne dass ihr genau wisst, warum, ist das ein gültiges Urteil.

Ratet vor dem Auflösen, welche Antwort von welchem Tool stammt. Das macht Spass und zeigt, wie stark Erwartungen das Urteil prägen.

Wer mag: eine zweite Meinung durch eine KI

Lasst die Antworten zusätzlich von einer KI bewerten, die nicht im Test ist, und gebt ihr eure Kriterien mit. Liegt sie oft anders als ihr, lasst sie weg. Euer eigenes Urteil zählt mehr.

Was ihr nicht wisst, kann euer Urteil nicht verzerren.

5

Schritt 5 von 5

Entscheiden

Der Testsieger ist noch nicht der Entscheid. Wägt ab, wie viel besser ein Tool ist und was der Wechsel kostet.



Wechseln

wenn ein neues Tool bei euren wichtigsten Aufgaben deutlich besser ist und nichts dagegen spricht.

Aufteilen

wenn verschiedene Tools bei verschiedenen Aufgaben vorne liegen, zum Beispiel eines für Texte, eines für Recherchen.

Bleiben

wenn der Vorsprung klein ist oder Vorgaben dagegen sprechen. Auch das ist ein gutes Ergebnis.

Als Faustregel lohnt sich ein Wechsel erst, wenn ein Tool bei euren wichtigsten Aufgaben klar vorne liegt, im Schnitt etwa eine ganze Note.

Achtet auch auf Kosten und Tempo. Viele Abos sind heute günstig, oft weil die Anbieter einen Teil der Kosten selbst tragen. Das kann sich ändern. Und ein etwas schwächeres, aber schnelleres Tool kann im Alltag die bessere Wahl sein.

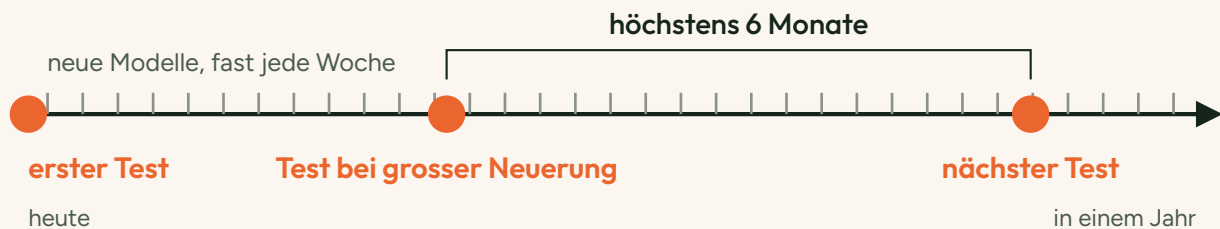
Aus der Praxis

Im Test von Nufar Gaspar, auf deren Methode diese Anleitung beruht, bekamen zwei Modelle im Schnitt dieselbe Note, obwohl das eine in der Nutzung rund zehnmal günstiger war. Und ihr Lieblingstool lag längst nicht überall vorne.

Wechselt nur, wenn es sich klar lohnt.

Dranbleiben

Ein guter Test ist eine einmalige Investition. Die Aufgaben bleiben, nur die Tools wechseln.



Nicht jedes neue Modell braucht einen Test. Wiederholt ihn, wenn ein Tool erscheint, das ihr ernsthaft in Betracht zieht, spätestens aber nach sechs Monaten. Prüft einmal im Jahr, ob eure Kriterien noch stimmen.

Wenn es hakt

Problem	Was hilft
Alle Antworten wirken gleich gut.	Eine schwierigere Aufgabe oder eine von der Wunschliste ergänzen.
Die Noten gehen weit auseinander.	Im Aufgabenblatt genauer festhalten, was ein gutes Ergebnis ist.
Ein Tool schneidet überraschend schlecht ab.	Einstellungen prüfen. Braucht die Aufgabe eure Dateien, im echten Tool mit Zugriff testen.
Das Bewerten dauert ewig.	Höchstens vier Tools und vier bis sechs Aufgaben.
Niemand traut dem Ergebnis.	Die Aufgaben mit den Fachleuten überarbeiten.

Wenn niemand dem Ergebnis traut, liegt es fast immer an den Aufgaben.

Arbeitsblatt 1

Aufgabenblatt

Pro Aufgabe ein Blatt. Am besten mit der Person ausfüllen, die diese Arbeit im Alltag macht.

Aufgabe

Bereich

Wer bewertet?

Auftrag an die KI, für alle Tools Wort für Wort gleich

Gut ist es, wenn ...

Auf keinen Fall ...

--

Art der Aufgabe

☐ häufig ☐ heikel ☐ neu ☐ überarbeiten ☐ Wunschliste

Abkürzung: Lasst euch von der KI helfen

Stelle mir bis zu acht Fragen zu unseren häufigsten Aufgaben und zu dem, was dabei schiefgehen kann. Schlage danach 4 bis 6 Testaufgaben vor, jeweils mit einem genauen Auftrag an die KI und mit Kriterien für ein gutes Ergebnis.

Arbeitsblatt 2

Bewertungsbogen

Pro Person ein Blatt. Die Antworten tragen nur Buchstaben, keine Namen. Kreuzt eine Note an und schreibt einen kurzen Satz dazu.

Name

Datum

--	--

Aufgabe	Antwort A	Antwort B	Antwort C	Antwort D
1	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
2	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
3	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
4	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
6	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Mein Tipp Wer ist wer?				

5 sofort brauchbar

4 mit kleinen Anpassungen

3 nach Überarbeitung

2 kaum brauchbar

1 unbrauchbar

Arbeitsblatt 3

Entscheid

Nach dem Auflösen gemeinsam ausfüllen, mit den Fachleuten und der Geschäftsleitung.

Ergebnis pro Aufgabe

Aufgabe	Bestes Tool	Note	Note heute
1			
2			
3			
4			
5			
6			

Was sonst noch zählt

Punkt	spricht dafür	spricht dagegen
Kosten		
Tempo		
Datenschutz und Vorgaben		
Aufwand für die Umstellung		

Unser Entscheid

☐ wechseln ☐ aufteilen ☐ bleiben

Warum?

Nächster Schritt und wer ihn macht

--

Nächster Test am

--

Ihr wollt den Test nicht allein machen?

Wir begleiten euch dabei, ganz oder nur in einzelnen Schritten.

Start

Ziel klären, Rollen verteilen, Zeitplan festlegen.

Aufgaben

Testaufgaben mit euren Fachleuten erarbeiten.

Testleitung

Tests durchführen, Namen verdecken, auflösen.

Entscheid

Workshop mit der Geschäftsleitung.

Im Strategiegelgespräch klären wir, wo ihr steht und ob eine Begleitung sinnvoll ist. 45 Minuten, online und kostenlos.

Strategiegelgespräch vereinbaren

Nach dem Entscheid: KI im Team

Ein neues Tool allein ändert noch keine Arbeitsweise. Mit unserem Angebot «KI im Team» einigt sich ein Team auf klare Regeln für KI und übt an den eigenen Aufgaben.

Zum Weitergeben

Die aktuelle Ausgabe gibt es kostenlos unter younite.consulting/impulse/ki-tools-vergleichen

younite

younite consulting GmbH
Hirschengraben 40
6003 Luzern

Kai Christen

Partner, Autor dieser Anleitung

team@younite.consulting
041 508 23 93
younite.consulting

Grundlage dieser Anleitung ist die Folge «A System for Testing New AI Models» des Podcasts «The AI Daily Brief» vom 7. Oktober 2026 mit Nufar Gaspar und Nathaniel Whittemore, auf YouTube unter dem Titel «The Best Way to Test New AI Models». younite consulting hat die Methode auf den Einsatz in Organisationen übertragen und vereinfacht. Produktnamen sind Marken ihrer jeweiligen Inhaber. Ausgabe Oktober 2026.