

Klären, bevor KI generiert

Ein Praxistest. Echte Projekte, alle Belege im Original.



“AI is jagged -

**superhuman genius
in some spots and
a dumb toddler in others.”**

Sam Altman

Claude Code | 09.09.2026:

“I need to correct something I told you earlier. I said the settings file was gitignored so the key "never reached GitHub." That was wrong.”

Don't believe the hype...



Wer hier steht: kein AI-Experte

Tagsüber

Führungsjob bei einem
Schweizer Energieversorger.

Ich bin kein Informatiker.

Feierabend

Baue seit Mai 2026 ein KI-
Tool für Solo-Selbstständige.

Alles währenddessen gelernt.

Jeden Freitag



Ein öffentlicher Bericht, was
wirklich passiert ist.

Auch die Fehlschläge.

Der Feldtest.

Annahmen

Was stillschweigend vorausgesetzt wird.

Klärungsfragen

Was gefragt wird, bevor generiert wird.

Erfolgskriterien

Woran “gut” erkennbar ist, vorab definiert.

Das Tool.

Business Clarity Check

ÜBERSICHT EN DE

5 Business Clarity Check

25 Fragen in 5 Bereichen. Antworte ehrlich.
Dauert 10-15 Minuten. Es gibt keine richtigen oder falschen Antworten.

01 POSITIONIERUNG	02 MINDSET
03 SICHTBARKEIT	04 GLAUBWÜRDIGKEIT
05 KOMMUNIKATION	

DEIN VORNAME

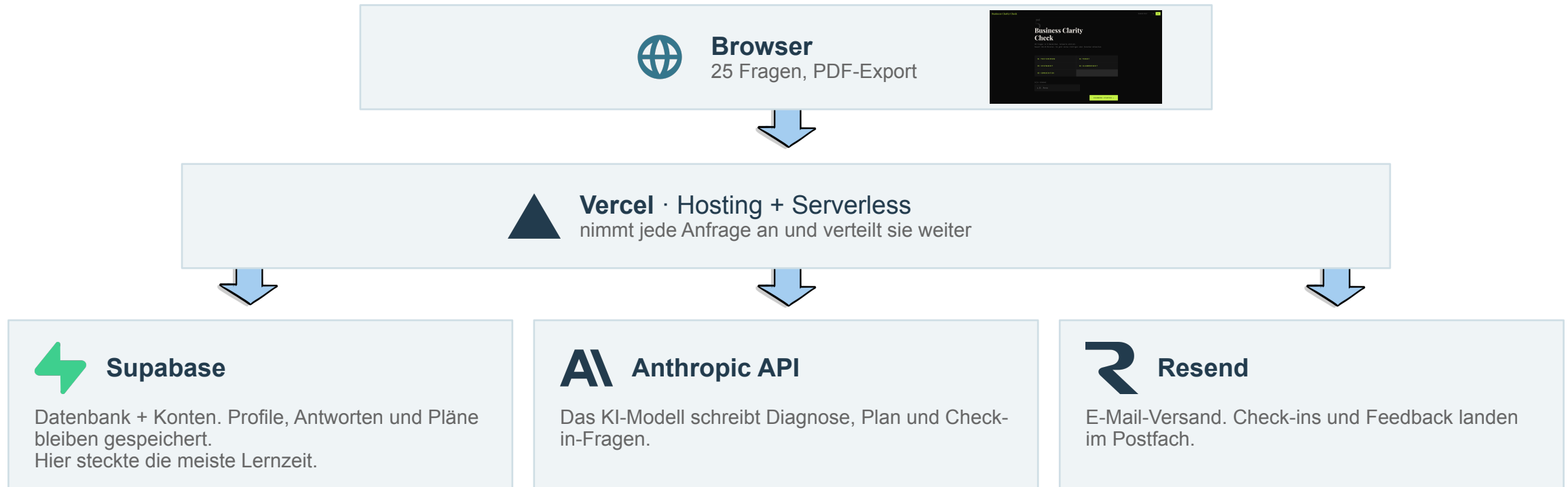
z.B. Anna

DIAGNOSE STARTEN →

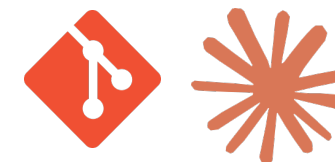
Output

- KI-Diagnose zum Status quo
- 30-Tage Aktionsplan

Was unter der Haube steckt



Vier Dienste, eine PDF-Bibliothek, zwei Bau-Werkzeuge.
Jedes davon heisst: Konto, Doku, Fehlversuche.
Mit Bordmitteln ist so ein Tool nicht getan.



*Gebaut mit git und Claude Code.
Alles nach Feierabend gelernt.*

Ein Lauf. Ein Tool. Zwei Ansagen.

Die Diagnose sagt:

„Kein LinkedIn überarbeiten.“

Der 30 Tage Plan desselben Laufs sagt:

„Tag 4: LinkedIn-Profil aktualisieren.“

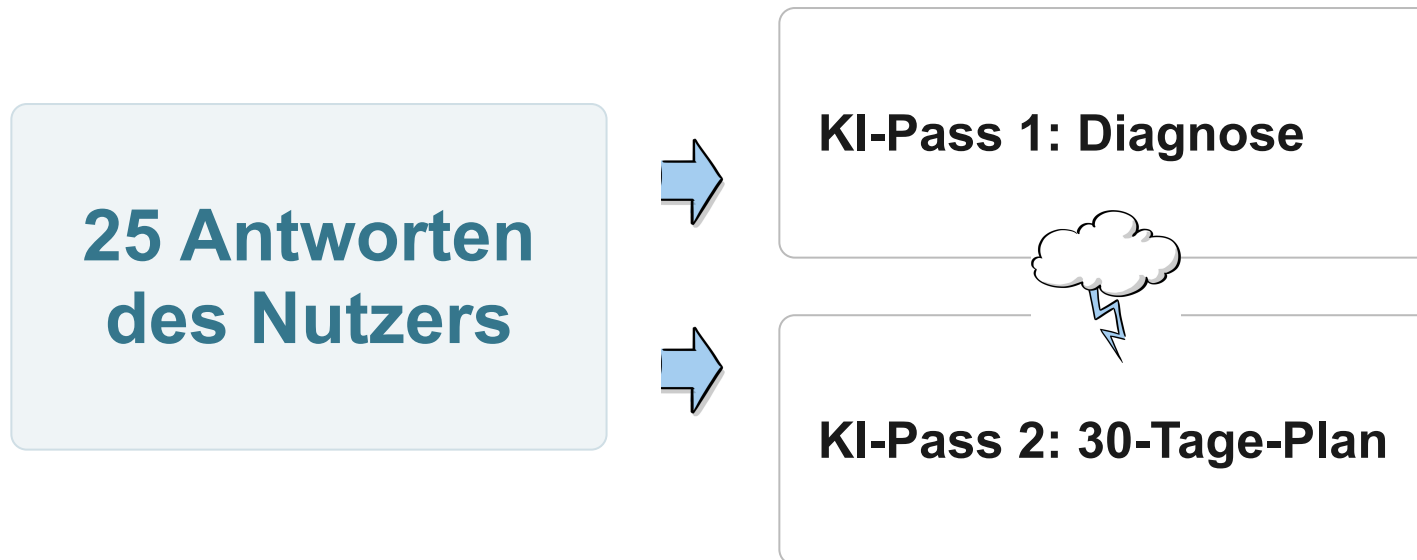
"Kein neues Profil aufbauen. Kein Content-Plan erstellen. Kein LinkedIn überarbeiten."

vs.

"Tag 4: LinkedIn-Profil aktualisieren – Passe Headline, About-Sektion und Angebot auf LinkedIn an"

BELEG · KI-Diagnose + 30-Tage-Plan, ein einziger Lauf, Business Clarity Check V1, 11.08.2026, wörtlich

Zwei Arbeiter, keine Absprache



Der Plan-Pass bekommt die Diagnose nie zu sehen.

Beide arbeiten mit denselben Rohdaten, aber blind zueinander.

Jeder Output klingt für sich souverän.

Der Widerspruch fällt keinem der beiden auf, weil keiner vom anderen weiss.

Wer füllt die Lücken in deinem Auftrag?

Ein Mensch

**kann gemeinsamen Kontext nutzen
oder nachfragen.**

Auch Menschen können Lücken übersehen
oder falsch ergänzen.

Ein KI-Modell

kann fehlende Angaben plausibel ergänzen.

Eine flüssige Antwort zeigt nicht, ob die
Ergänzung deinem Auftrag entspricht.

Alignment: wichtige Entscheidungen vor der Ausgabe klären.

Das Versuchsobjekt: ein KI-Geschäfts-Check

1 Die Idee

Solo-Selbstständige beantworten online 25 Fragen und checken sich selbst. Die KI schreibt eine Diagnose und einen 30-Tage-Plan. Kostenlos, Deutsch und Englisch.

2 Version 1

Ein starres Formular, ein Durchlauf, fertig.
Live seit Mai 2026, getestet von 6 echten Selbstständigen.

3 Gebaut

Nach Feierabend, mit KI-Hilfe, ohne Programmier-Vorwissen.
Das Bauen selbst war der Lernweg.

Ein starres Formular, sechs verschiedene Leben

6/6

Bei allen sechs Testern
passte das Formular nicht
zur Situation

Passt nicht

Neuausrichtung,
Zukunftsplan,
Kürzertreten:

25 fixe Fragen trafen
keine dieser
Ausgangslagen.

Erfindet Fakten

Die Diagnose
behauptete etwas, das
der Tester nie gesagt
hatte.

Die Lücke wurde
einfach gefüllt.

Entmutigend

30 Aufgaben an 30
Tagen.

Zu viel, zu komplex,
demotivierend statt
aktivierend.

"The form is the bottleneck, confirmed 6/6."

"AI hallucinated unstated facts. [A tester] never said he fails to use his network; the diagnosis asserted he doesn't."

BELEG · Feedback-Analyse aller 6 Tester, 05.06.2026, Original (anonymisiert)

Im Frust steckt die Produktidee

„Es hat irgendwie den Punkt verfehlt, dass ich meine USP grade komplett neu ausrichte.“

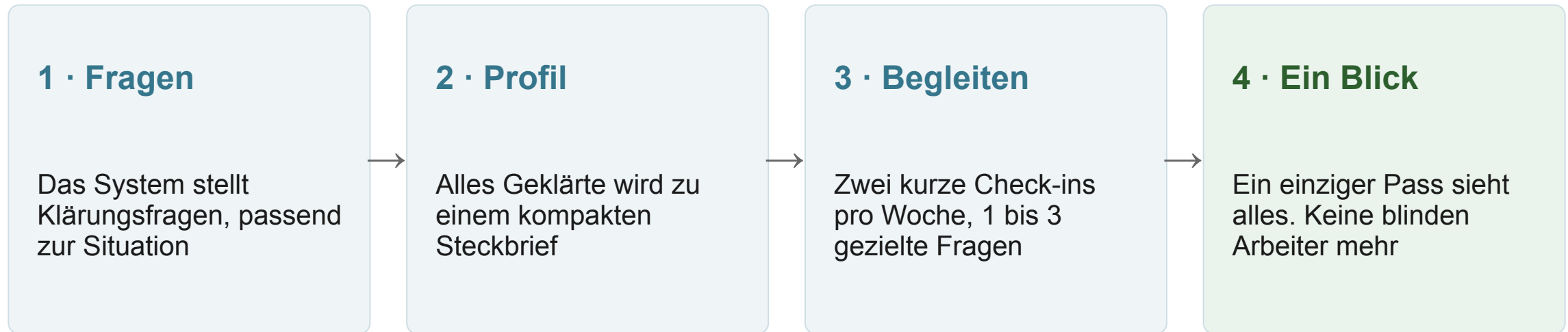
Tester, Mai 2026, Original

“Wahrscheinlich 30 Franken im Monat für ein System, dass meine Infos behält.”

Anderer Tester, Mai 2026, Original

Links der Frust, rechts der Wunsch: ein System, das den Kontext behält, statt jedes Mal blind zu raten. Genau das wurde Version 2.

Version 2: erst klären, dann urteilen



Klärung ist kein Prompt-Trick. Sie ist in den Bauplan des Tools eingelassen.

Erfinden verboten!

Jede Aussage über den Nutzer braucht eine Quelle: etwas, das er wirklich gesagt hat.

Im Zweifel bleibt ein Feld leer, statt dass geraten wird.
Fehlende Information ist nie ein Befund.

Die Regel existiert, weil die Lücke vorher einfach gefüllt wurde.

Jede Zeile im Regelwerk hat einen Insight als Ursprung.

TRUTH / NO INFERENCE: every field must be traceable to something the user actually said. Never invent facts, numbers, names, channels, or assets. Leave a field null rather than guessing. "Attends trade shows" does NOT imply "has trade-show contacts."

BELEG · Produktions-Regelwerk V2, Original, gekürzt

Erst der Massstab, dann der Ernstfall

Vor dem Start: drei erfundene Testfälle mit vorab definierten Kriterien. Erst wenn alle bestehen, geht es live.

Testfall	Situation	Ergebnis
Erfolg	Aufgaben erledigt, Momentum da: erkennt das System Fortschritt?	Bestanden, 2 Durchläufe
Stillstand	Zwei Wochen nichts passiert, Selbstzweifel: reagiert es ohne Druck und Scham?	Bestanden, 2 Durchläufe
Kurswechsel	Unerwartete Anfrage von aussen: baut es den Plan um, ohne zu erfinden?	Bestanden, 2 Durchläufe

Aber: Der Weg zum bestandenen Ergebnis waren sieben Anläufe mit vier dokumentierten Fehlern. Die kommen jetzt, und alle vier erzählen dieselbe Geschichte.

Die ignorierte Formatvorgabe

DIE LÜCKE

„Antworte nur mit Rohdaten“ stand zweimal im Auftrag. Wie ernst das gemeint war, blieb implizit.

MENSCH: MÖGLICHE PRÜFUNG

Das Format prüfen: Das nachfolgende Programm kann nur Rohdaten verarbeiten.

DIE AI HAT ...

die Antwort trotzdem dekorativ verpackt. In 3 von 3 Läufen. So nicht verwendbar -> Absturz.

DER FIX Format technisch festgelegt. Im berichteten Verlauf keine weiteren Formatfehler; keine allgemeine Garantie.

#1 Regel verschärft: "Antworte nur mit rohem JSON, keine Codeblöcke." → still 3/3 fenced.

#2 Doppelte Endserinnerung → 1/3 fenced.

#3 Format technisch erzwungen → 0 Fehler, dauerhaft. "Prompt-level 'no fences' rules fail on Haiku, tested 3/3."

BELEG · Iterationsprotokoll der Validierung, Original, gekürzt

Die übergangene Absage

DIE LÜCKE

Eine Aufgabe hatte den Status „übersprungen“. Dass das „der Nutzer hat schon Nein gesagt“ bedeutet, stand nirgends explizit.

MENSCH: MÖGLICHE PRÜFUNG

Nach einer Absage nicht erneut nachfragen.
Zum Beispiel zuerst prüfen, welche Aufgaben noch offen sind.

DIE AI HAT ...

genau das gefragt. In 3 von 3 Läufen, trotz expliziter Regel und Beispiel im Auftrag.

DER FIX Die Auswahl trifft jetzt der Code vorab. Die AI formuliert nur noch, sie entscheidet nicht mehr.

"picked a 'skipped' action in 3 of 3 runs; precomputing the list makes the selection mechanical" ·
Design-Regel: "cheap model formats, server filters."

BELEG · Produktions-Regelwerk + Iterationsprotokoll, Original

Meine eigene Mehrdeutigkeit

DIE LÜCKE

Zwei meiner Regeln sagten nie, welche von beiden zuerst gilt. Die Reihenfolge blieb offen.

MENSCH: MÖGLICHE PRÜFUNG

Die Absicht klären:
Welche Regel soll zuerst gelten?

DIE AI HAT ...

irgendeine Reihenfolge gewählt.
Formal regelkonform, faktisch falsch.

DER FIX Die Reihenfolge ist jetzt ausbuchstabiert. Die Lücke war bei mir, nicht im Modell.

"my own rules 2 and 3 didn't define their mutual order"

→ Fix: fester Ablauf "1st accountability, 2nd captures, 3rd focus" im Auftrag verankert.

BELEG · Iterationsprotokoll + Produktions-Regelwerk, Original

Die veraltete Prüfung

DIE LÜCKE

Das Produkt wechselte die Anrede von Sie zu du. Dass die automatische Prüfung noch auf Sie testet, hat niemand hinterfragt.

MENSCH: MÖGLICHE PRÜFUNG

Beim Review prüfen: Wir duzen inzwischen. Testet die Prüfroutine noch die alte Anrede?

DIE AI HAT ...

kein Problem gemeldet. Die Prüfung zeigt grün, gemessen am Kriterium von gestern.

DER FIX Erfolgskriterien altern. Die Nachprüfung ist als offene Schuld dokumentiert, nicht vergessen.

Prüfung vom 11.06.: bestanden, in Sie-Form.

Produktwechsel zu du: 14.07.

Befund 11.08.: Die Prüfroutine testet weiterhin auf Sie, aber ohne echten Bug.

BELEG · Prüfprotokoll + Code-Befund, zusammengefasst

Live-Betrieb, 18. August: die Lücke kehrt zurück

DIE LÜCKE

Das System empfiehlt eine Änderung an der Website des Nutzers. Wie diese Website aussieht, hat nie jemand gesagt.

MENSCH: MÖGLICHE PRÜFUNG

Vor der Empfehlung die Website prüfen oder nach ihrem Aufbau fragen.

DIE AI HAT ...

empfohlen: „füge es unter der Hauptüberschrift ein“. Eine Überschrift, die nie jemand erwähnt hat.

DER FIX Der Fund wurde am selben Tag als Verbotsbeispiel ins Regelwerk übernommen. Klärung ist Dauerbetrieb, kein Projekt-Setup.

"Real violation (2026-08-18): how_to said 'add it under the main headline' – presumes a headline exists. Correct: 'insert it where it fits; if nothing fits, note that and decide next session.'"

BELEG · Produktions-Regelwerk, am Fundtag ergänzt und ausgerollt, Original

Vier Lücken, eine Geschichte

- 1 Offene Angaben können unbemerkt durch plausible Annahmen ersetzt werden.
- 2 Regeln, die hart gelten sollen, werden erzwungen, nicht erbeten.
- 3 Erfolgskriterien stehen vor der Generierung fest. Und sie altern mit dem Produkt.

„Dann hättest du ja vorher alles wissen müssen“

1 Klärung gilt pro Auftrag, nicht pro Produkt

Jeder einzelne Lauf hatte Annahmen-Check und Kriterien. Was das Produkt insgesamt werden soll, wusste vorher niemand.

2 Das Produkt haben die Tester geklärt

Ohne V1 live gäbe es die 6/6-Erkenntnis nicht. „Ich weiss, was Nutzer brauchen“ ist selbst eine stillschweigende Annahme.

3 Beides zusammen ist der Weg

Kläre den Auftrag, bevor die KI generiert. Kläre das Produkt, indem du es echten Nutzern gibst.

Pause

15 Minuten. Danach: Lücken füllen, wenn ein Mensch antwortet.

Alignment mit einem Menschen

Zweites Projekt: die dokumentierte Begleitung einer Kleinunternehmerin.

Diesmal sitzt kein Modell auf der anderen Seite, sondern ein Mensch mit vollem Terminkalender. Gleiche Disziplin: Erst Zielkundin, Ausgangszahlen und Erfolgskriterien klären, dann handeln.

Aufgabe für dich (2 Minuten):

Welche Kundin soll öfter bei dir im Stuhl sitzen?

Streiche alles, was nicht passt. Kreise ein, was stimmt.

Frauen Mitte 40 bis 50: Ansatzfarbe, Glossing, Brauen alle 4 bis 5 Wochen

~~Balayage Kundinnen mit 2 bis 3 Stunden pro Termin~~

~~Brautkundinnen und Anlass-Styling~~

~~Frauen 25 bis 35, Trendfarben und Schnitt~~

~~Herren Kurzhaarschnitt im 3 Wochen-Takt~~

Gewählte Gruppe steht laut Kasse für rund 70 % vom Umsatz.

NACHBILDUNG | Original lief als WhatsApp-Streichliste. Inhalt identisch, Layout rekonstruiert.

Kassendaten, Baseline April bis Juni 2026

Drei Befunde aus dem echten Kassensystem

Anonymisiert: nur relative Werte, keine Beträge.

①

Weniger Services pro Besuch

Ø Dienstleistungen pro Damen-Besuch: 2,67 im Januar, rund 2,2 seit April.

②

Produktverkauf praktisch null

Unter 1 % vom Umsatz. Branchenüblich grob 5 bis 10 %.

③

Eine Kundengruppe trägt fast alles

Damen stehen für knapp 90 % vom Umsatz. Kapazität der Inhaberin = Engpass.

Quelle: Kassensystem-Export der Kundin. Absolute Beträge aus Vertraulichkeitsgründen entfernt.

In diesem Fall halfen konkrete Auswahlmöglichkeiten.

Produktionsaufgabe

„Beschreibe deine Wunschkundin.“

Bleibt liegen. Zu gross, zu offen, verliert gegen das Tagesgeschäft.

Reaktionsaufgabe

„Hier sind Varianten. Streiche, was nicht passt.“

Antwort kommt. Auswählen geht in Minuten.

Auswahl kann den Einstieg erleichtern. Offene Fragen bleiben sinnvoll, wenn die Optionen noch unklar sind.

Drei Anregungen aus den Fällen

1 Ausgangslage vor Massnahme

Erst Zahlen, dann Hebel. Ohne Ausgangswert ist jeder Erfolg nur eine Behauptung.

2 Erfolgskriterien vereinbaren

Vorher gemeinsam definiert. Nicht hinterher am Ergebnis entschieden.

3 Klärungsfragen gestalten

Passende Optionen anbieten. Offen fragen, wenn die Möglichkeiten noch nicht feststehen.

Zweimal dieselbe Aufgabe, live

Gleiche Notizen, gleicher Zweck, gleiches Modell. B ergänzt Kontext, Kriterien und Rückfragen.

Lauf A: Direktstart

Notizen rein, Aufgabenzeile dazu, Output raus.

Lauf B: Erst klären

Notizen und Auftrag klären:
Kontext ergänzen, Kriterien festlegen,
Rückfragen beantworten, dann schreiben.

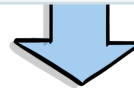
Der Fall: Notizen vom Projektmeeting | Lauf A

Getippt während des Meetings, unbearbeitet.

Projektmeeting 10.9., 14:00, Teams

dabei: Peter, Sandra, Reto, ich. Chefin entschuldigt.

- Lieferant Meier hat immer noch nicht geantwortet (3. Mal nachgefragt)
- Budget: Peter meint 40k reicht, Sandra nicht sicher, Reto sagt eher 55k
- Go-live: November wird eng, evtl. erst Q1. Nicht entschieden
- Schulungskonzept von Reto ist fertig, muss noch abgenommen werden (von wem?)
- Sandra übernimmt Kommunikation ans Team, wollte aber erst mit der Chefin reden
- Frage Budgeterhöhung: keiner wollte das der Chefin sagen, Peter macht es "nächste Woche"
- Reto: neue Tickets seit Montag, ca. 12, meist Login-Probleme
- Nächstes Meeting Do, gleiche Zeit
- Kaffeemaschine 3. Stock wieder kaputt



AUFGABENZEILE

Zusammenfassung des gestrigen Projektmeetings für die Chefin, die nicht dabei war

Die Vorlage für Lauf B

KONTEXT

Ich bin Projektleiter. Die Chefin ist meine Vorgesetzte, kennt das Projekt, war gestern nicht dabei. Sie liest die Mail auf dem Handy zwischen zwei Terminen.

AUFGABE

Zusammenfassung des gestrigen Projektmeetings für die Chefin, die nicht dabei war
-> Zusammenfassung des Meetings

KRITERIEN

max. 120 Wörter · zuerst, was sie entscheiden muss · nur, was in den Notizen steht, keine Ergänzungen · Unangenehmes bleibt unangenehm · Nebensächliches weglassen

BEVOR DU SCHREIBST

Stelle mir genau 3 Klärungsfragen, jede mit 2 bis 3 Antwortoptionen zum Auswählen. Erfinde nichts, was ich nicht beantwortet habe.

Verglichen wird das gesamte Vorgehen. Mehrere Eingaben ändern sich zugleich.

Jetzt ihr: zwei Fälle, eine Checkliste

1 · Allein · 15 Min

Teilt A und B zu. Arbeitet einzeln.

A bearbeitet Fall A, B bearbeitet Fall B.

Die Fälle stehen auf der nächsten Folie.

Beantwortet die sechs Fragen schriftlich, auf Papier und ohne KI.

Markiert eine Stelle mit einer wichtigen, bisher unausgesprochenen Annahme.

Fertig: zu jeder der sechs Fragen mindestens ein Satz.

2 · Zu zweit · 10 Min

Bildet Paare aus A und B.

Je 4 Minuten: Markierte Stelle vorlesen und die mögliche Annahme des Modells erklären. Das Gegenüber stellt genau eine Nachfrage.

Letzte 2 Minuten: Welche der beiden Annahmen hätte die teuersten Folgen?

3 · Blitzlicht · 5 Min

Je drei Freiwillige aus A und B teilen einen Satz nach dieser Vorlage.

„Meine stillschweigende Annahme war ..., beim nächsten Prompt werde ich ...“

Ein Satz reicht.

Die zwei Fälle

FALL A · PERSON A

Mail ans Team

Du leitest ein Team von 9 Personen. Ab 1. Oktober wechselt die Spesenabrechnung auf ein neues Tool. Du hast zehn Minuten und tippst in ChatGPT:

Schreib eine Mail an mein Team: ab 1. Oktober läuft die Spesenabrechnung über ein neues Tool.

Kontext:

Ich leite ein Team von 9 Personen. Ab 1. Oktober läuft die Spesenabrechnung über ein neues Tool. Das Tool heisst Expensify, der Zugang kommt per Mail von der IT, die Schulung ist ein 20-Minuten-Video. Zwei im Team reichen Spesen fast nie ein, drei reichen monatlich ein, einer davon chronisch zu spät. Was mit Spesen passiert, die bis 30.9. nicht eingereicht sind, hat mir Finance noch nicht gesagt.

FALL B · PERSON B

Absage an einen Anbieter

Du hast eine Produkt-Demo gesehen und dich für den Konkurrenten entschieden. Dein Chef will nicht, dass der Preis genannt wird. Du hast zehn Minuten und tippst in ChatGPT:

Schreib eine freundliche Absage an den Anbieter nach der Produkt-Demo.

Kontext:

Ich habe vor zwei Wochen eine Produkt-Demo gesehen, der Verkäufer war sympathisch, die Demo war gut. Wir haben uns für den Konkurrenten entschieden, 30% günstiger, mein Chef kannte ihn schon. Der Verkäufer hat zweimal nachgefasst und ein Folgegespräch vorgeschlagen. Ob wir in zwei Jahren nochmal evaluieren, ist offen.

Die sechs Fragen vor dem ersten Prompt

1 Welche Annahmen treffe ich stillschweigend?

4 Was muss ich vorher klären?
Welche Frage oder Auswahl hilft dabei?

2 Was weiss das Modell nicht, was es wissen müsste?

5 Wie sieht ein plausibles, aber *falsches* Ergebnis aus?

3 Woran erkenne ich ein gutes Ergebnis?
Kriterien vor der Ausgabe festlegen.

6 Was ist bewusst ausserhalb des Auftrags?

Fazit: Was ist geprüft, was bleibt offen?

Bisher erreicht

- Formatfehler im berichteten Testverlauf behoben
- Abgelehnte Aufgaben vorab im Code ausfiltern
- Gesprächsfolge im Auftrag festgelegt
- Quellenpflicht als Regel formuliert

Bewusst offen

- Interne Notizfelder: die weichste Stelle der Wahrheitsregel
- Die Prüfroutine testet noch das alte Kriterium, Nachprüfung geschuldet
- Geschwindigkeits-Optimierung eingebaut, Wirkung unbelegt

Regeln und bestandene Tests begrenzen Risiken. Sie belegen keine Fehlerfreiheit.

Woher kommen Kriterien, bevor es Output gibt?

1 · Aufgabe zu Ende denken

Aus den Anforderungen ableiten:

parsebar, widerspruchsfrei, nur Herleitbares, machbar neben dem Tagesgeschäft.

Der LinkedIn-Widerspruch war ab Tag 1 als Kriterium formulierbar.

2 · Bekannte KI-Fehlermodi

Geliehene Erfahrung:

Bekannte Fehler als Prüffälle nutzen: Fakten erfunden, Format verletzt, Regel übergangen. Für jeden Fall ein prüfbares Kriterium festlegen.

Deckt Lücke 1 bis 3 ab.

3 · Premortem, Frage 5

„Wie sieht ein plausibles, aber falsches Ergebnis aus?“

Drei plausible Fehlantworten entwerfen. Daraus konkrete Prüfkriterien ableiten.

Härtetest: einen guten Output selbst von Hand schreiben.

Aber:

„Passt nicht zu meiner Situation“, „30 Aufgaben entmutigen“, du statt Sie: das konnten nur echte Nutzer liefern.

**Kriterien aus dem Auftrag ableiten.
Anforderungen durchdenken (deduktiv).**

**Das Produkt mit Menschen erproben.
Aus Beobachtungen lernen (empirisch).**

Ein Testfall ist eine Klärungsfrage an dich selbst, bevor die KI generiert.

Und die wichtigste Vorarbeit für gute KI-Ergebnisse ist keine KI-Arbeit.

DANKE

Der Rest der Reise, jeden Freitag

Alles aus dieser Session stammt aus einem dokumentierten Lernweg: jeden Freitag ein Bericht, was diese Woche wirklich passiert ist. Inklusive der Wochen, in denen nichts funktioniert hat.

nottoolateforai.substack.com

Fragen jederzeit, auch nach heute.



QR: Substack öffnen

Exploriert, um zu lernen. Geklärt wird künftig vorher.

So lief es

Explorativ: fixe Idee, ständig neue Änderungen. Kriterien und Testfälle entstanden als Ergebnis von vier Fehlern.

Bewusst so gewählt, das Ziel war: KI im Bauen lernen. Dafür war der Weg richtig.

Vorschlag für den nächsten Durchlauf

Kriterien und Testfälle zuerst festlegen.
So können Fehler schon im Test auffallen.
Nicht jeder Fehler wird vorhersehbar sein.

Fast dieselben Regeln und Fixes. Der Unterschied ist die Reihenfolge.

**Auftrag und Tests gemeinsam entwickeln.
Hier als «Test Driven Design» bezeichnet.**

Rückblick: ein möglicher Ablauf mit früheren Tests

- | | | |
|----------|--|---|
| 0 | Kriterien als Testfälle | Drei erfundene Profile plus Kriterienliste existieren am Tag 1, nicht nach sieben Anläufen. |
| 1 | Format früh testen | Ein einfacher Test hätte die nicht verarbeitbaren Antworten früher zeigen können. |
| 2 | Pro Regel ein Testpaar | Positiv- und Verbotsfall. „Geht auf Messen“ darf nie zu „hat Messekontakte“ werden. |
| 3 | Schwankungen untersuchen | Mehrdeutige Regeln und Modellvariabilität prüfen; dieselben Fälle mehrfach ausführen. |
| 4 | Tests altern mit | Beim Wechsel zu du wäre der Sie-Test rot geworden. Rot schreit, fehlend schweigt. |
| 5 | Fehler als Prüffälle festhalten | Beispiel von Folie 21: eine nicht belegte Überschrift. Beim nächsten Lauf erneut prüfen. |

Rückblick: Welche Fehler waren vorher absehbar?

WISSBAR

Lücke 1 · Formatvorgabe ignoriert

Wissbar, Quelle 2: Format-Drift ist ein bekannter Fehlermodus.

WISSBAR

Lücke 2 · Absage übergangen

Wissbar, Quelle 2 und 3: ein Premortem findet den Fall „fragt Abgelehntes erneut“.

WISSBAR

Lücke 3 · Eigene Mehrdeutigkeit

Wissbar, Quelle 1: die eigenen Regeln einmal gegeneinander lesen.

NICHT VORHER

Lücke 4 · Veraltete Prüfung

Nicht ex ante: das ist Wartung. TDD hilft trotzdem, der alte Test wäre rot geworden.

NICHT VORHER

6/6 „passt nicht“, du statt Sie

Nicht ex ante: empirisch. Nur echte Nutzer konnten das liefern.

Drei von fünf Fehlerarten erscheinen rückblickend absehbar. Die Zeitersparnis wurde nicht gemessen.

Der Code, der funktionierte

DIE LÜCKE

Der Auftrag beschrieb, was das Tool tun soll. Wie oft es den Speicher fragen darf, stand nirgends. Kein Mengenkriterium, kein Kostenkriterium.

MÖGLICHE MENSCHLICHE PRÜFUNG

Bei 6 Nutzern und 34 Spieltagen fragen: «Wie viele Zugriffe sind das pro Klick?» Die Zahl vor dem Einsatz prüfen.

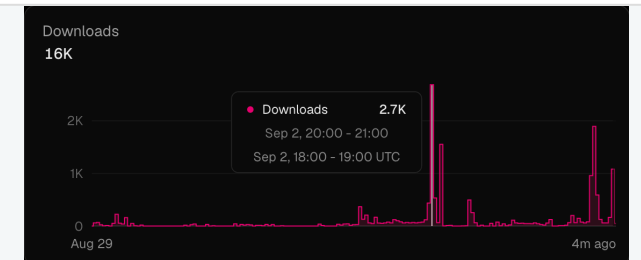
DIE AI HAT ...

funktionierenden Code geliefert, ohne Zwischenspeicher. Jeder Klick, jeder Nutzer, jeder Spieltag: neue Anfragen. In 4 Tagen fast das Monatskontingent verbraucht.

DER FIX Zwischenspeicher eingebaut und das Kriterium messbar gemacht: erster Aufruf 6 Anfragen, jeder weitere Tabwechsel 0. Vorher war die Zahl nie definiert, also nie geprüft.

Simple Operations: 9'149 von 10'000 im Monatskontingent. Spitze: 2'700 Lesezugriffe in einer Stunde, 2.9., 20 bis 21 Uhr, bei 6 Nutzern. Nach dem Fix: erster Aufruf 6 Anfragen, jeder Tabwechsel danach 0.

BELEG · Vercel Usage + Blob-Monitoring, Spieltags-Tool Bundesliga-Managerspiel, September 2026



18 Seiten Regeln vorab. Und trotzdem.

1 Auftrag geklärt, bevor der erste Prompt lief

Spieltags-Tool für 6 Nutzer: 18 Seiten Regelwerk mit Fristen, Kaderregeln, Aufstellungen, alles serverseitig erzwungen. Genau das, was der TDD-Ablauf fordert.

2 Was trotzdem schiefging

Regel-Auslegung nach Go-live (Sperrlogik, Formationswechsel) und der Speicher-Bug. Das Regelwerk klärte das Was. Wie viel und wie oft fehlten komplett.

3 Der Unterschied zum ersten Projekt

Keine erfundenen Fakten, kein Format-Absturz, keine übergangene Absage. Die Lücken lagen dort, wo das Regelwerk schwieg. Klärung verschiebt Fehler, sie beseitigt sie nicht.

Was Klärung kostet und ersparen kann

Nicht geklärt

Sieben Anläufe, vier dokumentierte Fehler.

Jede offene Lücke wurde erst gefunden, als der Output schon falsch war. Debuggen im Nachhinein.

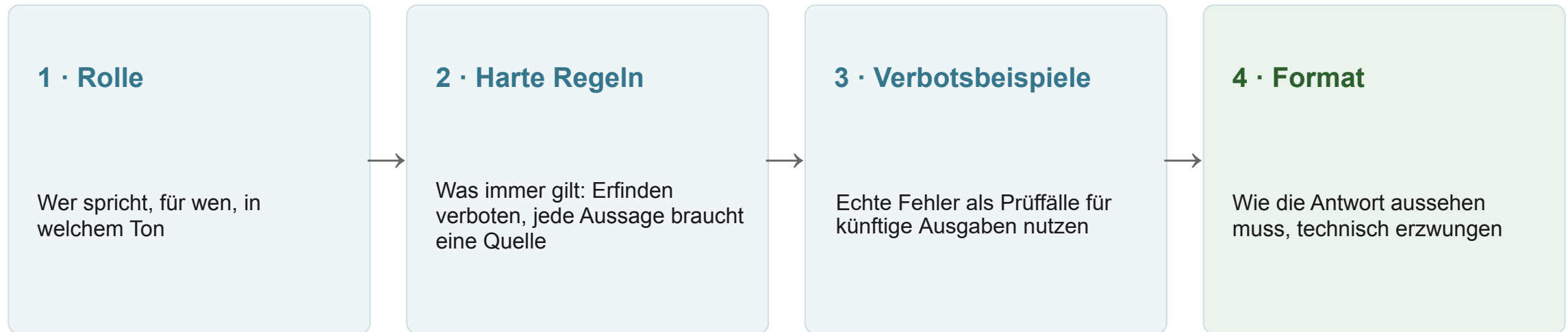
Vorab geklärt

Annahmen-Check, Kriterien, Klärungsfragen.

Braucht Vorarbeit. Kann helfen, falsche Annahmen vor der Ausgabe zu entdecken.

Vorbereitung und spätere Korrektur kosten Zeit. Den Aufwand im eigenen Fall vergleichen.

Anatomie eines Auftrags



Auftrag und Tests weiterentwickeln: aus Anforderungen und beobachteten Fehlern.