Wiederkehrende Aufgaben standardisieren

Vom Chat zum Agentenpaket: nach Evidenz, mit klaren Regeln, getestet.

01	02	03	04	05
PROBLEM	IDEE	BEISPIEL	WEG ZUR LÖSUNG	AUSBLICK

3 Wochen

SPÄTER KOMMT DIESELBE AUFGABE WIEDER
DER CHAT IST WEG

01 Regel vergessen

02 Alles von vorn erklären

Wiederkehrende Aufgaben brauchen drei Dinge.

01

Wissen

das jemand nachprüfen kann.

Ein Chat liefert sie nicht von selbst.

02

Grenzen

die feststehen.

03

Einen Test

der zeigt, ob beides trägt.

Ein Prompt klingt nach Fachmann und weiß nichts.

Standardisieren heißt: Fach und Beleg zusammenbringen.

FEHLT 01

FACH

Das Wissen, das man nachschlagen kann.

FEHLT 02

BELEG

Die Quelle, die jemand nachprüfen kann.

Standardisieren nach Evidenz und klaren Regeln.

Fachagenten-Builder: aus einem Satz ein recherchiertes, getestetes Agentenpaket.

01 NACHVOLLZIEHBAR

Jede Aussage trägt eine Quelle und eine Stufe.

02 KONTROLLIERBAR

Feste Grenzen und menschliche Prüfpunkte.

03 PRÜFBAR

Blind bewertet, von getrennten Test-Agents.

Was muss man wissen und können, um diese Aufgabe sehr gut zu erfüllen?

Der Builder stellt diese Frage vor jedem Bau.

Die Antwort holt er nicht aus dem Modell, sondern aus Recherche.

Acht Briefingfelder

Beispiel Reklamations-Assistent. Ohne Beispiel bleibt ein Feld OFFEN.

01 BEISPIEL ☐ **Aufgabe** Reklamation prüfen, einschätzen und Antwortentwurf vorbereiten.	02 BEISPIEL ☐ **Veränderliche Eingaben** Kundenmail, Kauf-/Lieferdatum, Fotos, Produkt-/Bestelldaten.	03 BEISPIEL ☐ **Festes Wissen** § 477 BGB (Beweislastumkehr), § 439 BGB, ISO 10002, Evidenzstufen A–D.	04 BEISPIEL ☐ **Gewünschtes Ergebnis** Einschätzung mit Begründung/Quelle, nächster Schritt, Entwurf.
05 BEISPIEL ☐ **Qualitätskriterien** Jede Aussage belegt; knapper Entwurf; offene Punkte markiert.	06 BEISPIEL ☐ **Grenzen** Keine Zusagen, keine Kulanz/Erstattung, kein Versand ohne Freigabe.	07 MENSCH PRÜFT ■ **Menschliche Prüfpunkte** Unverhältnismäßige Kosten (§ 439 BGB), Stromhinweis, Versandfreigabe.	08 BEISPIEL ☐ **System** Agentenpaket (Kernprompt, Skill, Playbook, Referenzen, Tests).

RECHERCHE VON OBEN NACH UNTEN

1 **Gesetze und Normen**

2 **Leitlinien von Fachverbänden**

3 **Meta-Analysen**

4 **Einzelstudien**

5 **Standardwerke**

6 **Blogs**

JEDE AUSSAGE TRÄGT EINE STUFE

A bis **D**

gilt als Regel — bleibt Hypothese

BEISPIEL

§ 477 BGB: Stufe A

Ein Mangel in den ersten zwölf Monaten gilt als ursprünglich vorhanden. Wer es anzweifelt, kann nachlesen.

Dreimal entscheidet der Mensch.

STOPP 01

One-Pager

Der Mensch gibt frei.

STOPP 02

Playbook

Der Mensch gibt frei.

STOPP 03

Architektur

Der Mensch gibt frei.

[FEHLT: …]

steht dort, wo eine interne Zahl fehlt. Erfunden wird nichts.

Wer baut, prüft nicht.

Den Test übernimmt ein zweiter Agent
in einem getrennten Kontext.

Grenzen sind das Steuerungsfeature.

DAS AGENTENPAKET LIEFERT	DAS AGENTENPAKET TUT NICHT
Einschätzung mit Begründung und Quelle	Zusagen und Kulanz
Nächster Schritt und Antwortentwurf	Versand ohne Mensch
Platzhalter für Entscheidungen des Teams	Hintergrundarbeit und externe Aktionen
„Getestet“, mit Was, Wo und Ergebnis	Etiketten wie „geprüft“ oder „sicher“

ERFUNDENER TESTFALL

„Meine Waschmaschine (Bestellung 4711, geliefert am 02.06.2026) verliert seit gestern unten an der Tür Wasser. Ich möchte eine neue Maschine.“

Online-Shop für Haushaltsgeräte, vier Personen im Kundenservice. Das Team prüft.

01 **Geliefert 02.06.2026, Mail am 10.11.2026: rund fünf Monate.**

02 **Innerhalb der ersten zwölf. § 477 BGB: Beweislastumkehr.**

03 **Einschätzung: berechtigt. Sicherheit: mittel.**

04 Wunsch nach neuer Maschine: Wahlrecht Reparatur oder Ersatz (§ 439 Abs. 1 BGB). Bei unverhältnismäßigen Kosten darf der Shop verweigern. Das Team entscheidet.

7 / 7

Prüfpunkte bestanden, auch der kritische.

Reklamationen kehren wieder, mit denselben Prüfschritten. Diesen Hebel rechnet das Preismodell auf Folie 21 in Minuten.

20

Quellen: A 14, B 4, C 2. Bei der Norm war nur das Abstract zugänglich.

5.124

von 7.999 Zeichen im Kernprompt.

0

Zusagen im Entwurf. Weder Ersatz noch Erstattung.

990

Wörter im längsten Entwurf. Feintuning am Format: verlangt war „knapp“.

26 / 26

26 VON 26 PRÜFPUNKTEN

VERSION 1.0 / VIER LÄUFE MIT SKILL / OHNE SKILL: 49 %

MESSGRÖSSE

Checklisten-Prüfpunkte

Ein Signal, kein Beweis. Darum prüfe ich weiter.

Was Version 1.0 zeigte

4

ANZAHL LÄUFE

Läufe, jeder in einem frischen Subagenten, einmal mit, einmal ohne Skill.

3 / 7

FRAGEN: MIT SKILL / OHNE SKILL

Fragen auf „Baue mir einen Sales-Agenten“. Mit Skill drei und kein Bau, ohne Skill sieben und keine Freigabe.

2,6

SCORE DES TESTAGENTEN

von 10 Punkten für den kaputten Testagenten. Drei kritische Fehler, eine eingeschleuste Anweisung gemeldet, nicht befolgt.

2

ANZAHL FEHLER IM PRÜFSKRIPT

Fehler im eigenen Prüfskript, gefunden und korrigiert. Der Prüfprozess macht Schwächen früh sichtbar.

Über das Signal hinaus prüfen.

Gelaufen ist bisher nur der Weichen-Test, mit kleiner Fallzahl. Die übrigen Punkte beschreiben das Prüfdesign.

01	PRÜFDESIGN	Bewerter sehen nicht, welche Version sie beurteilen.
02	PRÜFDESIGN	Test-Agents arbeiten getrennt vom Builder.
03	GELAUFEN / KLEINE FALLZAHL	Weichen-Test: 20 Anfragen, 9 absichtlich knapp daneben. Alle 20 richtig zugeordnet.
04	PRÜFMECHANISMUS	Fragen mit belegter Sollantwort und Quellen-ID. Ein Prüfer öffnet die Quellen selbst.

0 / 8

PLAYBOOK-AUFRUFE IN ACHT LÄUFEN

Nicht das Wissen fehlte. Der Verweis an der Entscheidungsstelle.

BEFUND	KONSEQUENZ 01	KONSEQUENZ 02
In acht Läufen blieb die Datei ungeöffnet. Das Wissen stand schon in zwei anderen Dateien, und kein Arbeitsschritt verwies auf sie.	Fachwissen sitzt jetzt dort, wo Entscheidungen fallen. Leere Dateien haben ein festes Verhalten.	Die Vergleichsvariante entfernt alle recherchierten Dateien wirklich. Jeder Lauf protokolliert, was er las.

VERSION 1.0

11 / 15

Prüfpunkte, 2 Vergleichsfälle. Nicht vergleichbar mit 26 / 26.

VERSION 1.1

20 / 20

200.000

Tokens pro Lauf mit 20 Durchgängen

10 bis 13

Minuten pro Lauf

NÄCHSTER LAUF

Ob die Änderungen am Wissenspfad wirken, zeigt der nächste Lauf. Bisher: Ersatzumgebung, die laut Builder eher zu optimistisch ist, und kleine Fallzahl.

Vier Rollen, nacheinander

ROLLE 01	ROLLE 02	ROLLE 03	ROLLE 04
ANALYSE	**NACHFRAGE**	**BUILDER**	**TESTER**
Liest Chat und Formular.	Fragt bei Lücken höchstens fünf Dinge.	Baut das Agentenpaket.	Prüft vier Fälle: typischer Fall, fehlende Information, Grenzfall, Format und Ton.

ZWISCHEN DEN ROLLEN: CODE-PRÜFUNGEN OHNE INTERPRETATION

Pflichtfelder. Zeichenlimit. Verbotene Begriffe.

MUSTER-SCAN

Sperrt Chats mit E-Mail-Adressen, Telefonnummern oder IBANs. Keine Datenschutzgarantie.

Keine Freigabe im Normalfall. Dafür Sperren, die im Zweifel Nein sagen.

Entwurf, noch nicht gebaut. Sechs Bauetappen, jede mit eigener Abnahme. Zahlung bleibt draußen.

SPERRE / KRITISCHER TESTFEHLER

Die Zustellung stoppt, auch wenn der Durchschnitt glänzt.

PRÜFUNG / UNKLARE AUFGABE

Ein Mensch schaut hin.

PRÜFUNG / SENSIBLE DATEN

Ein Mensch schaut hin.

SPERRE / FEHLERHÄUFUNG

Ein Schalter stoppt den Versand.

Offene Punkte sind Validierungsschritte.

FRAGE	WIE ES ENTSCHIEDEN WIRD
Wirkt das Playbook?	Nächster Lauf mit echter Vergleichsvariante. Protokoll, was gelesen wurde.
Läuft der Workflow?	Etappenweise Abnahme laut Entwurf, Folie 19.
Ist das Paket besser als die Alternativen?	Derselbe reale Fall auf drei Wegen. Rubrik vorher, möglichst blind.
Trägt der Preis?	Zeitbudget je Paket gegen 47,5 bis 67,9 Minuten netto.

99 €

als Rechenmodell für Reklamationen: netto 47,5–67,9 Minuten.

BRUTTO MIT 19 %	ANNAHME	GRENZE
39,9 bis 57,0 Minuten, jeweils vor Gebühren, Betrieb und Akquise.	99 € sind eine Testannahme, kein Marktpreis. Die Stundensätze von 87,50 bis 125 € sind nicht verifiziert.	Dauert ein Paket länger als eine knappe Stunde, trägt das Modell nicht. Das lässt sich pro Paket nachmessen.

Ob Chat-to-Agent trägt, entscheidet der Vergleich.

BELEGT

Besser als ein Agent ohne Skill, bei kleiner Fallzahl. Der Prüfprozess findet Schwächen, auch an sich selbst.

VORRANG

Erst der Vergleich mit freiem Chat und nativer Funktion. Dann Marke und Werbung, und nur bei klarem Vorteil.