

HUMAN APPROVAL CHECKPOINTS FÜR KI-WORKFLOWS

Du hast einen funktionierenden KI-Workflow, aber prüfst noch alles selbst. Mit Human Approval Checkpoints verschiebst du dein Urteil an die Stellen, an denen es wirklich gebraucht wird, und machst objektive Prüfungen später zu Regeln, Skills und Evals.

1 PROMPT

TOOL: CLAUDE

STAND 2026-09-08

Dein KI-Workflow läuft. Trotzdem liest du jedes Ergebnis komplett durch, suchst dieselben Fehler und gibst dieselben Korrekturen wieder und wieder. Das Problem ist nicht fehlende Automatisierung. Dir fehlt ein klares Gate an der Stelle, an der dein Urteil tatsächlich gebraucht wird.

Der bessere Ablauf: Du setzt Ziel und Standard, Claude arbeitet, ein Check prüft das Ergebnis, du entscheidest nur bei echten Ermessensfragen, Claude überarbeitet und am Ende wird noch einmal verifiziert.

WAS DU BRAUCHST

Einen bestehenden Workflow, den Claude bereits ausführen kann.

Ein klares gewünschtes Ergebnis und einen Standard, an dem du es messen kannst.

Zugriff auf die Quellen, Dateien oder Tests für objektive Prüfungen.

01 ZIEL UND STANDARD FESTLEGEN

Definiere zuerst, was am Ende richtig ist. Formuliere Ergebnis und Qualitätsstandard so konkret, dass du später zwischen einem prüfbar Fehler und einer Geschmacksentscheidung unterscheiden kannst.

Beispiel: „Jede Aussage muss durch eine echte Quelle gedeckt sein“ ist prüfbar. „Die Zusammenfassung soll nach mir klingen“ braucht dagegen dein Urteil.

PROMPT

```
Turn the workflow in this conversation into an AI approval loop. Place human judgment after any pass where truth, safety, taste, intent or an irreversible action can diverge. Use objective tests for facts, files, calculations and system state; reserve human review for consequential tradeoffs and taste. Return the workflow as: human sets outcome and standard → AI pass → check or approval → AI revision → final verification. For every checkpoint, state what evidence is required and what happens on failure. Then propose how repeated corrections become rules, examples, Skills or evals, and which single checkpoint could be removed first only after the workflow repeatedly passes.
```

01

DEN ERSTEN KI-PASS AUSFÜHREN

Lass Claude den bestehenden Workflow einmal komplett durchlaufen. Unterbrich nicht für Kleinigkeiten, die später objektiv geprüft werden können.

Der erste Pass soll Material für die Checks erzeugen: Entwurf, Quellen, Berechnungen, Dateien, Testergebnisse oder den aktuellen Systemzustand.

OBJEKTIVER CHECK ODER MENSCHLICHE ENTSCHEIDUNG?

Objektiver Check: Es gibt eine richtige Antwort, und etwas außerhalb deines Geschmacks kann sie bestätigen. Löst der Link auf? Stimmt die Zahl mit der Quelle überein? Läuft der Code?

Menschliche Entscheidung: Es gibt keine eindeutig richtige Antwort, sondern eine vertretbare. Was soll zuerst kommen? Passt der Ton? Ist der Trade-off akzeptabel?

Wenn Claude eine Geschmacksentscheidung als objektiven Check markiert, verschiebe sie zurück zu dir.

DER EHRliche TEIL

Ein guter Workflow bedeutet nicht, dass der Mensch verschwindet. Du willst ihn an die Stellen verschieben, an denen er tatsächlich einen Unterschied macht. Objektive Checks gehören langfristig in Tests, Skripte und Evals. Geschmack, Absicht und Verantwortung bleiben menschliche Arbeit.

Bei Gesundheit, Recht, Finanzen, Sicherheit, Datenschutz und irreversiblen Aktionen sollte das menschliche Gate bestehen bleiben. Ein langer fehlerfreier Lauf ist kein Beweis dafür, dass ein Agent dort zuverlässig genug ist.

01

DIE FREIGABE SETZEN

Platziere die Freigabe direkt nach dem Pass, an dem dein Urteil relevant wird. Zeige das Ergebnis, die relevanten Belege und die konkrete Frage, die entschieden werden muss.

Gut: „Sind Ton und Ask für diesen Empfänger passend?“ Schlecht: „Bitte alles prüfen.“

Wenn die Freigabe scheitert, benenne den Grund. Claude bekommt damit eine konkrete Korrektur und kann gezielt überarbeiten.

01

GEZIelt ÜBERARBEITEN

Nach einer Ablehnung soll Claude nicht den gesamten Workflow blind neu starten. Gib die konkrete Abweichung zurück und lass nur den betroffenen Teil überarbeiten.

Danach läuft derselbe Check erneut. So entsteht die Schleife: Pass, Check, Korrektur, neuer Check.

01

FINAL VERIFIZIEREN

Nach der Revision brauchst du einen letzten Check. Prüfe, ob die ursprüngliche Korrektur behoben wurde und ob dabei kein neuer Fehler entstanden ist.

Objektive Checks sollten möglichst automatisiert sein. Bei einer menschlichen Entscheidung reicht eine erneute Freigabe, wenn der Workflow tatsächlich verändert wurde.

AUS KORREKTUREN WERDEN AUTOMATISIERUNG

Wenn du denselben Fehler zweimal korrigierst, behandle ihn nicht mehr als Einzelfall. Verschiebe die Korrektur eine Stufe nach oben.

- 1. **Regel:** Schreibe eine klare Anweisung in Prompt oder Projektregeln.
- 2. **Beispiel:** Wenn die Regel wiederholt missverstanden wird, zeige ein gutes und ein schlechtes Beispiel.
- 3. **Skill:** Wenn dieselbe mehrstufige Korrektur wiederkehrt, speichere die komplette Prozedur.
- 4. **Eval:** Wenn sich das Ergebnis reproduzierbar prüfen lässt, baue einen Test daraus.
- 5. **Gate entfernen:** Entferne erst danach genau ein Gate und beobachte reale Durchläufe.

Entferne zuerst das Gate, das noch nie einen Fehler gefunden hat. Nicht das Gate, das dich am meisten nervt. Ein nerviges Gate findet oft deshalb Fehler, weil es noch nötig ist.

VIER WORKFLOWS, VIER GATES

WORKFLOW	MENSCHLICHES GATE	OBJEKTIVES GATE
Post	Absicht und Geschmack	Claims und Links
Outreach-Mail	Ton und Ask	Format und Pflichtfelder
Research	Ermessensfragen, falls vorhanden	Quellen und Belege
Code	Architektur und Ansatz	Tests, Build und erwartetes Verhalten

HÄUFIGE FRAGEN

KANN ICH OBJEKTIVE CHECKS SOFORT AUTOMATISIEREN?

Ja, wenn du eine verlässliche Prüfmethode hast. Ein Link-Check, ein Testlauf oder ein Abgleich mit einer definierten Quelle eignet sich besser als ein pauschales „sieht richtig aus“.

BRAUCHT JEDER WORKFLOW EINEN MENSCHLICHEN CHECK?

Nicht an jeder Stelle. Du brauchst ihn dort, wo Wahrheit, Sicherheit, Absicht, Geschmack oder eine irreversible Entscheidung auseinanderlaufen können.

WANN DARF ICH EIN GATE ENTFERNEN?

Erst wenn die entsprechende Prüfung über mehrere echte Durchläufe zuverlässig bestanden wurde. Entferne Gates einzeln, damit ein Fehler auf die Änderung zurückgeführt werden kann.

HINWEIS

Hinweis: Dieser Guide dient zu Informations- und Bildungszwecken. Alle Inhalte spiegeln den Stand bei Erstellung wider — Oberflächen, Funktionen und Preise der genannten Tools können sich ändern. Prüfe Ergebnisse von KI-Tools immer selbst nach. @GPTMarlon übernimmt keine Haftung für Entscheidungen auf Grundlage dieser Inhalte. Dieser Guide stellt keine Anlage-, Steuer-, Rechts- oder Gesundheitsberatung dar. Lass Ergebnisse bei Bedarf fachlich prüfen.