Healthy AI First, Folge 11: Praxis 3, Fable 5.2 auf low, gemessen statt geglaubt

Quellen und Belege zur Folge

Die dritte Praxisfolge zeigt, wie ich Claude Fable 5.2 auf der Effort-Stufe low gegen high gemessen habe, und fasst acht Prompt-Änderungen aus Anthropics Anleitung zusammen. Die einzigen eigenen Zahlen sind die Messwerte der sechs Läufe vom 17.11.2026. Alle anderen Zahlen stammen aus Anthropics Tests und sind in der Folge so gekennzeichnet.

Alle Links wurden am 17.11.2026 geprüft.

Was wir in dieser Folge erwähnen

Effort einstellen

- **Aussage:** Die Effort-Stufe bestimmt, wie viel das Modell nachdenkt, wie viele Werkzeuge es aufruft, wie lange ein Zug dauert und wie viele Token er verbraucht. Fable 5.2 kennt fünf Stufen; Standard der API ist high. Anthropic rät, bei high zu starten und an eigenen Evals hinauf oder herunter zu gehen.
  Quelle: Anthropic, Prompting Claude Fable 5.2 (Abschnitt "Calibrate effort"); Anthropic, Effort.
  Link: https://platform.claude.com/docs/en/build-with-claude/prompt-engineering/prompting-claude-fable-5-2 , https://platform.claude.com/docs/en/build-with-claude/effort
  Einordnung: Herstellerdokumentation, Stand 17.11.2026.
- **Aussage:** In der API steht die Stufe im Feld `output_config.effort`. In Claude Code gibt es den Schalter `--effort` beim Start und `/effort` in der Sitzung; die Statuszeile zeigt die aktuelle Stufe.
  Quelle: Anthropic, Effort; Claude Code, Modellkonfiguration; eigene Prüfung in Claude Code 2.1.326.
  Link: https://code.claude.com/docs/en/model-config
  Einordnung: Den Standard in Claude Code nennt die Folge bewusst nicht, weil die Quellen dazu nicht übereinstimmen.
- **Aussage:** In Anthropics Tests lag Fable 5.2 auf low nahe an dem, was der Vorgänger auf höheren Stufen erreicht hat. Ab high denkt das Modell vor fast jeder Antwort nach. Effort pro Nachricht (Beta) erhält den Prompt-Cache.
  Quelle: Anthropic, Prompting Claude Fable 5.2.
  Einordnung: Anthropics eigene Tests, nicht von uns nachgemessen.

Die Messung

- **Aussage:** Ein echter Bugfix aus unserem Kundenportal (15 Dateien einschließlich Tests), zurückgesetzt auf den Stand vor dem Fix. Je drei Läufe auf low und high mit Claude Code headless (`claude -p --model claude-fable-5-2 --effort <stufe> --output-format json`), nur Projekt-Einstellungen. Prompt: Tickettext plus ein Satz zu Tests und Typprüfung. Bewertung mit den 84 Tests des echten Fixes plus Typprüfung; vor dem Fix schlagen 5 der 84 fehl.
  Quelle: Eigene Messung (17.11.2026).
  Einordnung: Eine Aufgabe, drei Läufe je Stufe, gut beschriebenes Ticket. Vage Aufgabe und

medium nicht getestet. Die Tests prüfen Verhalten, nicht die Code-Struktur. Alle sechs Läufe liefen gleichzeitig. Kosten sind der Listenpreis der API.

- **Aussage:** Alle sechs Läufe bestanden. low im Mittel 5:28 Minuten, 4,01 US-Dollar, rund 31.900 Ausgabe-Token, 92 bis 98 Züge. high im Mittel 14:35 Minuten, 10,17 US-Dollar, rund 80.400 Ausgabe-Token, 133 bis 171 Züge.
  Quelle: Eigene Messung (17.11.2026), Einzelwerte in den Shownotes.
  Einordnung: Zeit ist die Wanduhrzeit des ganzen Aufrufs.
- **Aussage:** Bei langen Aufgaben mit vielen Schritten beendet Fable einen Zug manchmal, bevor die Arbeit fertig ist; auf low und medium häufiger.
  Quelle: Anthropic, Prompting Claude Fable 5.2 ("Finish the whole task").
  Einordnung: Unsere Läufe dauerten Minuten, nicht Stunden.

Acht Prompt-Änderungen

- **Aussage:** Fertig-machen-Text: in Anthropics Test auf medium etwa halb so viele Sitzungen mit unfertiger Arbeit, rund 30 Prozent mehr Ausgabe-Token; Enden, die auflisten, was fehlt, blieben.
  Quelle: Anthropic, Prompting Claude Fable 5.2 ("Finish the whole task").
- **Aussage:** Eine Zeile wie "answer directly without deliberating" stoppt das Nachdenken ab medium nicht mehr.
  Quelle: Anthropic, Prompting Claude Fable 5.2 ("Calibrate effort", "Review prompts and harnesses designed for earlier models").
- **Aussage:** Eine Freigabe-Regel für "sending messages to people outside this conversation" ließ das Modell in simulierten Aufgaben etwa die Hälfte seiner Routine-Anfragen zurückhalten.
  Quelle: Anthropic, Prompting Claude Fable 5.2 ("Approvals").
- **Aussage:** Eigentümer-Satz: doppelter Code zusammengeführt in etwa 75 statt unter 10 Prozent der Läufe, rund 70 bis 80 Prozent mehr Ausgabe-Token, dazu ungefragte Änderungen.
  Quelle: Anthropic, Prompting Claude Fable 5.2 ("Say who owns the code").
  Einordnung: Anthropics Test mit kleinen Coding-Aufgaben.
- **Aussage:** Zwischenmeldungen kommen als Thinking-Blöcke, in der Standardeinstellung leer; `thinking.display` auf `"updates"` (Beta) oder `"summarized"`. Mit dem Satz im System-Prompt eine Nachricht vor dem ersten Werkzeugaufruf in jedem Lauf, rund 20 Prozent mehr Ausgabe-Token.
  Quelle: Anthropic, Prompting Claude Fable 5.2 ("Keep users informed during long turns").
- **Aussage:** Wie viel Fable delegiert, hängt vom Harness ab. In Anthropics Test auf high halbierte eine längere, kostenbezogene Werkzeugbeschreibung die Zahl der Subagenten ungefähr.
  Quelle: Anthropic, Prompting Claude Fable 5.2 ("Set expectations for subagents").
- **Aussage:** Vorsicht häuft sich an; "nie" gilt als absolut; ein allgemeines "sei freier" ist zu vage. Harte Stilregeln machen Berichte abgehackt.
  Quelle: Anthropic, Prompting Claude Fable 5.2 ("Keep caution from piling up", "Make reports easy to follow").

Der Harness

- **Aussage:** Änderungen an System-Prompt oder Werkzeugen in einem fortgesetzten Gespräch können am Prefix-Check scheitern (Fehler 400). Harness-Code, der frühere Züge ändert, macht

Thinking-Blöcke ungültig. Ein im Prompt gezeigtes Token-Budget verändert den Verbrauch; in Anthropics Test auf high und xhigh senkte das Entfernen eines großen Budgets die Kosten bei gleichwertiger Leistung.
Quelle: Anthropic, Prompting Claude Fable 5.2 (Hinweisblock, "Review prompts and harnesses designed for earlier models").

Audit

- **Aussage:** `/claude-api prompt-audit` prüft Prompts, Skills, CLAUDE.md und Agenten-Dateien auf Text für ältere Modelle und liefert einen Bericht mit Datei, Zeile und Diff-Vorschlag, ohne etwas zu ändern.
  Quelle: Werkzeug aus Claude Code.
  Einordnung: Nicht Teil von Anthropics Prompting-Anleitung.

Leitfaden

- Unser Leitfaden mit allen Prompt-Texten: https://www.linkedin.com/feed/update/urn:li:activity:7528489810681044992/

Healthy AI First ist KI-recherchiert und KI-generiert, gesprochen von einer KI-Version meiner Stimme und von mir (Björn) geprüft.