Healthy AI First, Folge 12: Praxis 4, Das Prompting Playbook, mit und ohne Code

Quellen und Belege zur Folge

Die vierte Praxisfolge erzählt den öffentlichen Vortrag "The prompting playbook" von Margot van Laar (Anthropic) nach und überträgt die Schritte auf ein eigenes Beispiel ohne Code. Eigene Messungen enthält die Folge nicht. Alle Ergebnisse stammen aus dem Vortrag und sind in der Folge so berichtet. Grundlage ist unsere eigene Abschrift der Aufzeichnung.

Alle Links wurden am 18.11.2026 geprüft.

Die Quelle

- **Vortrag:** "The prompting playbook", Margot van Laar, Code with Claude London, 19. Mai 2026.
  Link: https://www.youtube.com/watch?v=G2B0YWuJUgI
  Einordnung: Öffentliche Aufzeichnung von Anthropic. Im Vortrag stellt sich die Sprecherin als Applied AI Engineer bei Anthropic in London vor.
- **Session-Seite:** https://claude.com/code-with-claude/session/ldn-the-prompting-playbook
  Einordnung: Nennt Titel, Sprecherin, Ort und Datum.

Was wir in dieser Folge erwähnen

Teil 1: bestehender Prompt nach einem Modellwechsel

- **Aussage:** Das Beispiel ist ein Support-Bot für "Meridian Mobile", verkleinert und an echte Prompts angelehnt. Viele haben am Prompt mitgeschrieben, niemand ist klar zuständig, Policy, Tonfall, Prozesse und Flicken für frühere Modelle sind gemischt.
  Quelle: Vortrag.
- **Aussage:** Nach einem Modellwechsel gibt es zwei Ursachen für schlechtere Ergebnisse: Das Modell ist fähig, verhält sich aber anders, oder es ist weniger fähig. Um das zu unterscheiden, braucht es ein Eval. Im Vortrag fünf Testfälle in drei Arten: Kontrollfall, Grenzfälle, Übergabe oder Ablehnung.
  Quelle: Vortrag.
  Einordnung: Die Sprecherin sagt selbst, dass ein echtes Eval viel mehr Fälle hat.
- **Aussage:** Aufräumen (ehrliche Rolle, kopierter Website-Text entfernt, XML-Tags für Rolle, Richtlinien, Policy, Tonfall) verbessert einen Fall; ein anderer schwankt zwischen den Läufen. Danach bestehen zwei Fälle durchgehend, drei Fehlerarten bleiben.
  Quelle: Vortrag.
- **Aussage:** Ausgabeformat im Prompt, Stop-Sequenz auf den schließenden Tag im API-Aufruf, Structured Outputs für komplexe Schemas. Eine Verbesserung im Eval erwartet die Sprecherin davon hier nicht.
  Quelle: Vortrag.

- **Aussage:** Hotspot-Fall: Die Kundendaten enthalten 5 GB, der Bot nennt die 4 GB des aktuellen Tarifs und verweist auf das Kundenkonto. Ursache ist die Zeile "Never give a customer the wrong plan details. Instead, point them to the URL." Nach der neuen Fassung bestehen alle Läufe dieses Falls.
  Quelle: Vortrag.
  Einordnung: Dass die Zeile ein Flicken für ein früheres Modell war, ist eine Vermutung der Sprecherin.
- **Aussage:** Abrechnungsfehler: Die Regel nennt nur die Kosten der Übergabe (rund 8 Dollar). Die neue Fassung nennt auch die Kosten einer falschen Antwort.
  Quelle: Vortrag.
  Einordnung: Die Folge stellt diesen Fall vor den Werkzeug-Fall; im Vortrag kommt er als letzter. Nach allen Korrekturen sagt die Sprecherin, dass alle Evals bestehen "sollten". Ein bestätigter Endlauf ist in der Aufzeichnung nicht zu hören.
- **Aussage:** Anteilige Rechnung: "Critical, always calculate any prorated amounts correctly." hilft nicht. Mit einem Tool (Zeile im Prompt, Schema im API-Aufruf, Implementierung) bestehen alle Läufe dieses Falls. "Instructions don't add capability."
  Quelle: Vortrag.

Teil 2: neuer Agent

- **Aussage:** Dienstplan für acht Beschäftigte mit harten Regeln, bewertet von einer Python-Funktion, fünf Durchläufe je Aufbau. Sonnet 4.6 mit einfachem Prompt: alle scheitern. Opus 4.7: alle scheitern, deutlich weniger Verstöße. Opus 4.7 mit Adaptive Thinking: offenbar verlässlich, im Wesentlichen dreifache Token und Latenz, genannt werden 100 Sekunden. Sonnet 4.6 mit besserem Prompt: zwei von fünf, der Rest erreicht das Ausgabelimit.
  Quelle: Vortrag.
  Einordnung: Modelle und Zahlen sind der Stand des Vortrags vom Mai 2026, kein aktueller Vergleich. Genaue Token- und Sekundenwerte nennt die Folge nicht.
- **Aussage:** Generate, Evaluate, Repair mit drei einfachen Prompts löst alle Durchläufe, mit deutlich weniger Token und geringerer Latenz als Sonnet 4.6 mit besserem Prompt. Weiche Anforderungen können zur Laufzeit in den Prüf-Prompt.
  Quelle: Vortrag.

Unsere Beispiele

- **Aussage:** Angebots-Assistent mit gewachsenen Projekt-Anweisungen, fünf Testfragen, "Nenne nie Preise", Reklamationen, Summe per Code, Prüfrunde in einem neuen Chat.
  Quelle: Eigene, erfundene Beispiele.
  Einordnung: Übertragungen der Schritte aus dem Vortrag. Nicht gemessen, keine Aussage von Anthropic.

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