

■ // ANTHROPIC COMMERCE AGENTS · DER AUFBAU

Der Shop-Agenten- Bauplan. zum Abschreiben zerlegt

Anthropic hat am 1. September ein Referenz-Repo für zwei Shop-Agenten veröffentlicht: einen für die Kundenseite, einen fürs Backoffice. Dieser Aufbau zerlegt die Architektur in ihre fünf Bausteine — Prompt, Skills, Tool-Verträge, Gates, Freigabe — und zeigt, wie du deinen eigenen Shopprozess in dasselbe Muster überträgst.

2 AGENTEN-ROLLEN

5 BAUSTEINE

FREIGABE-GATES

4 BRANCHEN-BEISPIELE

QUELLEN INKLUSIVE

Worum es geht: `anthropics/commerce-agents` ist eine Referenzimplementierung, keine fertige Plattform: alle Marken und Daten sind fiktiv, echte Bestellungen führt die Demo nicht aus, und aktiv gepflegt wird das Repo nicht. Wertvoll ist es trotzdem — es zeigt **die komplette Architektur zum Abschreiben**, inklusive der Stelle, an der ein Mensch freigeben muss. Am 4. September stand es bei 1.866 Sternen.

1 Zwei Agenten statt einem — und warum

Ein einzelner Agent, der beraten, Preise ändern und Bestellungen anfragen darf, macht jeden Fehler teuer. Das Repo trennt deshalb zwei Rollen mit verschiedenen Rechten:

ROLLE	ARBEITET FÜR	TYPISCHE AUFGABEN
Shopping-Agent	deine Kunden	Fragen beantworten, Produkte vergleichen, Bestände prüfen, Warenkorb vorbereiten
Merchant-Agent	dein Backoffice	Bestände auswerten, Kampagnen entwerfen, Preis- und Sortimentsänderungen vorbereiten

Der Shopping-Agent braucht nie Schreibrechte auf Preise, der Merchant-Agent redet nie mit Kunden. Diese Trennung ist die erste Entscheidung, die du übernehmen solltest — noch vor jeder Technik.

2 Die fünf Bausteine der Architektur

Jeder Agent im Repo besteht aus denselben fünf Teilen. Zusammen sind sie „der Aufbau“:

- 1 **Prompt** — die Rolle und ihre Grenzen in Textform: wofür der Agent zuständig ist, was er nie tut, wie er mit Unsicherheit umgeht.
- 2 **Skills** — wiederverwendbare Fähigkeits-Definitionen, etwa „Produktvergleich“ oder „Kampagnenentwurf“, die der Agent je nach Anfrage zieht.
- 3 **Tool-Verträge** — die exakte Liste, welche Funktionen der Agent mit welchen Parametern aufrufen darf. Was nicht im Vertrag steht, kann er nicht anfragen.
- 4 **Gates** — Bedingungen zwischen Absicht und Ausführung: Lese-Aktionen laufen durch, schreibende bleiben stehen.
- 5 **Menschliche Freigabe** — die letzte Stufe hinter jedem Gate für schreibende Änderungen. In der Demo bleibt jede vorbereitete Änderung stehen, bis ein Mensch sie bestätigt.

Der Clou daran: Diese Definition wird **einmal** geschrieben und läuft dann unverändert über drei Laufzeitwege von Anthropic — Messages API, Agent SDK und Managed Agents. Du entscheidest die Infrastrukturfrage später, nicht am Anfang.

3 Die Sicherheits-Mechanik im Detail

Die Safety-Dokumentation des Repos zieht zwei harte Linien, die du wörtlich übernehmen kannst:

- **Schreiben wartet immer.** Preisänderung, Sortimentsänderung, Bestellaktion — der Agent bereitet sie vor und legt sie zur Freigabe vor. Ausgeführt wird erst nach menschlicher Bestätigung.
- **Die Demo bestellt nie echt.** Der komplette Bestellpfad endet vor der Ausführung — auch das ist Absicht, nicht Unfertigkeit.

WARUM DAS DER WICHTIGSTE TEIL IST

Ein Agent, der berät, kostet dich bei einem Fehler ein unzufriedenes Gespräch. Ein Agent, der ungefragt Preise schreibt, kostet dich echtes Geld. Die Freigabe-Stufe ist der Unterschied zwischen „Assistent“ und „Risiko“ — und sie ist im Repo fertig vorgedacht.

4 Deinen eigenen Prozess in das Muster übertragen

So überträgst du einen bestehenden Shopprozess in den Aufbau — am Beispiel „Preisänderung“:

- 1 **Prozess wählen und in Schritte zerlegen.** Preisvorschlag berechnen, Auswirkung prüfen, Preis schreiben.
- 2 **Lesen und Schreiben trennen.** Berechnen und Prüfen sind Lese-Schritte, das Schreiben des Preises ist der eine kritische Schritt.
- 3 **Tool-Vertrag aufsetzen:** Der Agent darf `get_product`, `get_sales_stats` und `propose_price_change` aufrufen — ein direktes `set_price` existiert für ihn nicht.
- 4 **Gate mit Freigabe dahinter:** `propose_price_change` erzeugt einen Vorschlag in einer Warteschlange, die du abarbeitest. Erst deine Bestätigung schreibt.
- 5 **Prompt und Skill aus dem Branchen-Beispiel abschreiben,** das deinem Shop am nächsten kommt — vier fiktive Branchen liegen im Repo — und Produktlogik, Ton und Grenzen anpassen.

5 Quellen

- **Repository:** github.com/anthropics/commerce-agents
- **Safety-Dokumentation:** github.com/anthropics/commerce-agents/blob/main/docs/safety.md

■ // AUF EINEN BLICK

Die Kurzfassung

Vier Punkte, bevor du baust.

- ✓ **Zwei Rollen, getrennte Rechte** — der Kunden-Agent schreibt nie Preise
- ✓ Fünf Bausteine: **Prompt, Skills, Tool-Verträge, Gates, Freigabe** — einmal definiert, drei Laufzeitwege
- ✓ Schreibende Aktionen laufen **immer über menschliche Freigabe** — das ist der Kern des Bauplans
- ✓ Referenz, keine Plattform: **fiktive Daten, keine echten Bestellungen**, nicht aktiv gepflegt

■ // BLEIB DRAN

Agenten, die kontrollierbar bleiben.

Ich zerlege solche Referenzen in Bausteine, die du für dein eigenes Geschäft übernehmen kannst — inklusive der Stellen, an denen ein Mensch das letzte Wort behält. Folg mir für den nächsten Aufbau.

@ki.borish

Fakten verifiziert am 04.09.2026 aus dem Repository und der Safety-Dokumentation von anthropics/commerce-agents; Sterne-Zahl ist ein Snapshot vom 04.09.2026, 16:30 UTC. Das Repo ist eine Referenzimplementierung mit fiktiven Marken und Daten — ein geschäftlicher Erfolg damit gebauter Agenten ist nicht belegt. Funktions- und Beispielnamen in Abschnitt 4 sind zur Veranschaulichung gewählt; die konkreten Verträge stehen im Repo.