

KI-Agenten verstehen

Vom Chatbot zum autonomen Assistenten

Ein praxisorientierter Leitfaden für Entscheider und Teams

Reto Grau Consulting GmbH — März 2026

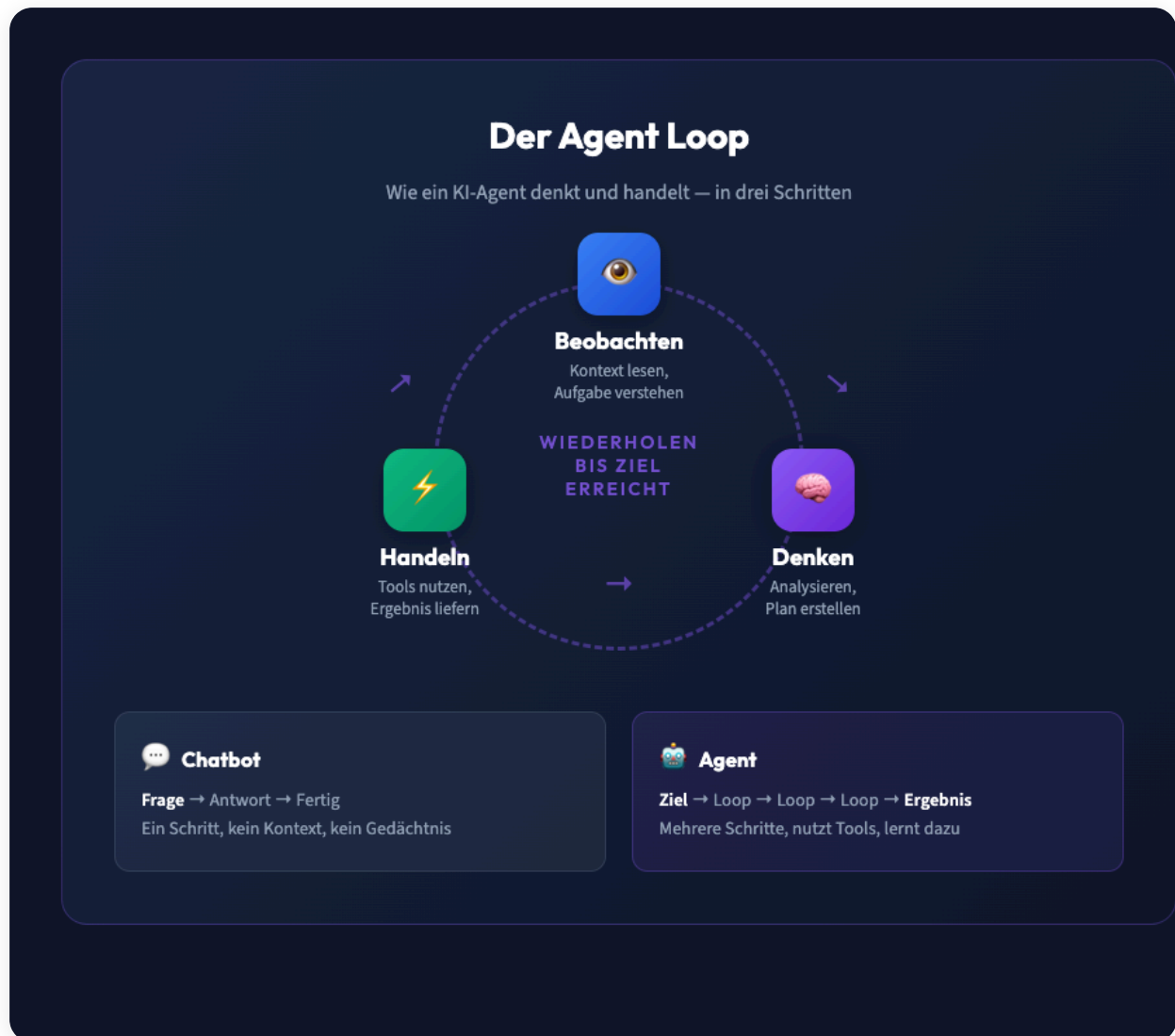
Einleitung

Künstliche Intelligenz hat sich in den letzten Jahren rasant entwickelt. Während die meisten Menschen KI als Chatbot kennen — man stellt eine Frage und erhält eine Antwort — geht die nächste Generation weit darüber hinaus: **KI-Agenten**.

Ein KI-Agent ist kein besserer Chatbot. Er ist ein digitaler Mitarbeiter, der eigenständig Aufgaben erledigt, Werkzeuge nutzt, aus Fehlern lernt und komplexe Workflows über mehrere Schritte hinweg ausführt.

Dieser Leitfaden erklärt die Konzepte hinter KI-Agenten — verständlich, praxisnah und ohne unnötigen Fachjargon.

1. Der Agent Loop — Wie ein KI-Agent arbeitet



Das Kernprinzip

Jeder KI-Agent folgt einem einfachen, aber mächtigen Zyklus aus drei Schritten:

 **Beobachten** — Der Agent liest seinen Kontext: Was ist die Aufgabe? Welche Informationen sind verfügbar? Was wurde bereits erledigt?

 **Denken** — Der Agent analysiert die Situation, wägt Optionen ab und erstellt einen Plan. Hier kommt die Intelligenz des Sprachmodells zum Einsatz.

 **Handeln** — Der Agent führt eine konkrete Aktion aus: eine Datei lesen, eine E-Mail senden, eine Websuche starten oder einen Bericht erstellen.

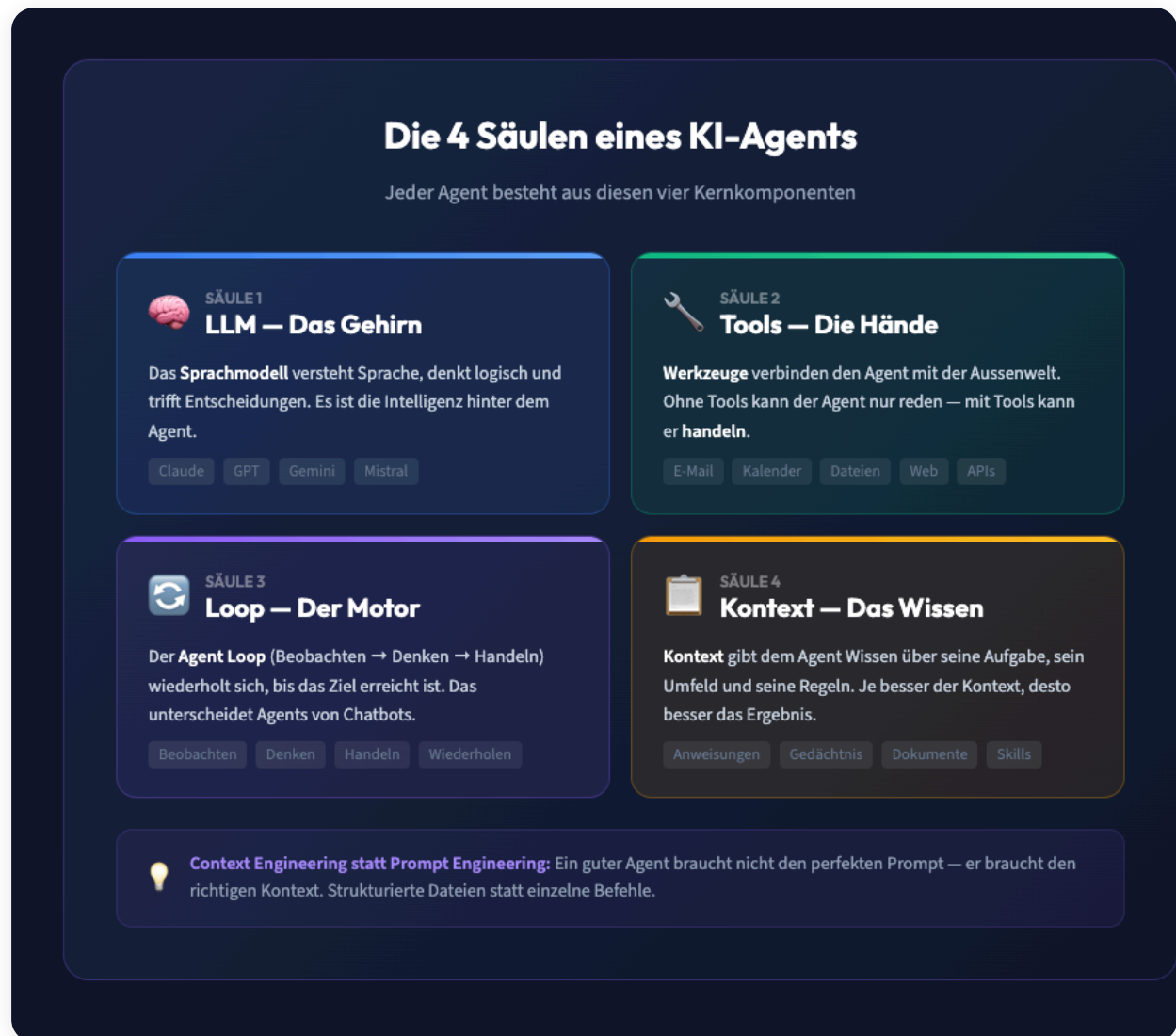
Der entscheidende Unterschied

Dieser Zyklus **wiederholt sich**, bis das Ziel erreicht ist. Das ist der fundamentale Unterschied zwischen einem Chatbot und einem Agenten:

	Chatbot	Agent
Ablauf	Frage → Antwort → Fertig	Ziel → Loop → Loop → Ergebnis
Schritte	Ein einzelner	Mehrere, iterativ
Kontext	Keiner (jede Frage ist neu)	Aufgebaut über die gesamte Aufgabe
Tools	Keine	Nutzt Werkzeuge aktiv
Gedächtnis	Keines	Lernt und merkt sich Präferenzen

Ein Agent arbeitet also nicht reaktiv (Frage-Antwort), sondern **proaktiv** — er verfolgt ein Ziel und arbeitet selbständig darauf hin.

2. Die vier Säulen eines KI-Agents



Jeder KI-Agent besteht aus vier Kernkomponenten. Fehlt eine davon, ist das System unvollständig.

Säule 1: LLM — Das Gehirn 🧠

Das **Large Language Model** (LLM) ist die Intelligenz hinter dem Agent. Es versteht natürliche Sprache, kann logisch denken, Texte verfassen und Entscheidungen treffen.

Bekannte Modelle sind Claude (Anthropic), GPT (OpenAI), Gemini (Google) oder Mistral. Jedes hat Stärken und Schwächen — die Wahl des richtigen Modells hängt vom Einsatzgebiet ab.

Wichtig: Das LLM allein ist nur ein Sprachmodell. Erst in Kombination mit den anderen drei Säulen wird es zum Agenten.

Säule 2: Tools — Die Hände

Tools (auch MCPs — Model Context Protocol — genannt) verbinden den Agent mit der realen Welt. Ohne Tools kann ein Agent nur Text generieren. Mit Tools kann er:

- **E-Mails** lesen und senden
- **Dateien** erstellen, bearbeiten und organisieren
- **Kalender** prüfen und Termine planen
- **Webseiten** durchsuchen und Informationen extrahieren
- **Datenbanken** abfragen und aktualisieren
- **APIs** ansprechen (CRM, ERP, Buchhaltung, etc.)

Die Anzahl und Art der verfügbaren Tools bestimmt, was ein Agent leisten kann. Ein gut ausgestatteter Agent hat Zugriff auf alle relevanten Systeme einer Organisation.

Säule 3: Loop — Der Motor

Der **Agent Loop** (Beobachten → Denken → Handeln → Wiederholen) ist das Herzstück. Er ermöglicht es dem Agent, komplexe Aufgaben in überschaubare Schritte zu zerlegen und iterativ zum Ziel zu arbeiten.

Ein Beispiel: Der Auftrag «Erstelle einen Marktbericht» erfordert mehrere Loops:

1. **Loop 1:** Nachrichten durchsuchen, relevante Artikel identifizieren
2. **Loop 2:** Marktdaten abrufen (Kurse, Indizes, Trends)
3. **Loop 3:** Analyse durchführen, Zusammenhänge erkennen
4. **Loop 4:** Bericht verfassen, formatieren, zur Freigabe vorlegen

Ohne den Loop wäre der Agent nach dem ersten Schritt fertig — wie ein Chatbot.

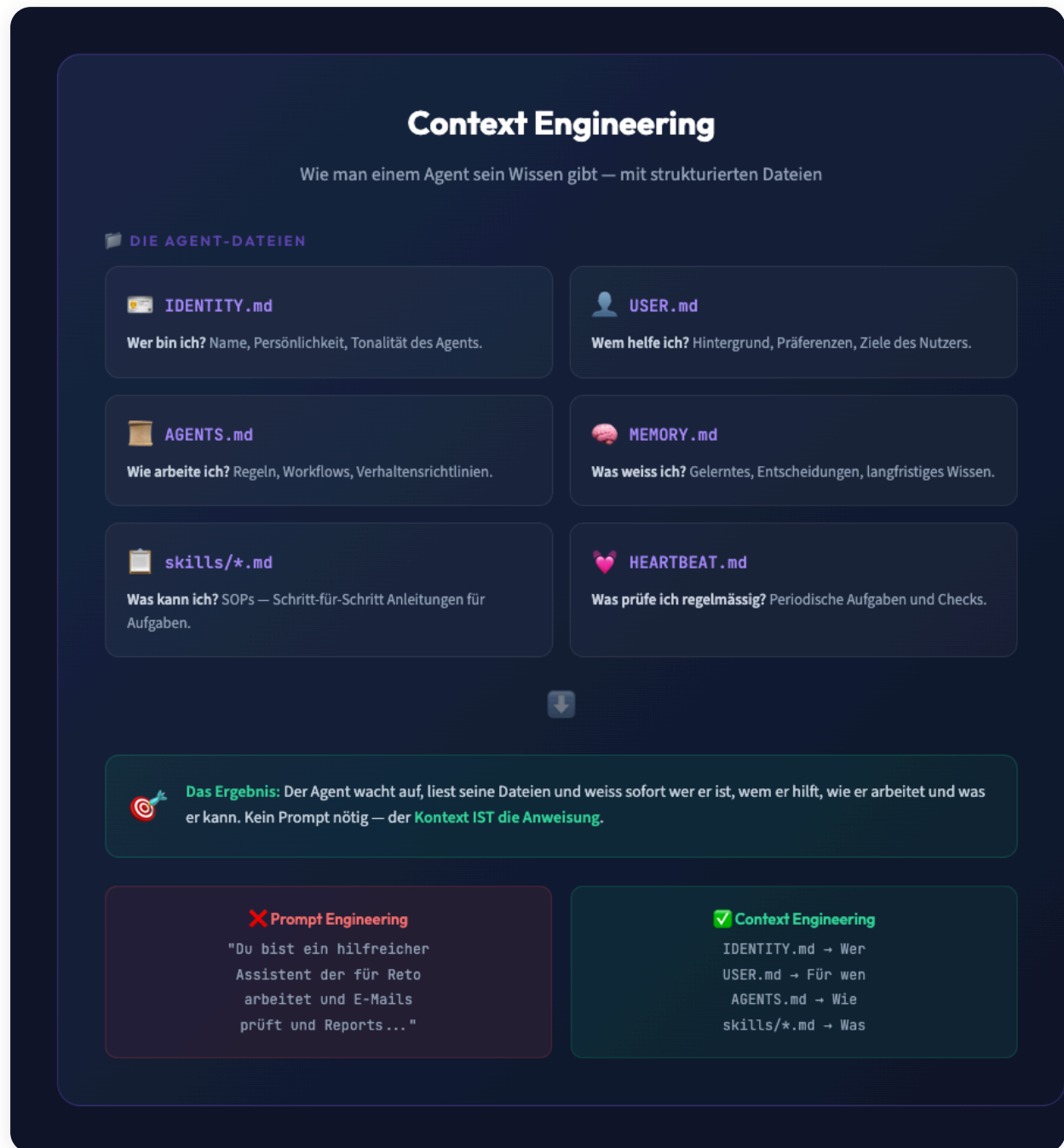
Säule 4: Kontext — Das Wissen

Kontext ist das Wissen, das der Agent braucht, um seine Aufgabe richtig zu erledigen. Dazu gehören:

- **Wer er ist** — Seine Rolle, sein Tonfall, seine Grenzen
- **Wem er hilft** — Hintergrund, Präferenzen, Ziele des Nutzers
- **Wie er arbeitet** — Regeln, Workflows, Prioritäten
- **Was er weiss** — Gelerntes, vergangene Entscheidungen, Fachwissen

Je besser der Kontext, desto besser das Ergebnis. Das führt uns zum nächsten Konzept.

3. Context Engineering — Die Kunst des richtigen Kontexts



Von Prompt Engineering zu Context Engineering

Die erste Generation von KI-Nutzung basierte auf **Prompt Engineering** — dem Versuch, die perfekte Eingabe zu formulieren, um das gewünschte Ergebnis zu erhalten. Das funktioniert für einfache Aufgaben, stösst aber schnell an Grenzen.

Context Engineering geht einen Schritt weiter: Statt einzelner Prompts erstellt man **strukturierte Dateien**, die dem Agent dauerhaft zur Verfügung stehen. Der Agent liest diese Dateien bei jedem Start und weiss sofort, wer er ist, was er kann und wie er arbeiten soll.

Die Agent-Dateien

Datei	Frage	Zweck
IDENTITY.md	Wer bin ich?	Name, Persönlichkeit, Tonalität
USER.md	Wem helfe ich?	Hintergrund, Präferenzen des Nutzers
AGENTS.md	Wie arbeite ich?	Regeln, Workflows, Verhaltensrichtlinien
MEMORY.md	Was weiss ich?	Gelerntes, Entscheidungen, langfristiges Wissen
skills/*.md	Was kann ich?	SOPs — Schritt-für-Schritt-Anleitungen
HEARTBEAT.md	Was prüfe ich?	Periodische Aufgaben und Checks

Warum ist das besser als ein Prompt?

1. **Persistent** — Der Kontext bleibt erhalten, auch wenn der Agent neu startet
2. **Strukturiert** — Verschiedene Aspekte sind in eigenen Dateien organisiert
3. **Erweiterbar** — Neues Wissen wird einfach in die richtige Datei geschrieben
4. **Wartbar** — Änderungen an einer Stelle wirken sich sofort aus
5. **Teilbar** — Teams können gemeinsam an den Kontext-Dateien arbeiten

Der Kontext IST die Anweisung. Kein einzelner Prompt kann die Tiefe und Breite eines gut strukturierten Dateisystems ersetzen.

4. Der Self-Improving Loop — Wie Agents aus Fehlern lernen



Das Prinzip der kontinuierlichen Verbesserung

Ein KI-Agent macht Fehler — besonders am Anfang. Das ist normal. Was einen guten Agent von einem schlechten unterscheidet, ist die Fähigkeit, **aus Korrekturen zu lernen und den gleichen Fehler nie zu wiederholen**.

Der Ablauf

1. **Agent macht einen Fehler** — Er wählt den falschen Ansatz, vergisst eine Präferenz oder macht eine falsche Annahme.
2. **Nutzer korrigiert** — «Nicht so, sondern so.» Ein einfaches Feedback genügt.
3. **Agent analysiert die Ursache** — Statt nur den Fehler zu beheben, fragt sich der Agent: *Warum* ist das passiert? War es eine falsche Annahme? Fehlendes Wissen? Eine unklare Regel?
4. **Agent aktualisiert seine Dateien** — Die richtige Kontext-Datei wird permanent geändert:
 - Falsches Verhalten → Update in **AGENTS.md**
 - Falsches Wissen → Update in **MEMORY.md**
 - Falsche Präferenz → Update in **USER.md**
5. **Fehler wiederholt sich nie** — Beim nächsten Start liest der Agent die aktualisierte Datei. Das System wird permanent besser.

Das Ergebnis

Über Zeit nimmt die Anzahl der Korrekturen drastisch ab. Nach wenigen Wochen kennt der Agent die Präferenzen, Regeln und Eigenheiten seines Nutzers so gut, dass Korrekturen zur Ausnahme werden.

Dieses Prinzip ist vergleichbar mit einem neuen Mitarbeiter: In der ersten Woche braucht er viel Anleitung. Nach einem Monat arbeitet er weitgehend selbständig. Der Unterschied: Ein KI-Agent vergisst nie, was er gelernt hat.

5. Praxisbeispiel: Ein Agent im Einsatz

Praxisbeispiel: Agent im Einsatz

So sieht ein typischer Agent-Workflow aus — von der Anfrage bis zum Ergebnis

Szenario: «Erstelle einen Marktbericht für unsere Kunden»

NUTZER

Erstelle den täglichen Marktbericht für heute.

AGENT — Beobachten

-  Lese skills/market-analysis.md... 
-  Lese memory/daily/gestern.md... 
-  Prüfe letzte Reports als Vorlage... 

AGENT — Handeln (Loop 1)

-  Suche aktuelle Crypto-News... 31 Artikel
-  Suche TradFi-Daten... S&P, Gold, Bonds
-  Suche AI-News... 34 Artikel

AGENT — Denken

-  Analysiere, gewichte, formuliere...
- Hauptthemen: GTC 2026, EU AI Act, BTC Konsolidierung

AGENT — Handeln (Loop 2)

-  Report erstellt (DE + EN)
-  E-Mail vorbereitet → wartet auf Freigabe
-  Fertig in 3 Minuten


Skill lesen
~2s


Web-Suche
~30s


Analyse
~60s


Report
~60s


Fertig
~3 Min



Research

Marktanalysen, Reports, Wettbewerbsvergleiche



Kommunikation

E-Mails, Follow-ups, Kundenkorrespondenz



Automatisierung

Monitoring, Alerts, wiederkehrende Tasks



Dokumentation

Protokolle, SOPs, Wissensmanagement



Compliance

Regulatorische Checks, Audit-Trails



Delegation

Sub-Agents für parallele Aufgaben

Szenario: Täglicher Marktbericht

Ein Unternehmen nutzt einen KI-Agenten für die tägliche Erstellung eines Marktberichts. Der gesamte Prozess dauert weniger als 3 Minuten — vollautomatisch.

Schritt 1 — Kontext laden (~2 Sekunden) Der Agent liest seine Skill-Datei «Market Analysis» und weiss genau, welche Quellen er nutzen soll, welches Format der Bericht haben muss und welche Themen relevant sind.

Schritt 2 — Recherche (~30 Sekunden) Der Agent durchsucht automatisch Nachrichtenquellen: Crypto-News, TradFi-Daten (S&P 500, Gold, Anleihen), AI-News. Er sammelt über 100 aktuelle Artikel.

Schritt 3 — Analyse (~60 Sekunden) Die gesammelten Informationen werden gewichtet, kategorisiert und in Kontext gesetzt. Der Agent identifiziert die 3-5 wichtigsten Themen des Tages.

Schritt 4 — Report erstellen (~60 Sekunden) Der Bericht wird in zwei Sprachen (Deutsch und Englisch) verfasst, formatiert und als PDF generiert.

Schritt 5 — Zustellung Der Report wird per E-Mail vorbereitet und wartet auf die Freigabe des Verantwortlichen. Nach Bestätigung erfolgt der Versand automatisch.

Einsatzgebiete für KI-Agenten

Bereich	Beispiele
Research	Marktanalysen, Wettbewerbsvergleiche, Due Diligence
Kommunikation	E-Mail-Triage, Follow-ups, Kundenkorrespondenz
Automatisierung	Monitoring, Alerts, wiederkehrende Reports
Dokumentation	Protokolle, SOPs, Wissensmanagement
Compliance	Regulatorische Checks, Audit-Trails, Datenschutz
Delegation	Sub-Agenten für parallele Aufgaben

Zusammenfassung

KI-Agenten sind die nächste Stufe der künstlichen Intelligenz — vom passiven Chatbot zum aktiven digitalen Mitarbeiter. Die wichtigsten Konzepte:

1. **Der Agent Loop** — Beobachten, Denken, Handeln, Wiederholen
2. **Die vier Säulen** — LLM, Tools, Loop, Kontext
3. **Context Engineering** — Strukturierte Dateien statt einzelner Prompts
4. **Self-Improving Loop** — Aus Korrekturen lernen, Fehler nie wiederholen

Der Schlüssel zum Erfolg liegt nicht in der Technologie allein, sondern in der **richtigen Konfiguration und dem richtigen Kontext**. Ein gut eingerichteter Agent kann innerhalb weniger Wochen zu einem unverzichtbaren Teammitglied werden.

Über Reto Grau Consulting

Reto Grau Consulting GmbH unterstützt Unternehmen und Organisationen bei der Einführung und Nutzung von KI-Agenten — von der Strategie über die Implementierung bis zum laufenden Betrieb.

Unsere Schwerpunkte:

- KI-Strategie und Implementierung
- On-Premise AI-Lösungen (Datensouveränität)
- Agent-Entwicklung und -Training
- Compliance und Datenschutz (CH/EU)

 kontakt@reto-grau-consulting.ch  reto-grau-consulting.ch

© 2026 Reto Grau Consulting GmbH, Langnau am Albis, Schweiz. Alle Rechte vorbehalten. Dieses Dokument dient der Information und stellt keine Beratung dar.