Von AI-Hype zu agentischer Betriebsfähigkeit

Ein Positionspapier für Reto Grau Consulting

Die Diskussion rund um agentische KI-Systeme dreht sich derzeit vor allem um immer leistungsfähigere Modelle. Diese Sicht ist nachvollziehbar, greift aber zu kurz. Wer versucht, Agentensysteme in realen Organisationen produktiv einzusetzen, erkennt rasch, dass Modellintelligenz nur ein Teil der Gleichung ist. Der eigentliche Engpass liegt weit häufiger in der Art und Weise, wie Modelle in sichere, steuerbare und belastbare Arbeitsumgebungen eingebettet werden.

Diese Beobachtung ist nicht abstrakt. Anthropic und OpenAI haben nach eigener Darstellung ein ganzes Jahr damit verbracht, herauszufinden, warum Unternehmen ihre Werkzeuge nicht produktiv einsetzen konnten — obwohl sie intern damit enorme Geschwindigkeit erreichten. Die Erkenntnis war ernüchternd: Die Technologie allein reicht nicht. Was fehlte, war die operative Fähigkeit, Agenten in bestehende Strukturen, Prozesse und Verantwortlichkeiten einzubetten. Genau deshalb binden beide Anbieter jetzt große Beratungsfirmen ein — nicht aus Liebe zum Consulting, sondern weil sie erkannt haben, dass die eigentliche Arbeit jenseits des Modells liegt.

Dieses Positionspapier fasst diese Sichtweise zusammen und leitet daraus ein Zielbild für Reto Grau Consulting ab. Es verfolgt drei Ziele: Erstens die Einordnung der aktuellen Agenten-Debatte, zweitens die Präzisierung der Konsequenzen für unser eigenes Setup mit OpenClaw, DGX und Mission Control, und drittens die Ableitung einer belastbaren Perspektive für kommunale und andere sensible Einsatzfelder mit full-on-prem Fokus.

1. Die Grundthese: Produktive Agentik ist ein Betriebsmodell

Die entscheidende Einsicht lautet: Viele Probleme, die heute als neuartige und besonders schwer beherrschbare Herausforderungen der Agentenwelt präsentiert werden, sind im Kern bekannte Probleme der Software- und Datenwelt in neuer Form. Lange Kontexte müssen strukturiert und verdichtet werden. Systeme müssen beobachtbar und messbar sein. Code und Prozesse benötigen Leitplanken. Mehrere ausführende Einheiten müssen

koordiniert werden. Und vor allem braucht es klare Spezifikationen dessen, was ein System leisten soll.

Damit verschiebt sich der Fokus. Nicht das Modell allein schafft Wert. Wert entsteht dort, wo ein Unternehmen in der Lage ist, einer KI präzise Ziele, saubere Datenzugänge, klare Freiräume, belastbare Sicherheitsgrenzen und nachvollziehbare Qualitätskriterien zu geben. Gute Agentensysteme sind deshalb keine reine Frage von Prompts oder Modellwahl. Sie sind eine Frage operativer Architektur.

Für Reto Grau Consulting ist diese Perspektive strategisch bedeutsam. Unsere Chance liegt darin, agentische Systeme nicht als nächste technische Spielerei zu positionieren, sondern in eine Form zu bringen, die für reale Organisationen kontrolliert, souverän und nachvollziehbar funktioniert.

2. Was die klassischen Systemprinzipien weiterhin lehren

Die aktuelle Agentenwelle verändert die Benutzeroberfläche der Informatik, nicht aber ihre grundlegenden Gesetzmäßigkeiten. Gute Engineering-Prinzipien behalten ihre Gültigkeit — gerade für Entscheider und Projektverantwortliche, die nicht aus der Softwareentwicklung kommen, lohnt sich der Blick auf diese bewährten Grundsätze, weil sie die agentische Welt greifbar und steuerbar machen.

Erstens lassen sich Engpässe selten zuverlässig im Voraus erraten. In agentischen Systemen ist das besonders relevant. Was zunächst wie ein Modellproblem wirkt, ist oft ein Problem der Kontextstruktur, der Tool-Integration, der Dateiorganisation oder der fehlenden Messung. Wer zu früh an der falschen Stelle optimiert, verschiebt Komplexität, ohne das eigentliche Problem zu lösen. Ein konkretes Beispiel aus unserer eigenen Praxis: Bei der Entwicklung unserer Portfolio-Bewertung (Mark-to-Market) lag das eigentliche Problem nicht in der KI-Qualität, sondern in fehlerhaften Ticker-Mappings und unklaren Preisquellen. Der Agent lieferte stabile Ergebnisse, sobald die Datengrundlage stimmte — nicht umgekehrt.

Zweitens gilt weiterhin, dass Systeme gemessen werden müssen, bevor sie sinnvoll optimiert werden können. Agentische Workflows benötigen Baselines, Testfälle, Vergleichswerte und klare Qualitätskriterien. Ohne Metriken bleibt die Beurteilung von Qualität zu subjektiv, und Organisationen diskutieren über Eindrücke statt über Leistung. Factory.ai, ein auf agentische Entwicklung spezialisiertes Unternehmen, hat dies systematisch untersucht und festgestellt: Teams, die ihre Agentenleistung nicht messen,

können sie auch nicht verbessern. Die einfache Disziplin, eine Baseline zu definieren und Ergebnisse dagegen zu prüfen, liefert oft mehr Fortschritt als ein Modellwechsel.

Drittens bleibt Einfachheit ein strategischer Vorteil. Viele Teams neigen dazu, bei Agenten sofort mit komplexen Architekturen aus Rollen, Schleifen und Spezialpfaden zu operieren. In der Praxis zeigt sich jedoch häufig, dass einfache, robuste und gut verstandene Strukturen überlegen sind. Rob Pike, einer der Schöpfer von Unix und Go, hat dieses Prinzip schon vor Jahrzehnten formuliert: Einfache Algorithmen sind nicht nur schneller zu verstehen, sie sind auch weniger fehleranfällig. In der Agentenwelt gilt das verstärkt, weil komplexe Systeme schwer zu debuggen sind — man weiß oft nicht, ob ein Problem am Prompt, am Kontext, am Tool oder an der Orchestrierung liegt.

Viertens gilt weiterhin, dass strikte Leitplanken Qualität schaffen. Im klassischen Engineering geschieht dies durch Linting, Standards und Tests. In der Agentenwelt braucht es diese Disziplin ebenfalls, teilweise sogar stärker als zuvor. Agenten arbeiten schnell und produktiv, aber nicht automatisch elegant, nachhaltig oder revisionssicher. Factory.ai beschreibt es treffend: Ohne strikte Linting-Regeln verhält sich ein Agent wie ein Entwickler, der nur schnell abliefern will, ohne sich um Codequalität zu kümmern. Erst die formale Durchsetzung von Standards — im Code wie in der Wissensarbeit — macht agentische Produktivität verlässlich.

Fünftens bleibt die zentrale Einsicht der Informatik bestehen: Daten und Struktur dominieren das Ergebnis. Ein leistungsfähiger Agent in einem schlecht organisierten Informationsraum wird nicht produktiv sein. Gute Kontextstrukturen, klare Regeln, zugängliche Projekthierarchien und saubere Wissensbausteine sind oft wertvoller als das nächste Modell-Upgrade. Auch hier zeigt die Praxis den Unterschied: Wenn ein Chatbot für eine Gemeinde auf sauber strukturierte, aktuelle und korrekt klassifizierte Wissensdokumente zugreift, steigt die Antwortqualität messbar — unabhängig davon, welches Modell dahinter läuft.

---

3. Die fünf realen Produktionsprobleme agentischer Systeme

Wer Agenten jenseits von Demos produktiv einsetzen will, trifft früher oder später auf fünf wiederkehrende Problemfelder.

3.1 Kontextkompression

Lange Agentensitzungen füllen jedes Kontextfenster. Auch große Kontextfenster beseitigen dieses Problem nicht, sondern verschieben es lediglich. Deshalb müssen

Arbeitsstände verdichtet, segmentiert und strukturiert werden. Jede Verdichtung ist verlustbehaftet. Die operative Herausforderung besteht darin, Ziele, Entscheidungen, Änderungen und nächste Schritte so zu sichern, dass Arbeit über längere Zeiträume stabil fortgesetzt werden kann.

Factory.ai hat drei verschiedene Kompressionsstrategien systematisch verglichen — ihre eigene inkrementelle Zusammenfassung, OpenAIs kompakten Endpunkt und Anthropics SDK-Kompression — und festgestellt, dass alle Ansätze beim Tracking konkreter Artefakte Schwächen zeigen. Die praktische Konsequenz: Langfristige Agentenarbeit erfordert explizite Meilensteine und strukturierte Zwischenstände, nicht nur bessere Kompressionsalgorithmen.

3.2 Instrumentierung und Messbarkeit

Agenten benötigen eine Umgebung, die beobachtbar ist. Dazu gehören reproduzierbare Builds, Baselines, Testdaten, Monitoring und nachvollziehbare Erfolgskriterien. Viele angebliche Agentenprobleme sind in Wahrheit Folge mangelnder Softwarehygiene. Eine unvorbereitete Umgebung erzeugt instabile Resultate, unabhängig davon, wie leistungsfähig das Modell ist.

3.3 Linting und formale Leitplanken

Wenn Agenten in Code oder Wissensarbeit eingreifen, brauchen sie explizite Grenzen. Im technischen Bereich sind das Linting-Regeln, Stilvorgaben, Tests und statische Prüfungen. Im weiteren Sinne gehören dazu aber auch formale Standards für Berichte, E-Mails, Analysen und Freigaben. Leitplanken reduzieren nicht die Produktivität, sondern erhöhen Verlässlichkeit, Konsistenz und Wartbarkeit.

3.4 Multi-Agent-Koordination

Sobald mehrere Agenten zusammenarbeiten, entsteht ein Koordinationsproblem. Rollen, Übergaben, Zuständigkeiten und Eskalationen müssen klar definiert werden. Viele Teams machen den Fehler, vorschnell hochkomplexe Schwarmmodelle zu bauen. Praktikabler ist häufig ein klarer Planner-Executor-Ansatz, der erst bei nachgewiesenem Bedarf erweitert wird.

3.5 Spezifikationen und menschliche Ermüdung

Das schwierigste Problem ist die Spezifikation. Ein Agent ist nur dann wirklich produktiv, wenn klar definiert ist, was er erreichen soll, welche Informationen er nutzen darf, welche Grenzen gelten, wie Erfolg gemessen wird und wann menschliche Freigaben nötig sind. Gute Spezifikationen verlangen Disziplin, Präzision und strukturiertes Denken.

Gerade hier zeigt sich, dass Agenten Denk- und Führungsarbeit nicht einfach abschaffen. Sie machen sie sichtbarer. Die häufig anzutreffende Vorstellung, Menschen könnten sich zurücklehnen, während Agenten den Rest erledigen, ist operativ irreführend. Agenten reduzieren bestimmte Tätigkeiten, erhöhen aber zugleich die Anforderungen an saubere Zieldefinition, Informationsstruktur und Qualitätskontrolle.

Diese fünfte Dimension ist deshalb entscheidend, weil sie nicht allein technisch gelöst werden kann. Sie verlangt Fachverständnis, organisatorische Klarheit und gutes Requirements Engineering. In vielen Fällen liegt genau hier der eigentliche Hebel professioneller KI-Einführung.

---

4. Konsequenzen für unser eigenes Setup

Für unser eigenes Setup liefert diese Analyse einen nützlichen Abgleich zwischen dem, was wir bereits aufgebaut haben, und dem, was noch fehlt.

Mit OpenClaw verfügen wir über eine agentische Laufzeitumgebung, die Tool-Nutzung, Skills, Sessions, Memory und Steuerlogik miteinander verbindet. Mit dem DGX Spark verfügen wir über einen lokalen Compute- und Souveränitätsanker. Mit Mission Control entwickeln wir eine operative Oberfläche, in der Sichtbarkeit, Steuerung und Kontrollierbarkeit zusammenlaufen. Unsere Memory-Strukturen, Daily Files, Skills und Approval-Regeln zeigen, dass wir Agentik nicht als lose Chat-Funktion, sondern als Betriebsumgebung verstehen.

Gleichzeitig zeigt der ehrliche Blick auch, wo wir noch Lücken haben. Unsere Messbarkeit ist heute überwiegend qualitativ — wir sehen, ob etwas funktioniert, aber wir messen Erfolgsquoten, Korrekturschleifen oder Policy-Hits noch nicht systematisch. Einige Regeln und Zuständigkeiten existieren bereits, aber häufig noch implizit statt als formalisierte Policy-Schicht. Und unsere Multi-Agent-Koordination funktioniert in der Praxis, ist aber noch nicht so standardisiert, dass sie ohne Kontextwissen sauber repliziert werden könnte.

Genau aus diesem Abgleich ergeben sich die nächsten Reifeschritte.

Erstens: Präzisere Spezifikationen. Für zentrale Arbeitsformen muss klarer und wiederverwendbar beschrieben sein, was ein gutes Ergebnis ist, welche Artefakte erwartet werden, welche Quellen zulässig sind und welche Entscheidungen autonom getroffen werden dürfen. Je präziser diese operative Sprache wird, desto produktiver und verlässlicher arbeiten Agenten.

Zweitens: Systematischere Multi-Agent-Koordination. Rollen, Übergabeformate, Ownership, Erfolgskriterien und Eskalationsregeln sollten schrittweise standardisiert werden. Das erhöht Skalierbarkeit, reduziert Reibung und macht längere Arbeitsketten robuster.

Drittens: Messbarkeit und Observability. Wenn wir langfristig eine wirklich professionelle agentische Arbeitsweise aufbauen wollen, brauchen wir nicht nur gute Outputs, sondern auch ein besseres Verständnis ihrer Entstehung. Erfolgsquoten, Korrekturschleifen, Laufzeiten, Fehlerklassen und Policy-Hits sollten mit der Zeit stärker formalisiert werden.

Viertens: Explizitere Policy-Schicht. Rechte und Grenzen sollten deutlicher klassifiziert werden: lesen, intern schreiben, extern kommunizieren, publizieren, Infrastruktur ändern oder finanzrelevante Aktionen anstoßen. Diese Unterscheidungen sind keine Bürokratie, sondern Teil eines belastbaren Betriebsmodells.

Fünftens: Modulare Kontextarchitektur und systematisches Projektmanagement. Längere Projekte und mehrschichtige Aufgaben erfordern saubere Informationspfade, retrieval-fähige Wissensbausteine und strukturierte Zwischenstände.

In diesem Zusammenhang gehört auch unser systematisches Projektmanagement ausdrücklich in den Mittelpunkt. Wir arbeiten bereits heute mit einer Kanban-Logik, Feature-Zerlegung, definierten Statusübergängen und klar zugeordneten Arbeitspaketen. Das ist kein Zufall, sondern eine bewusste Entscheidung: Gerade bei längeren Vorhaben verhindert eine saubere Board- und Feature-Struktur, dass Kontext implizit im Chat oder in Einzelköpfen hängen bleibt. Sie macht Fortschritt sichtbar, reduziert Übergabeverluste und bildet die operative Brücke zwischen Spezifikation, Umsetzung, Testing und Freigabe. Für agentische Systeme ist das besonders wichtig, weil Agenten — anders als menschliche Teammitglieder — keinen informellen Flurfunk haben. Was nicht explizit strukturiert ist, existiert für sie nicht. Ein Kanban-Board mit klaren Statusübergängen ist deshalb kein Projektmanagement-Overhead, sondern die Grundlage dafür, dass Mensch und Agent verlässlich am selben Vorhaben weiterarbeiten können.

5. Konsequenzen für ein Gemeinde-Setup mit full-on-prem Fokus

Für Gemeinden und andere sensible Organisationen wird diese Sichtweise noch relevanter. Dort steht nicht maximale Modellneuheit im Vordergrund, sondern Vertrauen.

Ein kommunales KI-System muss nicht nur nützlich sein, sondern kontrollierbar, nachvollziehbar und politisch verantwortbar.

Gerade deshalb ist ein full-on-prem oder souveränes hybrides Setup strategisch attraktiv. Es verbindet Datenhoheit, begrenzte Zugriffsrechte, kontrollierte Modellpfade und klare Governance. Für Gemeinden geht es nicht primär darum, die spektakulärste KI zu betreiben, sondern eine belastbare KI-Betriebsumgebung zu schaffen, die mit öffentlicher Verantwortung vereinbar ist.

Hier können wir bereits substanziell ansetzen. Unsere Arbeit rund um DGX, lokale Inferenz, Governance, Memory, Tool-Grenzen und operative Klarheit passt zu den Anforderungen kommunaler Organisationen. Gleichzeitig zeigt die Analyse, dass wir das Gemeindeangebot noch präziser entlang von Runtime-Sicherheit, Spezifikation, Auditierbarkeit und einfacher Betriebslogik zuschneiden sollten.

Ein überzeugendes Gemeindeangebot muss klar beantworten können: - Welche Daten bleiben lokal? - Was darf ein Agent lesen, schreiben oder nie tun? - Welche Antworten sind freigegeben, welche benötigen Eskalation? - Wie werden Quellen, Versionen und Entscheidungen nachvollziehbar gemacht? - Wie bleibt das System für Verwaltung und Politik verständlich?

Eine kommunale KI-Lösung ist nicht einfach ein Modell mit Chatfenster, sondern eine Betriebsarchitektur. Und genau diese Betriebsarchitektur zu bauen und verständlich zu machen, ist unser Kerngeschäft.

---

6. Fazit

Die eigentliche Transformation der KI-Welt liegt nicht nur in besserer Modellleistung, sondern in der Professionalisierung ihrer Laufzeit- und Betriebsumgebung. Wer diese Ebene beherrscht, wird produktive Agentensysteme aufbauen können, die über Showcases hinausgehen. Wer sie vernachlässigt, wird trotz guter Modelle an Unklarheit, Unsicherheit und organisatorischer Reibung scheitern.

Für Reto Grau Consulting bedeutet das ein klares Zielbild: sichere, steuerbare, souveräne und produktive agentische Betriebsmodelle entwickeln — intern als Referenzarchitektur, extern als Angebot für Organisationen, die Kontrolle und Nachvollziehbarkeit ernst nehmen. Unser Setup mit OpenClaw, DGX und Mission Control bildet dafür bereits ein starkes Fundament. Die nächsten Reifeschritte sind benannt: Spezifikation, Messbarkeit, Policy-Schärfung, Multi-Agent-Koordination und Kontextarchitektur.

Der nächste konkrete Schritt ist, diese Reifeschritte als interne Roadmap zu operationalisieren und an den ersten beiden Prioritäten — Spezifikationen und Multi-Agent-Koordination — messbare Fortschritte zu erzielen. Damit machen wir den Unterschied zwischen Zielbild und gelebter Praxis kleiner.