

Warum Multi-Agenten-Systeme nur mit menschlicher Kontrolle Nutzen stiften

Multi-Agenten-Systeme übernehmen zunehmend komplexe Aufgaben in Unternehmen. Nutzen stiften sie jedoch nur dann, wenn der Mensch die Kontrolle behält und ihre Qualität nachweisbar wird. KI-Agenten arbeiten längst nicht mehr isoliert, sondern agieren im Team: Sie planen, prüfen und berichten. Doch je mehr Verantwortung sie übernehmen, desto dringender wird die Frage, wer ihre Arbeit kontrolliert. Die Antwort entscheidet darüber, ob aus einem Experiment ein belastbares Werkzeug wird.



Joachim Schü
Geschäftsführer



Sebastian Wagmann
Produktmanager AI-Lösungen



Max Hirsch
AI-Ingenieur

Heute liegt schon nach sehr kurzer Zeit eine fundierte Erstbewertung der Sachlage vor. Erstellt wurde sie aber nicht von den Mitarbeitern, sondern von einem Team spezialisierter KI-Agenten.

und gegenseitige Kontrolle zu besseren Ergebnissen. Denn für Unternehmensentscheider zählt nicht nur die Geschwindigkeit, sondern vor allem auch die Verlässlichkeit.

Denn so groß die Vorteile durch den Einsatz von KI auch sind – mit der Arbeitsteilung wächst ein neues Risiko: Ein Fehler in einem frühen Schritt wird weitergereicht, und alle folgenden Agenten bauen darauf auf. „Am Ende steht ein überzeugend formuliertes Ergebnis – nur leider ist es falsch“, beschreibt Sebastian Wagmann eine große Gefahr.

An diesem Punkt stellt sich eine entscheidende Frage: Wer haftet, wenn der Agent irrt? Wagmann: „Ganz gleich, ob in dem Prozess KI eingesetzt wurde oder nicht, die Antwort bleibt stets dieselbe: das Unternehmen. Hier zeigt sich, dass Autonomie ohne Kontrolle in kritischen Prozessen kein Fortschritt ist, sondern ein Risiko“. Deshalb ist ein Multi-Agenten-System nur dann nützlich, wenn der Mensch an entscheidenden Stellen prüfen und korrigieren kann. Man spricht vom Prinzip des „Human-in-the-Loop“.

Human-in-the-Loop: Gezielte Kontrolle statt Vollprüfung

Doch, so betont der Produktmanager: „Menschliche Kontrolle heißt nicht, jede Zeile nachzulesen, denn das würde den Effizienzgewinn vernichten“. Gut gestaltete Systeme binden den Menschen dort ein, wo Entscheidungen Folgen haben:

- **Freigabepunkte:** Eine Risikoeinstufung oder Maßnahmenempfehlung wird erst wirksam, wenn ein verantwortlicher Mitarbeiter sie bestätigt. Das sichert eine klare fachliche Verantwortung („Ownership“) und verhindert Schuldzuweisungen zwischen Fachbereich und IT.
- **Belege statt Behauptungen:** Spezialisierte Agenten stützen jede Aussage auf konkrete Stellen in Dokumenten, Gesetzestexten oder Richtlinien. Das Modell liefert die Logik, nicht die Fakten. Dies senkt das Risiko erfundener Antworten und erlaubt eine Prüfung per Klick.
- **Eskalation bei Unsicherheit:** Ist sich das System nicht sicher, legt es den Fall Fachleuten vor, statt zu raten.
- **Korrektur mit Wirkung:** Ändert ein Mitarbeiter eine Bewertung, wird dies dokumentiert und fließt in die Qualitätsmessung ein.

- **Lückenlose Nachvollziehbarkeit:** Jeder Arbeitsschritt jedes Agenten wird für Revision und Prüfer protokolliert.
- **Deterministische Entscheidungsregeln:** Bei festen Vorgaben wie Fristen oder Schwellenwerten entscheidet nicht die KI, sondern eine deterministische Regel. KI und feste Regeln arbeiten Hand in Hand.

Die EU-KI-Verordnung verlangt für Hochrisiko-Anwendungen genau solche Möglichkeiten menschlicher Aufsicht. Doch solche Kontrollpunkte alleine genügen nicht. Die Qualität der KI muss hoch, nachweisbar und dokumentiert sein. Schließlich: „Kein Unternehmen würde eine Produktionslinie ohne Kennzahlen betreiben – für KI gilt nichts anderes“, betont Wagmann.

Das Beratungsunternehmen Consileon hat dafür ein Qualitätsframework aufgebaut. Um eine Selbstkontrolle der KI zu vermeiden, stellt ein Input-Modul einmalig menschliche Musterlösungen bereit, gegen die das System fortlaufend bewertet wird. Spezifische Module bewerten die Genauigkeit bei Kategorien, Freitexten und Textrecherchen. Das Framework überwacht täglich zentral definierte Kennzahlen:

1. **Fehlklassifizierungen:** Erfasst falsche Entscheidungen, wobei ein falscher Alarm günstiger ist als ein übersehenes Risiko.
2. **Wiederholbarkeit:** Überprüft, ob dieselbe Frage an zwei Tagen dieselbe Antwort liefert.
3. **Rauschen:** Misst überflüssige Informationen bei Recherchen.
4. **Abdeckung:** Kontrolliert, ob alle relevanten Informationen gefunden werden. Übersehene Details führen zu schweren Folgefehlern.

Max Hirsch, AI-Ingenieur bei Consileon, erklärt: „Nutzer können diese Werte im Dashboard ablesen und erkennen, wie verlässlich das System arbeitet. Meta-KI-Agenten überwachen die Kennzahlen und schlagen gezielte Verbesserungen vor“.

Unter dem Begriff „Trusted AI“ bündelt Consileon Human-in-the-Loop-Werkzeuge und Qualitätstransparenz, ergänzt durch eine verankerte AI Governance. Sie regelt Verantwortlichkeiten und die Einhaltung der EU-KI-Verordnung.

Wie dies in der Praxis funktioniert, zeigt der Consileon Compliance Manager. Er unterstützt Unternehmen dabei, regulatorische Anforderungen wie NIS-2, DORA, DSGVO oder die EU-KI-Verordnung zu bewerten. „Dahinter arbeitet ein Team aus Agenten mit klar getrennten Rollen: Ein Steuerungs-Agent zerlegt den Auftrag. Fach-Agenten gleichen Anforderungen mit Unternehmensunterlagen ab, während der Mensch stichprobenartig Begründungen prüft“, erklärt Max Hirsch. Qualitätssicherungs-Agenten markieren Widersprüche für Eskalationsentscheidungen, und ein Kommunikations-Agent entwirft Berichte zur menschlichen Freigabe. Der Compliance Manager wurde zudem selbst gegen die Transparenzpflichten der EU-KI-Verordnung bewertet.

„Multi-Agenten-Systeme stiften echten Mehrwert bei komplexen Aufgaben wie dem Compliance-Management oder der Regulierungsprüfung“, betont Sebastian Wagmann. Der entscheidende Kundennutzen durch Consileon liegt in der nachweisbaren Verlässlichkeit: „Über ein eigenes Qualitätsframework vergleichen wir die Agenten mit menschlichen Musterlösungen und messen täglich Kennzahlen zu Genauigkeit, Wiederholbarkeit und Vollständigkeit.“ Kunden sehen die Leistungswerte transparent auf einem Dashboard und erhalten belegte Fakten statt halluzinierter Antworten. So verwandelt Consileon risikobehaftete KI-Experimente in rechtssichere, verlässliche Werkzeuge.

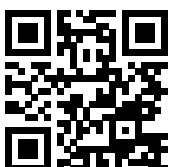
Und so kommt Joachim Schü, Geschäftsführer von Consileon, zu dem Ergebnis: „Unsere Kunden fragen nicht mehr, ob KI funktioniert. Sie fragen, woran sie erkennen, dass sie heute funktioniert. Genau diese Antwort können wir geben“.

Fünf Fragen an Ihren KI-Anbieter

Vor dem Einsatz von KI-Agenten in kritischen Prozessen sollten Führungskräfte folgende Fragen klären:

1. An welchen Stellen soll ein Mensch Ergebnisse prüfen, freigeben oder korrigieren?
2. Wie oft übersieht das System ein echtes Risiko – und woher wissen Sie das?
3. Liefert dieselbe Anfrage morgen dasselbe Ergebnis?
4. Ist jeder Arbeitsschritt für Revision und Prüfer nachvollziehbar dokumentiert?
5. Wer im Unternehmen trägt die Verantwortung für die Freigabe von KI-Ergebnissen?

Weitere Informationen unter:
consileon.de



CONSILEON

Auf Unternehmen prasseln in unseren schnelllebigen Zeiten ständig neue Aufgaben und Herausforderungen herein, die meistens nach einer raschen Lösung rufen. Dann sind die jeweiligen Fachabteilungen gefragt und müssen möglichst schnell die Lösung liefern. Ein Beispiel aus dem Alltag: Wenn ein Unternehmen mal wieder mit einer neuen regulatorischen Vorgabe konfrontiert wird, ist das ein Fall für die Compliance-Abteilung. Ein komplizierter und aufwendiger Prozess wird in Gang gesetzt, für den die Kolleginnen und Kollegen früher Wochen benötigten, denn immerhin müssen betroffene Prozesse, Richtlinien und Verträge geklärt werden. Viel Arbeit also.

Doch das hat sich radikal geändert: „Heute liegt schon nach sehr kurzer Zeit eine fundierte Erstbewertung der Sachlage vor. Erstellt wurde sie aber nicht von den Mitarbeitern, sondern von einem Team spezialisierter KI-Agenten“, sagt Sebastian Wagmann, Produktmanager AI-Lösungen bei Consileon.

Ein solches Multi-Agenten-System funktioniert nicht anders als ein gut organisiertes Projektteam: Ein steuernder Agent zerlegt den Auftrag und koordiniert die Aufgaben, spezialisierte Fach-Agenten, ausgestattet mit entsprechenden IT-Werkzeugen, recherchieren und analysieren. Schließlich prüft ein Qualitätssicherungs-Agent die Ergebnisse, bevor ein Kommunikations-Agent sie verständlich für den Auftraggeber aufbereitet. Die Entscheidungskette verläuft dabei nicht rein linear, sondern schließt bei Bedarf mehrere Abstimmungsrunden zwischen den Agenten ein.

Dass Unternehmen nicht auf ein einziges KI-Modell setzen, hat organisatorische Gründe: Sie sind auf unterschiedliche Aufgaben trainiert, nutzen spezifische Werkzeuge und unterscheiden sich in Kreativität und Regelbasierten Wenn-Dann-Systemen. So führen Arbeitsteilung, Spezialisierung