

RAG: Firmenwissen für KI nutzbar machen

Retrieval-Augmented Generation verstehen, Angebote beurteilen, Alternativen erkennen

Dauer	2 Stunden als Überblick, Halbtage mit Anforderungsübung
Format	Vortrag oder Workshop, Präsenz oder online
Zielgruppe	Geschäftsführung, IT-Verantwortliche, Prozess- und Qualitätsverantwortliche im Mittelstand
Flughöhe	Entscheidungsebene: wozu, wann und woran erkennbar. Keine Entwicklerschulung , kein Umsetzungsformat
Vorkenntnisse	Grundverständnis KI hilfreich, keine Technikenkenntnisse nötig
Gruppengröße	bis 20, im Workshop bis 12

Worum es geht

Jedes Unternehmen sitzt auf Angeboten, Verträgen, Protokollen, Handbüchern und E-Mails. Die Frage ist nicht, ob eine KI daraus Antworten geben kann. Die Frage ist, wie das technisch funktioniert, was es kostet und wo es verlässlich scheitert.

Das Verfahren dahinter heißt RAG (Retrieval-Augmented Generation). Das Sprachmodell wird nicht neu trainiert. Es bekommt zur Laufzeit die passenden Stellen aus den eigenen Dokumenten vorgelegt und antwortet belegt daraus. Wer die vier Schritte dieses Verfahrens kennt, liest ein Angebot fachlich statt entlang von Begriffen. Der Kurs richtet sich an die, die über solche Vorhaben **entscheiden und sie beauftragen**, nicht an die, die sie umsetzen.

Der Kurs behandelt ausdrücklich die Fälle, in denen RAG die falsche Wahl ist. Fragen, die gezählt, summiert oder verglichen werden, gehören in eine Auswertung, nicht in eine Dokumentensuche. Diese Trennung früh zu ziehen ist billiger als jede spätere Korrektur.

Inhalt

- **Das Grundproblem:** warum ein Sprachmodell die eigenen Firmendokumente nicht kennt und warum Nachtrainieren fast nie die Antwort ist
- **Die vier Schritte:** Einlesen, Speichern, Wiederfinden, Antworten, und an welcher Stelle welche Entscheidung fällt
- **Warum das Einlesen über alles Weitere entscheidet:** strukturbewusstes Zerlegen der Dokumente steigert die Trefferqualität in einer Literaturübersicht um bis zu 33 Prozent gegenüber dem naiven Verfahren und ist dabei rechengünstiger. Warum ausgerechnet PDF das schwierigste Eingangsformat ist
- **Die Überraschung bei der Suche:** in einem Vergleich über 23.088 Anfragen auf **Finanzdokumenten** schlug die klassische Stichwortsuche die moderne Bedeutungssuche. Welche Dokumentart welches Verfahren braucht und warum die Kombination beider fast immer gewinnt
- **Der teuerste Qualitätshebel:** die Nachbewertung der gefundenen Stellen wirkt am stärksten und kostet ein Vielfaches der reinen Suche. Wo die wirtschaftliche Grenze verläuft
- **Berechtigungen:** warum ein solches System sauber belegte Antworten aus Dokumenten geben kann, die der Fragende nicht sehen darf, und was dagegen hilft
- **Belegpflicht statt Vertrauen:** jede Aussage mit Fundstelle. Warum das die Bedingung für Akzeptanz im Betrieb ist
- **Was RAG strukturell nicht kann:** die Frage, wie viele Verträge dieses Jahr auslaufen, ist unbeantwortbar, unabhängig von der Qualität des Systems. Genau diese Frage kommt in der Praxis zuerst
- **Die Alternativen im Verfahren:** großes Kontextfenster, Suche ohne KI-Antwort, Datenbankauswertung, Schnittstellenabfrage. Wo beides möglich ist, liefert das große Kontextfenster die bessere Qualität und RAG den niedrigeren Preis je Anfrage. Große Kontextfenster nutzen ihre Mitte allerdings schlecht

KI-BERATUNG SCHROELL

- **Die Alternativen im Bezug:** selbst bauen, ein fertiges Produkt mit eingebauter Funktion kaufen, den Betrieb mieten. Drei Wege mit unterschiedlichen Folgekosten
- **Abnahme:** woran vorab festgelegt wird, wann das Ergebnis gut genug ist

Was die Teilnehmenden mitnehmen

Sie können ihre eigenen Anwendungsfälle in dokumentenbasiert und datengetrieben trennen. Diese Sortierung verhindert die teuerste Fehlentscheidung in solchen Projekten, nämlich eine Dokumentenplattform an Aufgaben zu messen, für die sie nie gebaut wurde.

Sie erkennen an einem Angebot, ob es vollständig ist: ob Berechtigungen vor der Suche greifen, ob beide Suchverfahren kombiniert werden, ob die Nachbewertung überhaupt konfiguriert ist. Die drei Stellen, an denen solche Projekte typischerweise scheitern, sind benannt, ebenso die Fragen, die einem Anbieter vorab zu stellen sind.

Sie können begründen, ob der eigene Fall RAG braucht oder ob eine einfachere Lösung reicht, und sie können festlegen, woran der Erfolg gemessen wird, bevor gebaut wird.

Im Halbtagesformat entsteht zusätzlich eine Anforderungsskizze für den eigenen Betrieb: Fragen-Triage, Bestandsprofil, zehn Prüffragen mit erwarteter Fundstelle, Anbieterfragenkatalog. Vier Seiten, ausgefüllt, als Grundlage für das nächste Ausschreibungsgespräch.

Hinweise

Der Inhalt beruht auf einem eigenen Fachbuch zur RAG-Referenzarchitektur und auf laufenden Beratungsmandaten, nicht auf Herstellerunterlagen. Zentrale Aussagen sind gegen begutachtete Fachliteratur belegt, einzelne zusätzlich durch eigene dokumentierte Messreihen.

Der Kurs ist herstellernerneutral. Werkzeuge werden verglichen, nicht empfohlen.

Das Format bleibt bewusst auf der Überblicksebene: Es erklärt, wozu das Verfahren dient, wann es die falsche Wahl ist und woran ein tragfähiges Angebot erkennbar ist. Wie ein System gebaut wird, ist nicht Gegenstand.

Datenschutzfragen werden praktisch eingeordnet, nicht juristisch behandelt. Der Kurs stellt Urteilsfähigkeit her, er ersetzt keine Rechtsberatung und keine Umsetzungsbegleitung.

KI-Beratung Schroell · Fernand Schroell · Mondseeweg 7 · 94542 Haarbach

Tel +49 157 82728500 · KI-Beratung@schroell.net · www.schroell.net