

RAG in der Praxis: Warum die meisten Systeme scheitern, bevor sie wirken

02. Juli 2026

28. CrossMediaForum

Datenqualität & Informationsverlust

**Für KI Systeme gilt mehr denn je das Prinzip:
„Garbage in, Garbage out“**

Selbst die besten Sprachmodelle bekommen Sie kaputt, indem Sie:

- > Unvollständige Dokumente einsetzen
- > Fehlerhafte OCR-Ergebnisse nicht erkennen
- > **Dokumentstrukturen nicht beachten**



Fehlende Dokumentstruktur

KI Systeme interessieren sich fast nur für Content. Aktuell werden viele semantische Strukturen nahezu vollständig verworfen!

KI Systeme setzen „Proximity Boosting“ ein, damit sie effizient arbeiten können:

- Der Text wird in beliebigen „Portionen“ zerstückelt. Sogenannte **Chunks**
- Chunks, die näher zum gesuchten Begriff sind, werden höher gewichtet

Culpa in contrahendo – Haftung aus vorvertraglichem Schuldverhältnis

Die culpa in contrahendo gehört zu den wichtigsten und bedeutsamsten Haftungsgrundlagen des deutschen Zivilrechts, weil sie eine Schutzpflicht begründet, die den Abschluss eines Vertrages begründet.

I. Rechtsgrundlage

Nach § 311 Abs. 2 BGB entsteht ein Schuldverhältnis, wenn ein Teilnehmer gemäß § 241 Abs. 2 BGB bereits durch die Aufnahme von Vertragsverhandlungen (Nr. 1), durch die Ablehnung eines Vertrages, bei welcher der eine Teil im Hinblick auf eine etwaige rechtsgeschäftliche Beziehung dem anderen Teil die Möglichkeit zur Einwirkung auf seine Rechte, Rechtsgüter und Interessen gewährt oder ihm diese anvertraut (Nr. 2), oder durch ähnliche geschäftliche Kontakte (Nr. 3). Verletzt eine Person diese Rücksichtnahmepflichten schuldhaft, haftet sie nach § 280 Abs. 1 BGB auf Schadensersatz.

II. Voraussetzungen

Erforderlich sind (1) ein vorvertragliches Schuldverhältnis im Sinne des § 311 Abs. 2 BGB, (2) die schuldhaft Verletzung einer Schutz-, Aufklärungs- oder Sorgfaltpflicht aus § 241 Abs. 2 BGB, (3) ein hierdurch verursachter Schaden sowie (4) fehlende Exkulpation nach § 311 Abs. 1 Satz 2 BGB, wonach der Schädiger die Beweislast für fehlendes Verschulden trägt.

III. Typische Fallgruppen

In der Praxis relevant sind insbesondere die Verletzung von Aufklärungspflichten über wertbildende Faktoren beim Unternehmens- oder Grundstücksverkauf, der grundlose Abbruch von Vertragsverhandlungen nach Erweckung berechtigten Vertrauens auf den Vertragsschluss sowie die Verletzung von Körper, Gesundheit oder Eigentum eines potenziellen Vertragspartners bereits im Vorfeld des Vertragsschlusses, etwa der vielzitierte Sturz über eine herabgefallene Salatblatt-Verpackung im Selbstbedienungsladen (vgl. BGH, Urteil vom 28.01.1976 – VI ZR 104/74).

IV. Rechtsfolge

Der Geschädigte ist gemäß §§ 249 ff. BGB grundsätzlich auf das negative Interesse zu stellen: Er ist so zu stellen, wie er stünde, wäre er nie in die Vertragsverhandlungen eingetreten. Ein Anspruch auf Erfüllung des – mangels Vertragsschluss nicht existentem – Hauptvertrages besteht demgegenüber grundsätzlich nicht.

Hinweis: Dieser Text dient der allgemeinen Information und stellt keine Rechtsberatung im Einzelfall dar.

¹ Culpa in contrahendo (c.i.c.), lat. für „Verschulden bei Vertragsverhandlungen“; von Rudolf von Jhering 1861 entwickeltes Rechtsinstitut, das eine Haftung für Pflichtverletzungen bereits im vorvertraglichen Stadium begründet – also bevor überhaupt ein wirksamer Vertrag zustande gekommen ist. Seit der Schuldrechtsreform 2002 ist sie in § 311 Abs. 2 i.V.m. § 241 Abs. 2 und § 280 Abs. 1 BGB ausdrücklich kodifiziert.

Konsequenzen Schlechter Qualität

Halluzinationen

Falsche oder erfundene Antworten durch unvollständige Eingabedaten.

Fehlentscheidungen

Unvollständige Informationen führen zu falschen Schlussfolgerungen.

Vertrauensverlust

Anwender verlieren das Vertrauen in die KI und akzeptieren sie nicht mehr.

Informationsverlust bleibt oft unbemerkt

- Es reicht nicht, nur sichtbaren Text zu extrahieren**
- Die semantische Struktur muss erhalten bleiben.**

Verlust fachlicher Informationen

Relevante Details aus Tabellen und Diagrammen gehen verloren.

Fehlinterpretationen

Falsche oder unvollständige Antworten durch fehlende Kontextdaten.

Sinkende Wissensqualität

Verlust von Quellen und Kontextinformationen schwächt die Wissensdatenbank.

Vendor Lock-in



Eingeschränkte Verhandlungsmacht

Der Anbieter kann Preise erhöhen oder Konditionen ändern, ohne dass der Kunde leicht ausweichen kann.



Hohe Wechselkosten

Migration zu einem anderen System ist teuer und aufwändig: Datenkonvertierung, Schulungen, Prozessanpassungen.



Abhängigkeit von der Roadmap

Weiterentwicklung, Support und Preisgestaltung liegen vollständig in der Hand des Anbieters.



Risiko bei Anbieterproblemen

Insolvenz, Strategiewechsel oder Support-Einstellung treffen den Kunden besonders hart.



Kurz gesagt: Der Kunde verliert Unabhängigkeit und Verhandlungsspielraum, während der Anbieter mehr Kontrolle gewinnt.

Das Blackbox-Problem

Kunde

Lädt Dokumente hoch (Word, PDF usw.) und sendet sie an den Anbieter.

- Eigene Dokumente
- Kein Einblick in Verarbeitung
- Nur Antworten als Output

Anbieter (Blackbox)

Erzeugt intern Chunks aus den Kundendaten. Die Chunks verbleiben beim Anbieter.

- Proprietäre Chunk-Erstellung
- Daten liegen beim Anbieter
- Vollständiger Vendor Lock-in

Chunking mit Octopus

Der souveräne Ansatz

Kunde (Schritte 1–4)

01. Dokumente hochladen

Word, PDF und weitere Formate werden in Octopus geladen.

02. Strategie wählen

Z. B. *Split by Sentence*, *by Paragraph* oder *by Fixed Length*.

03. Chunks erhalten

Der Kunde bekommt die fertigen Chunks als Dateien zurück.

04. Vektordatenbank befüllen

Die eigene Datenquelle wird mit den selbst erstellten Chunks bestückt.

Anbieter (Schritt 5)

Der Kunde verbindet seine eigene Datenquelle mit dem gewählten Sprachmodell und kann produktiv arbeiten – ohne Abhängigkeit.

- ✔ Volle Datenkontrolle
 - ✔ Freie Anbieterwahl
 - ✔ Berücksichtigung semantischer Struktur
- u.v.m

„Eine KI ohne Quellenangabe ist nur eine gut formulierte Behauptung.“

Fehlende Zitierfähigkeit

Problem:

Während der Konvertierung oder Datenaufbereitung gehen Seitenreferenzen, Kapitelverweise oder andere Quelleninformationen verloren.

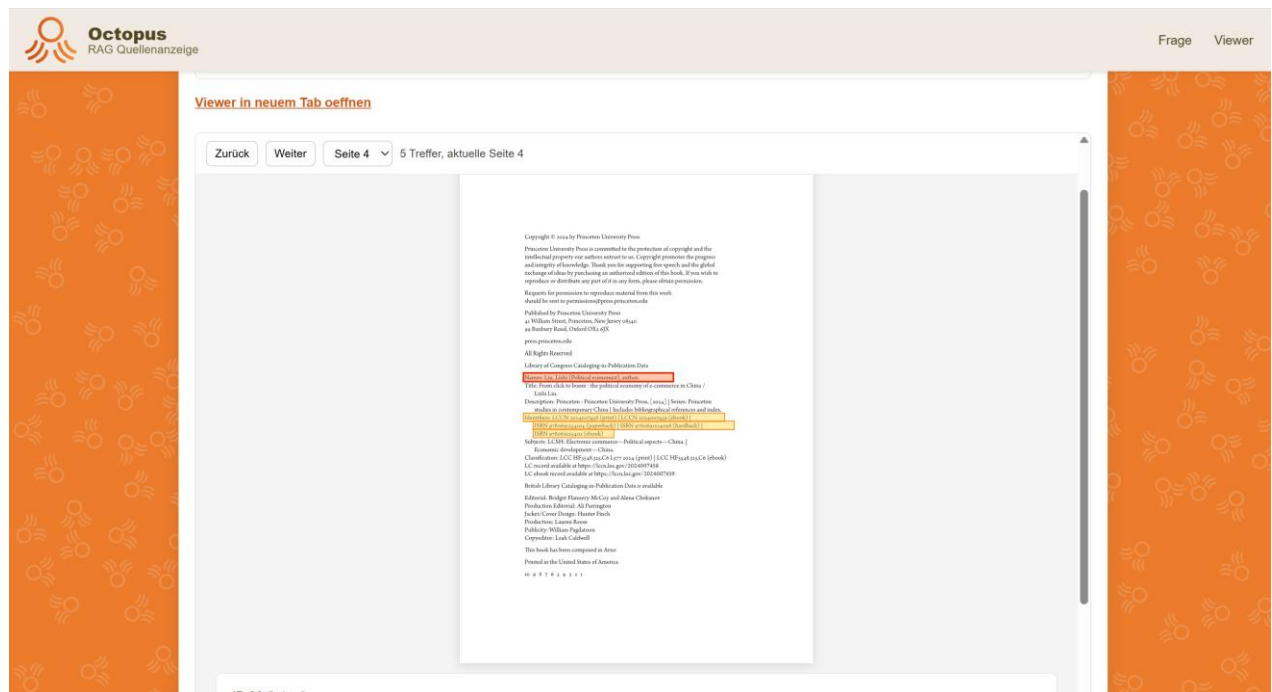
Folgen:

- > Antworten können nicht auf die Originalquelle zurückgeführt werden. Wissenschaftliches Arbeiten wird erschwert oder unmöglich.
- > Quellenangaben in KI-Antworten sind unvollständig oder fehlen.
- > Die Qualität der Wissensbasis sinkt.
- > Halluzinationen können schwerer gefunden werden.
- > Vertrauensverlust beim Nutzer

Zitierfähigkeit

Aktueller Stand

- Einfache RAGs zeigen die Chunks aus denen die Antwort generiert wurden.
- Bessere das dazugehörige Dokument. Z.B. 200 Seiten PDF
- Gute zeigen nur die Seiten.
- **Professionelle die Stelle**



Datenaufbereitung

Die Datenaufbereitung erfolgt über externe KI Anbieter wie OpenAI, Google oder Mistral. Die RAG Systemanbieter sagen das meist nicht, bzw. machen die Tragweite nicht klar! Selbst wenn dann ein lokales LLM verwendet wurde sind die Daten längst dort.

=> ALLE DATEN GEHEN GEORDNET AN DIE GROSSEN KI FIRMEN

- > Verlust der Kontrolle über den gesamten Wissensbestand
- > Compliance und Datenschutzrisiken
- > Abhängigkeit von externen Dienstleistern

Wie geht es besser?

Daten nur lokal verarbeiten. Eine Studie dazu haben wir erstellt und wird demnächst bei www.data2type.de und www.octopus.service.de veröffentlicht.

Wer aus Faulheit oder Unwissenheit Urheberrechte verletzt, ist keine KI Innovation. Er ist ein Haftungsrisiko.

Bei Fragen, gerne fragen!
www.octopus-service.de



Im Breitspiel 6a
69126 Heidelberg
T: +49-(0)6221-7391264
[E: montero@data2type.de](mailto:montero@data2type.de)



Karlstr. 3 -5
76133 Karlsruhe
T: +49-721-402485-12
[E: mehrschad.zaeri@parsQube.de](mailto:mehrschad.zaeri@parsQube.de)