



KI-GESTÜTZTER WISSENSBOT FÜR VERSICHERUNGSBEDINGUNGEN

ZUVERLÄSSIGE ANTWORTEN UND QUELLEN AUF BASIS KOMPLEXER VERTRAGSDOKUMENTE UND DATENBANKEN

AUSGANGSSITUATION

Der Kunde setzte eine Chatbot-Lösung auf Basis von Open WebUI ein, um Versicherungsbedingungen aus PDF Dokumenten durchsuchbar zu machen. Das integrierte Retrieval sollte relevante Textstellen finden und mit Quellen belegen. In der Praxis blieben die Ergebnisse jedoch hinter den Erwartungen zurück: Die Antworten waren häufig ungenau, Quellen passten inhaltlich nicht, Tabellen aus den Dokumenten wurden unvollständig oder falsch interpretiert. Ziel des Projekts war eine deutlich höhere Antwortqualität bei nachvollziehbarer und korrekter Quellenzuordnung. Voraussetzung dafür waren eine saubere Verarbeitung komplexer PDF Strukturen sowie eine robuste Anbindung an die bestehende Open WebUI Umgebung.

VORGEHEN

Zu Beginn sammelte unser Consultant strukturiertes Anwender Feedback zur bestehenden Chatbot Nutzung und optimierte darauf aufbauend das System Prompting, um kurzfristig die Antwortführung zu verbessern. Parallel erfolgte eine detaillierte Analyse der Schwächen des integrierten Retrievals. Anschließend wurde eine vollständig eigenständige RAG Pipeline konzipiert und implementiert, die das Standard Retrieval ablöst.

Im nächsten Schritt wurde eine PDF Ingestion entwickelt, die Fließtexte und Tabellen getrennt extrahiert, Überschriftenhierarchien erkennt und Tabellenstrukturen explizit abbildet. Darauf aufbauend wurde eine kombinierte Suche aus Vektor- und Volltextsuche umgesetzt, deren Ergebnisse über Reciprocal Rank Fusion zusammengeführt und per Cross Encoder nachgerankt werden. Für komplexe Nutzerfragen wurde eine automatische Zerlegung implementiert. Eine Retrieval Critique prüft, ob die gefundenen Passagen als Antwort ausreichen, und startet bei Bedarf weitere Suchläufe. Ergänzend entstand eine Ingest API zur Dokumentenverwaltung sowie ein Evaluations Framework zur automatisierten Qualitätsmessung. Die gesamte Lösung wurde containerisiert und über eine Manifold Pipeline in Open WebUI integriert.

ERGEBNIS

Das Projektergebnis ist eine produktiv einsetzbare, eigenständige RAG Pipeline, die qualitativ belastbare Antworten auf Fragen zu Versicherungsbedingungen liefert. Die Antwortqualität ist durch das integrierte Evaluations Framework mess- und reproduzierbar abgesichert, die Quellenzuordnung fachlich konsistent. Tabellen und strukturierte Inhalte aus PDFs werden korrekt berücksichtigt, auch bei komplexen Fragestellungen. Für den Kunden steht damit eine stabile Such- und Antwortlogik zur Verfügung, die unabhängig von den Einschränkungen des Standard Retrievals arbeitet und sich gezielt weiterentwickeln lässt. Im Arbeitsalltag profitieren die User von präziseren Antworten und einer deutlich besseren Nachvollziehbarkeit der zugrundeliegenden Dokumentstellen.

VERSICHERUNG

Die bestehende Chatbot Lösung des Kunden lieferte bei der Suche in Versicherungsbedingungen ungenaue Antworten und unpassende Quellen. Durch eine systematische Prompt Optimierung und den Aufbau einer eigenständigen RAG Pipeline wurde die Dokumentensuche qualitativ neu aufgestellt. Das Ergebnis ist eine belastbare, evaluierbare Grundlage für präzise und passende Antworten auf komplexe Fachfragen.

TECHNOLOGIEN & METHODEN

- **Python, FastAPI**
- **PostgreSQL mit pgvector**
- **Deutschoptimierte Sentence Transformer Embeddings**
- **Cross Encoder Reranking**
- **Kombinierte Vektor- und Volltextsuche mit Reciprocal Rank Fusion**
- **PDF Parsing mit separater Text und Tabellenextraktion**
- **Retrieval Critique und automatische Anfragezerlegung**
- **Evaluations Framework für RAG Qualität**
- **Docker / Podman (Containerisierung)**
- **Open WebUI mit Manifold Pipeline Anbindung**