Privates IT-Praxisprojekt

Lumos AI und private Wissenssuche

Selbst gehostete KI-Oberfläche mit lokaler Modellanbindung und einer kontogebundenen Suche im Nextcloud-Wissensspeicher.

Stand 09.11.2026 | WebUI und Modellcontainer geprüft

Ziel und Nutzen

Eine eigene KI-Oberfläche soll allgemeine Gespräche und den gezielten Zugriff auf freigegebenes Wissen verbinden. Das Projekt konzentriert sich auf eine nachvollziehbare Weiterleitung der Anfragen und eine klare Grenze zwischen Modellantwort und tatsächlich abgerufenem Inhalt. Private Quellen sollen nur dem zugeordneten Benutzer zur Verfügung stehen.

Technischer Aufbau

Open WebUI stellt die Chatoberfläche bereit. Ollama bildet die lokale Modellanbindung. Ein eigener Router ergänzt Funktionswege, darunter die Wissenssuche. Für diese Suche verwendet das Backend eine geprüfte Kontozuordnung und einen gesonderten Nextcloud-Zugang. Markdown-Seiten und geeignete PDF-Anhänge werden unter den jeweiligen Rechten gelesen.

Open WebUI und Ollama	Chatoberfläche und lokale Modellausführung
Wissensrouter	Auswahl des passenden Funktionswegs und Quellenangaben
PDF-Worker	Begrenzte Textextraktion aus zugelassenen Anhängen

Umsetzung

1. Die bestehende WebUI wurde unter dem Namen Lumos gestaltet und an die lokale Modellumgebung angebunden.
2. Die Wissenssuche verwendet die authentifizierte Backend-Identität und eine ausdrückliche Cloud-Kontozuordnung.
3. Berechtigte Wissensseiten und PDF-Anhänge werden anhand des Anliegens durchsucht. Quellen verweisen auf tatsächlich abgerufene Inhalte.
4. Für PDF-Verarbeitung sind Größen-, Seiten-, Zeichen- und Laufzeitgrenzen sowie kontogetrennte Zwischenspeicherung eingerichtet.

Lumos AI und private Wissenssuche

Prüfung und Ergebnis

Die Einrichtungsprüfungen umfassen echte Wissensfragen, Quellenverweise, fehlende Treffer und die Abweisung unzugeordneter Benutzer. Zusätzlich wurde geprüft, dass erfundene Benutzerkennungen im Anfrageinhalt die Kontozuordnung nicht ersetzen. Am 9. November 2026 liefen WebUI und Modellcontainer. Die öffentliche WebUI antwortete erfolgreich über HTTPS.

Betrieb und Wartung

Für eine Fehlersuche werden Anmeldung, Kontozuordnung, Nextcloud-Zugriff und die konkrete Quelldatei geprüft. Die aktuellen Berechtigungen werden bei erneuter Nutzung einer PDF-Auswertung berücksichtigt. Änderungen am Router werden zusammen mit WebUI-Datenbank und Laufzeitkonfiguration gesichert.

Grenzen und nächste Schritte

Die dokumentierte Wissenssuche verwendet Stichwörter und Wortvarianten. Eine semantische Vektorsuche ist im beschriebenen Stand nicht eingerichtet. PDF-Scans benötigen eine Textebene oder vorgelagerte Texterkennung. Wetter- und Bildrouten hängen von der jeweiligen Konfiguration und den erreichbaren Anbietern ab und wurden für diese Dokumentation nicht neu ausgeführt.

Praktische Kenntnisse

Lokale KI-Dienste, Funktionsrouting, Quellenbezug, kontogebundene Autorisierung, PDF-Verarbeitung und Begrenzung von Hintergrundaufgaben.

Verknüpfte Projekte

Nextcloud - PaynsCloud · Schiedsrichterportal mit Cloudablage · Linux-Server mit Docker

Projektseite und technische Quellen

Diese Dokumentation online
https://payns.info/dokumentation/lumos.html

Open WebUI Pipe Functions
https://docs.openwebui.com/features/plugin/functions/pipe/

Ollama-Dokumentation
https://docs.ollama.com/

Die Umsetzung und die Aufbereitung der Dokumentation wurden KI-gestützt unterstützt. Die Beschreibung stützt sich auf die vorhandene Projektkonfiguration, Betriebsnachweise und die genannten Prüfungen.