

— NEXTCLOUD SCHLESWIG-HOLSTEIN TAG 2026 • KIEL

LLMs sicher mit dem Nextcloud Assistant verbinden.

Tadeusz Nikitin

Founder, Splitbot GmbH · Geschäftsführer, Kontor Business IT GmbH

kbit.de splitbot.ai



LINKEDIN



Von der Oberfläche bis unter die Motorhaube.

01

Die Oberfläche

Was der Anwender vom Assistant tatsächlich sieht.

02

Die Fähigkeiten

Text, Sprache, Bilder — und was davon gar nicht das LLM macht.

03

Die Architektur

Wer redet mit wem, und wo genau hört Nextcloud auf.

04

Der Aufwand

Was davon geklickt wird und was Betrieb bedeutet.

05

Das Modell

Woher es kommt — und wem die Daten dabei gehören.

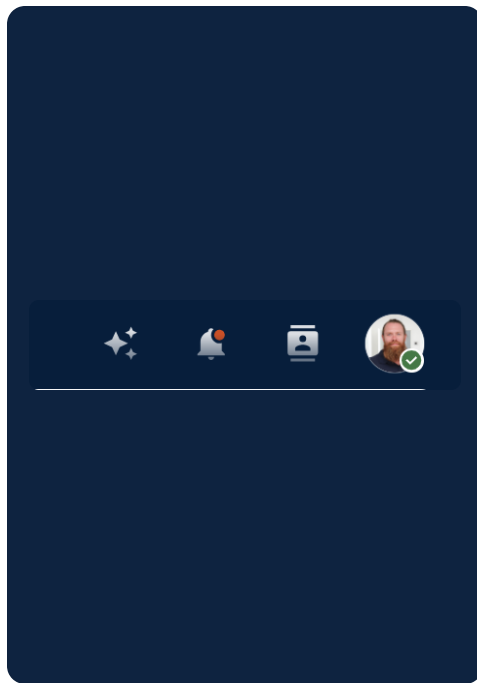
06

Das Wissen

Wie unsere eigenen Daten überhaupt in die Antwort kommen.

Was ist überhaupt der Nextcloud Assistant?

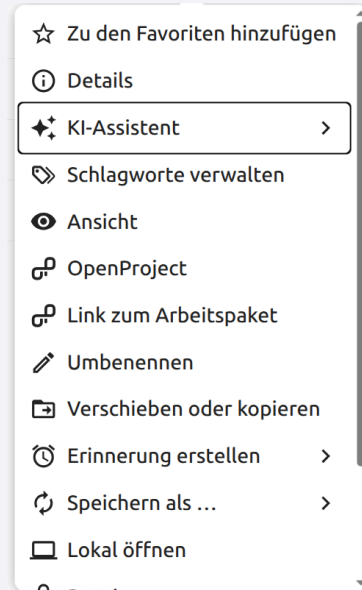
Keine Technik, keine Architektur. Erst einmal nur: Wo taucht er auf und wie fühlt er sich an?



01

Immer in der Kopfzeile

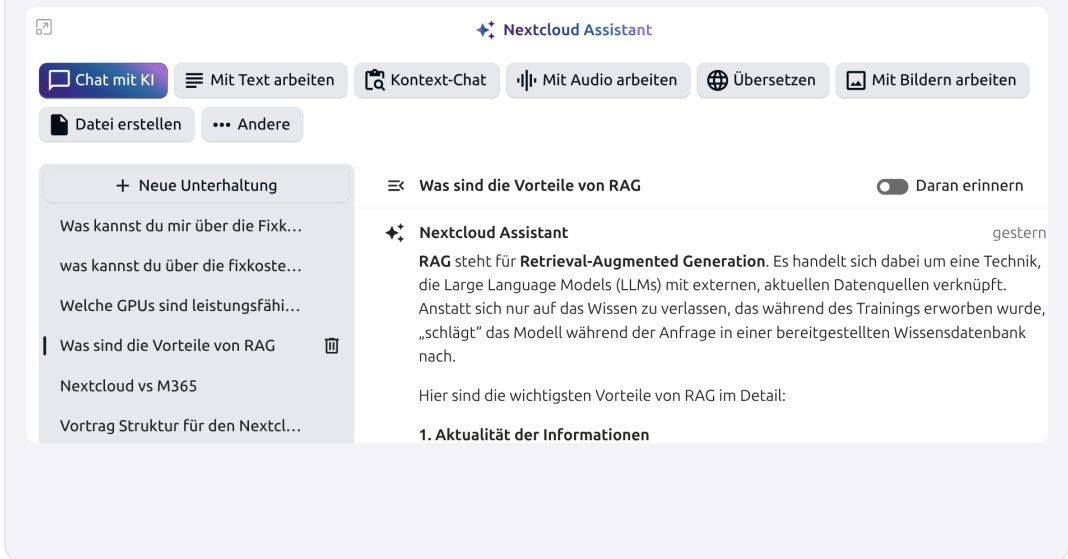
Ein Klick, aus jeder App heraus.



02

Direkt am Dokument

Im Kontextmenü, dort wo die Datei liegt.



03

Ein Dialog für alles

Chat, Text, Audio, Übersetzen, Bilder — plus Verlauf.

Was kann der Assistant **konkret?**

☰ Text erstellen ⓘ

Eine Aufgabe beschreiben, die der Assistant erledigen soll, oder eine Frage stellen

Lübeck bei Nacht

Der vom Assistenten erstellte Text

Lübeck bei Nacht hat einen ganz besonderen, fast magischen Charme. Wenn die Tagestouristen weg sind und die warme Straßenbeleuchtung die mittelalterlichen Backsteinfassaden und die spitzen Giebel der Altstadt anstrahlt, verändert sich die Atmosphäre der Hansestadt grundlegend.

Hier sind verschiedene Facetten, wie man Lübeck bei Nacht erleben kann:

1. Die romantische Architektur-Tour (Spaziergang)

Kopieren

⚠ Dieses Ergebnis wurde von KI generiert. Überprüfe die Angaben und passe sie gegebenenfalls an.

01

Text

- Zusammenfassen, schreiben, umformulieren
- Übersetzen, freie Fragen im Chat

Audioeingang

Von Gerät hochladen Aus dem Speicher auswählen Aufnahme beginnen

▶ 0:00 / 0:13 🔊 ⋮ ✕

Der transkribierte Text

Das sind Notizen zu einem wichtigen Gespräch der Geschäftsführung der Kontor Business IT. Wir sitzen hier heute Abend zusammen, um folgenden Sachverhalt zu klären.

Kopieren

⚠ Dieses Ergebnis wurde von KI generiert. Überprüfe die Angaben und passe sie gegebenenfalls an.

02

Sprache

- Audio hochladen oder direkt aufnehmen
- Transkript geht anschließend ans LLM



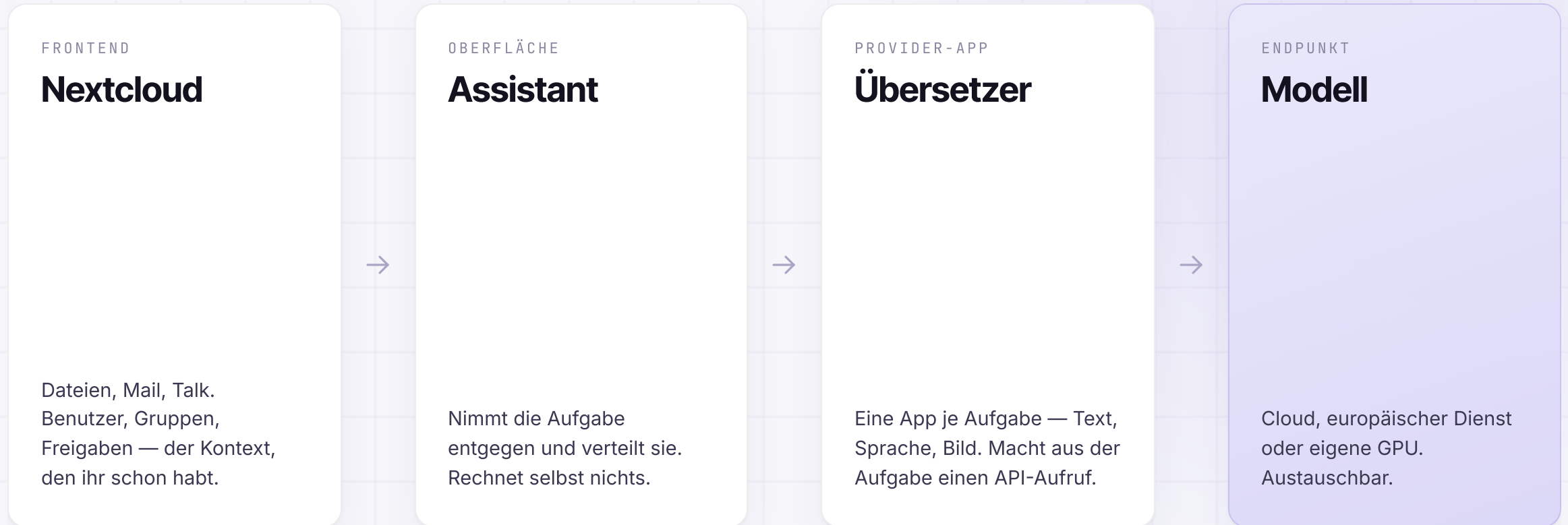
03

Bilder

- Bildverständnis / Vision
- Texterkennung in Dokumenten
- Bildgenerierung (optional)

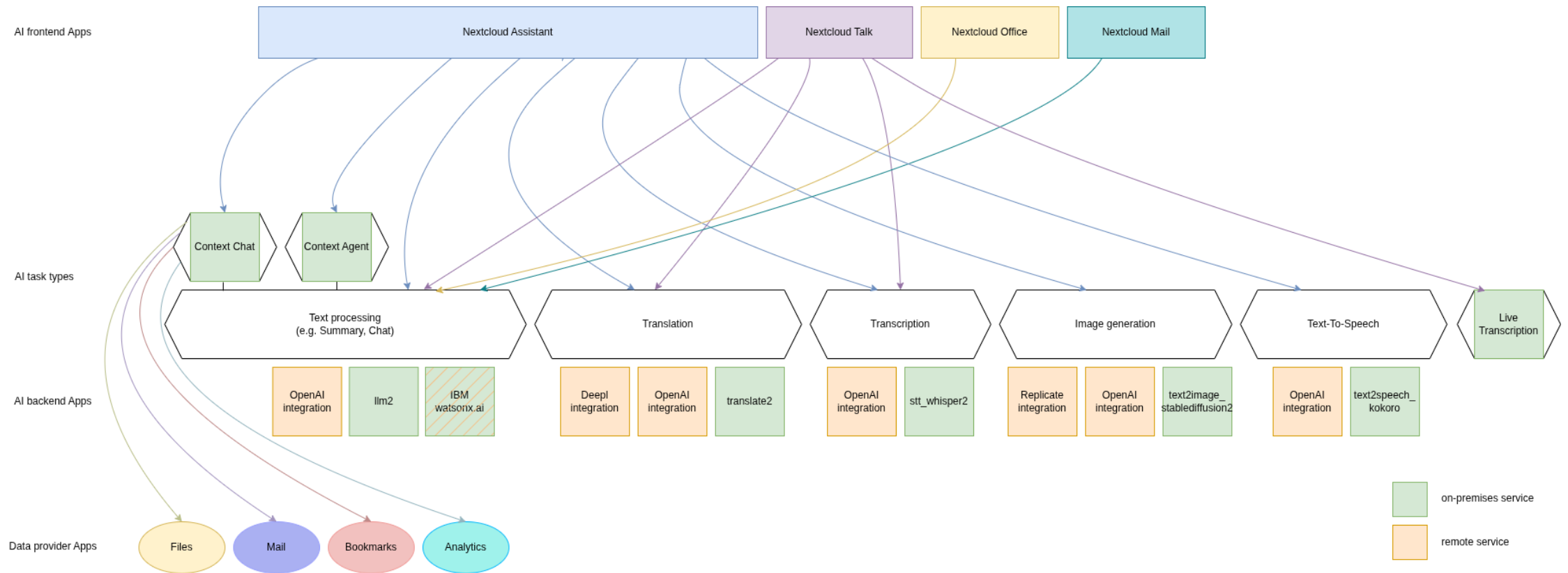
Je nach eingerichtetem Provider

Wie bekomme ich das **zum Laufen?**



**Nextcloud ist Frontend und Integrationsschicht.
Das Modell darf woanders laufen.**

Vier Ebenen: Frontend, Task, Backend, Datenquelle.



Grün = on-premises

Orange = remote

Quelle: Nextcloud Administration Manual – AI Overview · docs.nextcloud.com

Vom Klick zum Plattformbetrieb.

STUFE 01

Nextcloud-Admin

assistant

Aus dem App Store installieren und aktivieren. Fertig.



STUFE 02

Admin mit
Dienstverständnis

integration_openai

Endpunkt, Key, Modellwahl, Quotas. Noch Oberfläche — aber man muss wissen, was man einträgt.



STUFE 03

Systemadministration

AppAPI / ExApps

Zugriff aufs Betriebssystem. Container, Deploy-Daemon, Netz, Zertifikate.



STUFE 04

Betrieb & Datenhaltung

Context Chat / Agent

Vektorindex, Background Jobs, Indexierungslauf, Versionsgleichstand, Ressourcen.



STUFE 05

Eigene KI-Plattform

LLM selbst betreiben

GPU-Betrieb, Treiber, Inference-Server, Kapazitätsplanung, Updates.



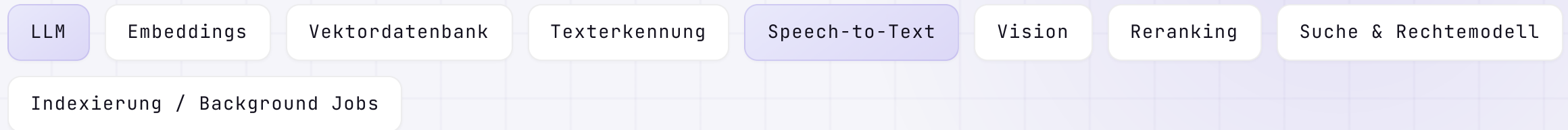
GEKLICKT



BETRIEBEN

Alles davon ist gut dokumentiert. Nichts davon ist eben mal geklickt.

LLM $\neq$ KI-Infrastruktur

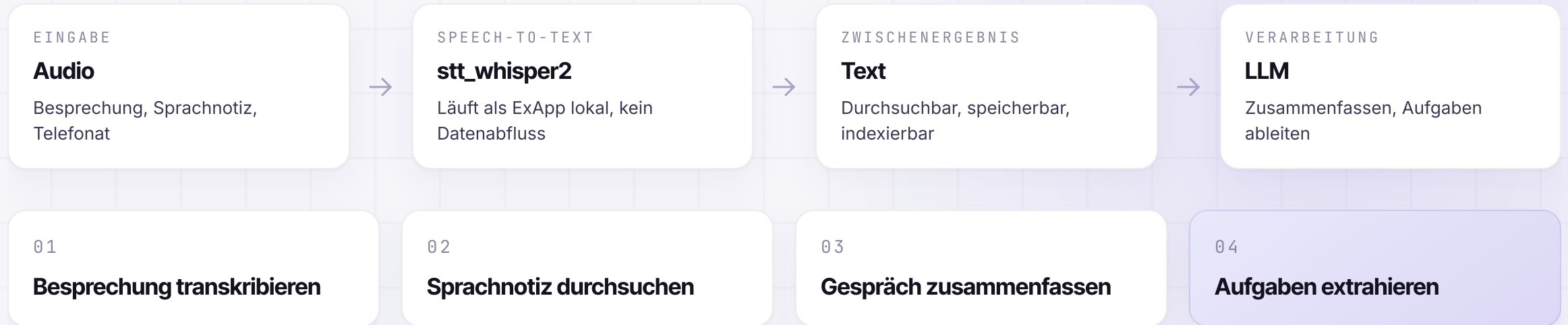


**„Ein 70B-Modell hilft mir überhaupt nicht,
wenn es meine E-Mails nicht kennt.“**

DAS MODELL IST EINE KOMPONENTE – NICHT DAS SYSTEM

- NICHT ALLES MACHT DAS LLM

Beispiel: Voice Transcription



Wo bekomme ich **mein Modell** her?

01

⚠ RECHTLICH PRÜFEN

Cloud

OpenAI & Co. Schnell startklar, keine eigene Hardware.

- Beste Verfügbarkeit
- Daten verlassen das Haus
- Vertragsfrage, keine Technikfrage

02

⚠ RECHTLICH PRÜFEN

Europäische Anbieter

Inference in europäischer Infrastruktur, betrieben nach EU-Recht.

- Kompromiss aus Aufwand und Kontrolle
- Meist OpenAI-kompatible API
- Kein eigener GPU-Betrieb nötig

03

Selbst hosten

Open-Weight-Modelle auf eigener Hardware.

- Volle Kontrolle über die Daten
- GPU, Betrieb und Updates liegen bei uns
- Modell jederzeit austauschbar

WEG 03 KOSTET HARDWARE

Was dahinter steht

INFERENCE

RTX A6000

48 GB VRAM je Karte, mehrere Karten im Verbund

ARBEITSSPEICHER

1 TB RAM

Nicht für das Modell — für Indexierung, Queues, Parallelität

STORAGE

Ceph, Flash-only

Vektorindex und Dokumente wollen IOPS, keine Kapazität

Nextcloud legt mich nicht auf ein Modell fest.

GEMEINSAMER NENNER IST DIE OPENAI-KOMPATIBLE SCHNITTSTELLE

Technisch möglich ist nicht **rechtlich zulässig**.



Nur weil ich ein Modell anbinden kann, darf ich meine Unternehmensdaten noch lange nicht dorthin schicken.

Zu klären, bevor der Endpunkt eingetragen wird

- Auftragsverarbeitung: Vertrag, Weisungsbindung, Unterauftragnehmer
- Wo verarbeitet der Anbieter tatsächlich — und wer hat Zugriff darauf?
- Werden Eingaben zum Training verwendet? Wie lange werden sie gespeichert?
- Mitbestimmung: Betriebsrat oder Personalrat frühzeitig einbinden

Gilt für Weg 01 und Weg 02 — bei Weg 03 verschiebt sich die Frage, sie verschwindet nicht.



Orientierungshilfe der Datenschutzkonferenz

KI-Systeme mit Retrieval Augmented Generation · 18 Seiten · Oktober 2025



Dr. Heinke Bastek

Qualitätsmanagement und Datenschutz
Kontor Business IT GmbH

Ansprechbar beim Networking im Anschluss.

Wie bekommt die KI unser Wissen?

SCHRITT 01

Frage

„Dienstleistungen der Kontor rund um die Nextcloud?“

Assistant · Oberfläche



SCHRITT 02

Suche

Produktblätter, Anleitungen, Verträge, Angebote

Embeddings · Vektorindex · ExApp



SCHRITT 03

Kontext

Vier Dokumente statt der ganzen Ablage

Nextcloud · Gruppen und Freigaben



SCHRITT 04

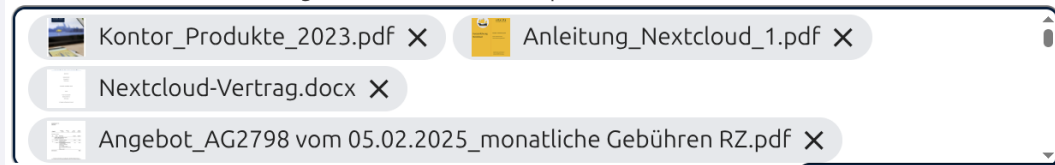
Antwort

Leistungsübersicht — jede Aussage belegt

Provider-App · Modell nach Wahl

Die Antwort nennt ihre Quellen

The sources referenced to generate the above response



context_chat + context_chat_backend

Rechtemodell ist Pflicht

Der Benutzer darf durch die KI nicht plötzlich Dinge wissen, die er vorher nicht wissen durfte.

Der Index kennt Gruppen und Freigaben. Datenschutz entsteht hier, nicht im Modell.

Nextcloud bleibt die Ablage — die **KI kommt von außen.**

Statt Modelle in Nextcloud zu betreiben

- **Sync über WebDAV, pro Benutzer mit dessen eigenen Zugangsdaten.** Die Plattform kann nichts holen, was der Benutzer nicht selbst sehen darf.
- **Eigene Vektordatenbank je Benutzer.** Kein gemeinsamer Index, der zur Abfragezeit gefiltert werden müsste.
- **Nächtlicher Vollabgleich.** Entzogene Freigaben und gelöschte Dateien fliegen aus dem Index.
- **Gemeinsames LDAP für die Identität.** Autorisierung bleibt trotzdem bei Nextcloud.

Antworten verweisen per Link zurück in die Nextcloud. Die Quelle ist die Datei — nicht deren Kopie.



Der ehrliche Preis: eine zweite Plattform, ein zweiter Betrieb, ein zweiter Vertrag. Die Fragen von vorhin gelten wieder — diesmal mit uns als Anbieter.

01

Der Benutzer verbindet selbst

02

Quelle zeigt zurück in die Nextcloud

01

**Die Daten sind wertvoller
als das Modell.**

02

Das Modell ist austauschbar.

03

**Digitale Souveränität ist eine
Architekturentscheidung.**

Fragen?

KONTAKT

Tadeusz Nikitin

Founder, Splitbot GmbH

Geschäftsführer, Kontor Business IT GmbH

kbit.de splitbot.ai



• VERNETZEN