



Technisches Datenblatt

Für IT-Verantwortliche und Datenschutzbeauftragte: Infrastruktur, Datenverarbeitung und Schutzmaßnahmen der Plattform Alvar Knowledge — das Dokument hinter den Fußnoten 4 und 5 unserer Website.

i. Hosting & Infrastruktur

Rechenzentrum	Google Cloud Platform, EU-Region europe-west1 . Sämtliche Speicherung und Verarbeitung erfolgt ausschließlich in der EU; es findet keine Übermittlung von Kundendaten in Drittländer statt.
Architektur	Containerisierte Dienste (Docker); Web-Anwendung, API-Backend (Python/FastAPI), asynchrone Verarbeitung (Celery), relationale Datenbank (PostgreSQL 16), Vektor-Suchindex (Qdrant) und Cache/Queue (Redis), jeweils logisch getrennt je Mandant.
Dokumentablage	Originaldokumente liegen unverändert in einem zugriffsbeschränkten Cloud-Storage-Bucket (EU); Verschlüsselung at rest und in transit (TLS 1.2+).
Geheimnisse	Zugangsdaten und Schlüssel werden zentral im Secret Manager verwaltet; kein Klartext in Code oder Konfiguration.
Backups	Automatisierte Datenbank-Sicherungen mit definierter Aufbewahrung, Wiederherstellung regelmäßig erprobt.

ii. KI-Verarbeitung

Modelle	Sprachmodelle der Gemini-Familie über Vertex AI (Google Cloud, EU-Verarbeitung). Die Modellanbieter verwenden Kundendaten vertragsgemäß nicht zum Training .
Dokumentverarbeitung	Mehrstufige Pipeline: native PDF-Extraktion → Layout-Analyse für komplexe Dokumente → OCR für Scans → Vision-Modelle nur als letzte Stufe für schwer lesbare Vorlagen. Kostenkontrolle pro Verarbeitungsaufgabe mit Budgetgrenzen.
Suche & Antworten	Hybride Suche (semantische Vektoren + Volltext) mit Reranking; jede Antwort trägt Quellenverweise, deren Existenz systemseitig validiert wird — nicht belegbare Aussagen werden verworfen.

iii. Anonymisierung (Common Pool)

- Mehrstufige Erkennung personenbezogener Daten: Microsoft Presidio kombiniert mit spaCy-Sprachmodellen für Deutsch und Englisch sowie eigenen, domänenspezifischen Erkennungsmustern.
- Zwingende Unterdrückung von Namen, Organisationen und Kontaktdaten; konfigurierbare feldgenaue Unterdrückung weiterer Angaben.
- Kontextuelle Generalisierung statt bloßer Schwärzung: Datum→Jahr, Budget→Spanne, Region→Land, Personenzahlen→gerundet.

- Re-Identifikations-Risiko-Bewertung auf Basis von k-Anonymität mit expliziten Warnungen vor der Freigabe.
- Ausschlussregeln für Konfliktregionen und sensible Sektoren, zweifach geprüft (Vorschau und Bestätigung), nicht umgebar.
- Menschliche Freigabe vor jeder Veröffentlichung; Originaldokumente verlassen die Organisation zu keinem Zeitpunkt — geteilt werden ausschließlich anonymisierte, strukturierte Projektobjekte. Jede Freigabe ist widerruflich.

iv. Zugriff, Mandantentrennung & Protokollierung

- Rollenbasierte Zugriffe (Vorstand/Administration, Mitarbeitende, Partner); Authentifizierung mit signierten Tokens.
- Datenbankseitige Zugriffsabsicherung bis auf Zeilenebene (Row-Level-Security) — Berechtigungen werden in der Datenbank erzwungen, nicht nur in der Anwendung.
- Strikt getrennte Datenhaltung je Organisation; der Common Pool ist ein separater, ausschließlich anonymisierter Datenbestand.
- Lückenloser Audit-Trail über Zugriffe, Freigaben und Exporte — auswertbar für Vorstand, Prüfung und Fördergeber.
- Strukturiertes Monitoring und Protokollierung des Systembetriebs inklusive Kosten- und Qualitätsüberwachung der KI-Aufrufe.

v. Auftragsverarbeiter

Google Cloud EMEA

Infrastruktur, Speicherung, KI-Verarbeitung (Vertex AI) — Verarbeitung in der EU-Region europe-west1, EU-Standardvertragsklauseln bzw. EU-Datenverarbeitungsbedingungen.

Weitere

Eine vollständige, aktuelle Liste der Unterauftragsverarbeiter stellen wir im Rahmen des Auftragsverarbeitungsvertrags (Art. 28 DSGVO) bereit.

Dieses Datenblatt beschreibt den technischen Stand zum angegebenen Datum; Änderungen, die das Schutzniveau wahren oder erhöhen, bleiben vorbehalten. Für Auftragsverarbeitungsvertrag, TOMs nach Art. 32 DSGVO und Detailfragen: y.bekri@dail.info.