1 Leistungsbeschreibung – Lector Hybrid-LLM-Dienste

Stand: 05.02.2026

1.1 1. Vertragsgegenstand

Bereitstellung von Cloud-basierten KI-Sprachmodellen (Large Language Models) zur Integration in lokale Lector-Installationen des Auftraggebers.

1.2 2. Leistungsumfang

1.2.1 2.1 Technische Komponenten

- **LLM-Router**: Installation und Konfiguration auf Servern des Auftraggebers
- **API-Endpunkt**: `https://llm.lector.ai/v1`
- **Authentifizierung**: API-Key basiert, OAuth 2.0 kompatibel
- **Verschlüsselung**: TLS 1.3 für alle Verbindungen

1.2.2 2.2 Verfügbare Modelle

Die verfügbaren Modelle werden vom Auftragnehmer festgelegt und können folgende Anbieter umfassen:

- Google Vertex AI
- Microsoft Azure OpenAI
- Weitere nach Vereinbarung

1.2.3 2.3 Nicht im Leistungsumfang enthalten

- Bereitstellung der lokalen Infrastruktur des Auftraggebers
- Internetanbindung des Auftraggebers

1.3 3. Service Level

1.3.1 3.1 Verfügbarkeit

Komponente	Verfügbarkeit	Messperiode
LLM-Router (zentral)	99,5 %	Monatlich

Komponente	Verfügbarkeit	Messperiode
API-Endpunkt	99,5 %	Monatlich

Ausgenommen: Wartungsfenster (werktags 18–22 Uhr, Wochenende), Ausfälle der LLM-Anbieter, Internetverbindung des Auftraggebers.

1.3.2 3.2 Reaktionszeiten

Priorität	Beschreibung	Reaktionszeit	Wiederherstellung
Kritisch	Totalausfall	2 Stunden	8 Stunden
Hoch	Wesentliche Einschränkung	4 Stunden	24 Stunden
Normal	Teileinschränkung	8 Stunden	48 Stunden

1.3.3 3.3 Support

- **Supportzeiten**: Mo–Fr 09:00–17:00 Uhr (außer Feiertage)
- **Kontakt**: support@lector.ai
- **Ticketsystem**: Über Kundenportal

1.4 4. Leistungskennzahlen

1.4.1 4.1 Durchsatz

- Maximale Anfragen pro Minute: gemäß gebuchtem Paket
- Maximale Token pro Anfrage: modellabhängig (typisch 4.000–128.000)

1.4.2 4.2 Latenz

- Ziel-Latenz (Time to First Token): < 500 ms (95. Perzentil)
- Abhängig vom gewählten LLM-Anbieter

1.5 5. Paketoptionen

Die verfügbaren Pakete, inkludierten KI-Seitenaufrufe, Stundenlimits und Preise sind der jeweils aktuellen **Preisliste des lector.ai Cloudservices** zu entnehmen.

1.6 6. Datenschutz

- Verarbeitung gemäß Anlage 3 des AVV
- Keine Zwischenspeicherung von Inhaltsdaten
- Metadaten-Speicherung nur für Abrechnungszwecke

1.7 7. Mitwirkungspflichten des Auftraggebers

- Bereitstellung stabiler Internetverbindung
- Einhaltung der technischen Voraussetzungen
- Sichere Verwahrung der API-Zugangsdaten

1.8 8. Ergänzende Regelungen

Es gelten ergänzend die Allgemeinen Geschäftsbedingungen (AGB) des lector.ai Cloudservices, insbesondere Abschnitt “Hybrid-LLM-Dienste”.

Anlagen:

- Anlage 1: Technische Schnittstellenbeschreibung (API-Dokumentation)
- Anlage 2: Preisliste Hybrid-LLM-Dienste
- Anlage 3: Auftragsverarbeitungsvertrag (AVV)

© 2026 lector.ai GmbH - Alle Rechte vorbehalten