

Silent AI

Unternehmenswissen produktiv nutzen – lokal, sicher, ohne Cloud.

Silent AI macht das Wissen Ihres Unternehmens per KI-Assistent nutzbar – vollständig On-Premises, mit lückenloser Respektierung der bestehenden Zugriffsrechte. Kein Prompt, kein Token verlässt Ihr Netzwerk.

100%

lokal (On-Premises)

100%

sicher (Off-Cloud)

100%

privat (Rechteverwaltung)

Silent AI ist eine kompakte Appliance auf 2HE, die eine leistungsfähige GPU und schnellen NVMe-Speicher als lokales Storage enthält.



AUF EINEN BLICK



Zero Cloud by Design

Alle Daten und LLM-Inferenz laufen lokal auf der Appliance. Kein Token verlässt das Netz.



Identity Management

AD/LDAP-Integration: Mitarbeiter erhalten nur KI-Antworten zu Inhalten mit Lesezugriff.



Schnelle Integration

Konnektoren für M365, SharePoint, Confluence, Slack, PDF, Fileserver, SQL, ... – Deployment in Tagen.



Planbare TCO

Einmalige Investition. Kein Token-Pricing, keine Abo-Überraschungen, CARE-Wartung inklusive.



Compliance-ready

DSGVO-konform, NIS2-ready, EU AI Act kompatibel. Kein Transfer in US-Infrastruktur.



Deutscher Hersteller

Entwicklung, Fertigung und Support in Deutschland. Kein US CLOUD Act Risiko.

DIE HERAUSFORDERUNGEN



Datenschutzrisiken

Cloud-KI sendet Prompts und Dokumente an US-Server. DSGVO-Konformität und US CLOUD Act sind unvereinbar.



Berechtigungslücken

Cloud-KI erfordert zusätzliches Rechte-Management. Mitarbeiter sehen dadurch evtl. Inhalte ohne entsprechendes Leserecht.



KI-Druck ohne Strategie

Wettbewerber sind produktiver. NIS2 und EU AI Act erzwingen jetzt Entscheidungen zur KI-Governance.

Silent AI macht Unternehmenswissen produktiv nutzbar – heute, lokal und vollständig unter Ihrer Kontrolle.

LÖSUNG UND FUNKTIONEN

Eine vollständig lokale KI-Plattform für Ihr Unternehmenswissen.



Lokaler KI-Assistent

LLM-Inferenz auf dedizierter NVIDIA GPU. Keine Cloud-API-Calls – vollständig air-gapped möglich.



Identity Management

AD/LDAP-Integration: jede Antwort respektiert die bestehende Zugriffsrechtestruktur Ihres Unternehmens.



RAG auf internen Daten

Retrieval-Augmented Generation auf M365, SharePoint, Confluence, Fileserver, SQL und anderen – ohne Halluzination.



Schnelle Konnektoren

Fertige Konnektoren für alle gängigen Dokumentenquellen. Deployment in Tagen, nicht Monaten.



Compliance by Design

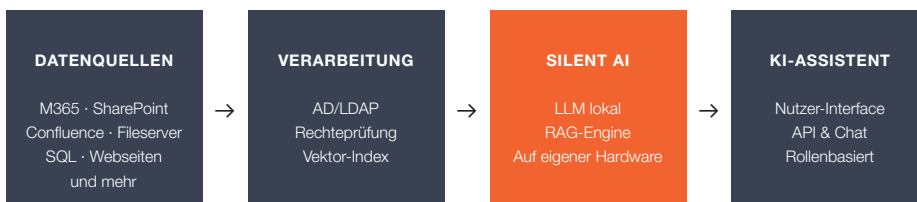
DSGVO-konform. Kein US CLOUD Act Risiko. NIS2-ready, EU AI Act kompatibel. Audit-sicher.



Zukunftssicher

Tauschbare LLMs, die immer leistungsfähiger werden. Bis zu 48 TB Silent AI Vault Storage für lokale Vektor-Daten.

ARCHITEKTUR AUF EINEN BLICK



IHR NUTZEN

- ✓ Kein Datenleck – garantiert by Design
- ✓ Produktivere Mitarbeiter bei Wissensarbeit
- ✓ Geringerer IT-Aufwand durch fertigen Stack
- ✓ Compliance ohne Kompromisse
- ✓ Zukunftssichere, ausbaufähige Plattform

SILENT ASSISTANTS

- Individuelle Kombinationen aus Data Set, Anweisungen und Prompts, z. B.
- Service-Assistent mit Zugriff auf Ticket-System
 - Finanz-Assistent mit Rechtsgrundlagen
 - Arzt-Assistent mit Hintergrundwissen

WARUM SILENT AI

01 ENTERPRISE KI

KI-Nutzung verteilter, sensibler Daten. Kein Cloud-Chatbot.

02 ZERO TRUST

Rechte vererbt aus den Quellsystemen. Ohne Ausnahmen.

03 ON PREMISES

Appliance aus einer Hand. Made in Germany.

04 FLAT RATE

Keine Token-Limits. Volle Kostenkontrolle.

05 NACHWEIS

Jede Antwort mit Quellen. Nachvollziehbar statt halluziniert.

Mehr Informationen und Demo-Termin:

www.fast-lta.de/silent-ai

FAST LTA GmbH | April 2026 | Änderungen vorbehalten



TECHNISCHE DATEN

PLATTFORM	
Bereitstellung	Als schlüsselfertige Appliance, On Premises
GPU	Aktuelle Nvidia GPU, 96 GB, fanless
SPEICHER FÜR LOKALE DATEN	
Storage	Silent AI Vault Speichermodul mit 12x NVMe, Triple Parity
Speicher-Slots	2 Slots für Silent AI Vault
Speicher-Kapazität	12 oder 24 TB (brutto) je Silent AI Vault
NETZWERK & SCHNITTSTELLEN	
Daten	Dual 10GbE (RJ45) oder 10GbE SFP+
Option	Dual 25/100GbE (QSFP)
Management	1x 1GbE Admin 1x 1GbE IPMI
TECHNISCHE MASSE	
Leistung	Idle: ca. 170 W Typisch: ca. 636 W (1 Silent AI Vault aktiv) Maximal: 1.200 W
Wärmeabgabe	Typisch: 2170 BTU/h Maximal: 4094 BTU/h
Maße	2HE x 19 Zoll x 640mm Gewicht ca. 17 kg
KI & INTEGRATION	
Architektur	RAG (Retrieval-Augmented Generation)
Identity Management	AD/LDAP-Integration (vollständig)
Nutzer-Interface	Web-UI · Chat API ab Mitte 2026
Authentifizierung	LDAP, SSO
Konnektoren	MS Office, MS SharePoint, MS Outlook/Exchange, Confluence, Slack, SQL, SMB/File Server, PDF-Upload, Web Scraper Weitere in Entwicklung / auf Anfrage API ab Mitte 2026
COMPLIANCE	
Datenschutz	DSGVO-konform, kein Drittländ-Transfer
Regulatorik	NIS2-ready, EU AI Act kompatibel
Herkunft	Deutsches Produkt · kein US CLOUD Act
Wartung	CARE 1/3/5 Jahre, feste Konditionen
Verfügbarkeit	Sofort lieferbar