The AI comes to your data.
Not the other way round.

A knowledge assistant that answers questions from your own documents — in your own data centre, fully air-gapped or on your workstations. **Your documents stay where they are. We do not store questions or answers.**

IN HOUSE – AND THAT IS WHERE IT STAYS

Applies to the deployment modes **in your own data centre** and **air-gapped**. When the assistant runs on your workstations, your documents stay where they are today; questions and answers, however, are produced on the computing power we provide — they are not stored there, see below. Anything processed beyond that is stated conclusively in the data processing agreement.

Your documents and records	The language model itself
Your staff's questions	The conversation histories
The generated answers	Every log
The search index	**Leaves your organisation: nothing**

AVAILABLE TODAY – FOR ROLLOUT AND REVIEW

MICROSOFT ENTRA · OKTA · KEYCLOAK

Enterprise sign-in

Staff use their organisation's existing sign-in. The confirmed identity is mapped to the right tenant.

USERS · GROUPS · DOCUMENTS

Permissions before search

Only released content is selected as a source for an answer. Content a user cannot access stays outside the search.

NINE SOURCE TYPES

Connect existing systems

Files and network drives, databases, IMAP, SharePoint, OneDrive, Outlook, Confluence, Jira and Zendesk.

REVIEWABLE BEFORE ROLLOUT

Evidence for answers and procurement

Answers show citations. Security questionnaire, deletion and contingency concept, operations manual, control mapping and software bill of materials are available on request.

THREE DEPLOYMENT MODES — YOU DECIDE WHERE THE COMPUTING HAPPENS

The same assistant, the same answers, the same functionality. The only difference is where the computing power sits. Private Deployment and Enterprise run on your infrastructure. Every deployment is quoted per project; hardware, integrations and service scope are agreed together before the quote.

Independent of the inference system: depending on the approved customer environment, Lokale AI can work with Ollama, llama.cpp, vLLM, NVIDIA NIM or another OpenAI-compatible server. Inference system, model and endpoints are verified per installation and set by the operations administrators.

In your data centre

Quoted per project

The installation sits in your network, under your access control. We hold only contract and billing data — no documents, no conversations, no logs. No document leaves your organisation. Enterprise adds extended integration and SLA scope.

- UP TO 4 WEEKS UNTIL GO-LIVE
- RUNS ENTIRELY ON YOUR INFRASTRUCTURE
- REQUIREMENTS AGREED WITH YOUR IT TEAM
- NAMED CONTACT PERSON

Air-gapped

Quoted per project

For environments that must never see the internet. Updates arrive on signed media, and the licence check works offline. Your audit team verifies for itself that the installation does not call out — before anything is installed.

- UP TO 8 WEEKS UNTIL GO-LIVE
- HARDWARE SIZED TO PROTECTION NEEDS AND DATA VOLUME
- ONBOARDING FOR YOUR OPERATIONS TEAM
- QUARTERLY VERIFICATION AUDIT

On your workstations

Quoted per project

The assistant is installed on your computers and reads your documents where they are today. We provide the computing power. Your question and the answer are neither stored nor logged there; the only thing counted is the number of requests, for billing, without content. Anything processed beyond that — in particular the documents themselves — is stated conclusively in the data processing agreement.

- IMMEDIATELY — INSTALLER FOR WINDOWS AND LINUX
- RUNS ON YOUR EXISTING COMPUTERS
- COMPUTING POWER PROVIDED BY US

AUDIT QUESTIONS — WHAT YOUR AUDITORS WILL ASK

1. Where is our data?

   That depends on the deployment mode.

   In your own data centre and in air-gapped operation, documents and processing stay in your environment. In the workstation model, the data path is set out in the contract.

2. Who has access?

   Your existing sign-in and your permissions remain authoritative.

   Microsoft Entra, Okta or Keycloak confirm the identity; user, group and document permissions limit the search.

3.	What happens when the contract ends?	**Deletion and return follow the deployment mode.** Where you operate the installation yourself, you delete the installation and the data in your environment; in the workstation model, the contract governs return and proof of deletion.
4.	Does the US CLOUD Act apply?	**It obliges US providers to hand over data they hold.** In your own data centre and in air-gapped operation, the data remains in your environment. In the workstation model, operator, storage location and legal basis are stated in the contract.
5.	Can we verify this before we install?	**Yes. Your own audit team proves it.** Verifiable, not promised — on the delivered software, before installation.
6.	What if the answer is wrong?	**You see which passages were used.** The answer shows the sources and passages it drew on. If there is not enough material, this is flagged visibly.

WHO IT IS FOR

Data-sensitive companies — Legal, compliance, finance and R&D work with knowledge that has to stay under your control. The assistant runs in your data centre, fully isolated or on your workstations.

Public organisations — Authorities, public enterprises and institutions keep case files, citizen enquiries and internal regulations in their own environment — with no network connection at all, if required.

Regulated sectors — Healthcare, critical infrastructure and other regulated sectors can use their internal knowledge without handing it to an external AI service.

Next step: 45 minutes, no obligation. We go through your use case and the requirements of your review — and tell you honestly if we are not the right fit.

Get in touch via lokaleai.de or kontakt@lokaleai.de.

As of September 2026 · Company based in the EU