

Offline Vision Analyzer

100% local • zero internet required after setup • private by design

A desktop multimodal RAG system that lets you upload scans, charts, diagrams, and PDFs, then ask natural-language questions about them — all powered by local models.

Github link: <https://github.com/santhimaddipudi/OfflineVisionAnalyzer>

Upload file → EasyOCR (text) + CLIP (image embedding) → ChromaDB



User question → hybrid retrieval → LLaVA (Ollama) → answer + citations

Flow:

Upload scan/chart/diagram



Preprocess image/PDF



Easy OCR extracts OCR



LLaVA answers visual questions



Embed text + image summaries



Store in ChromaDB



User asks question



Retrieve relevant text/images

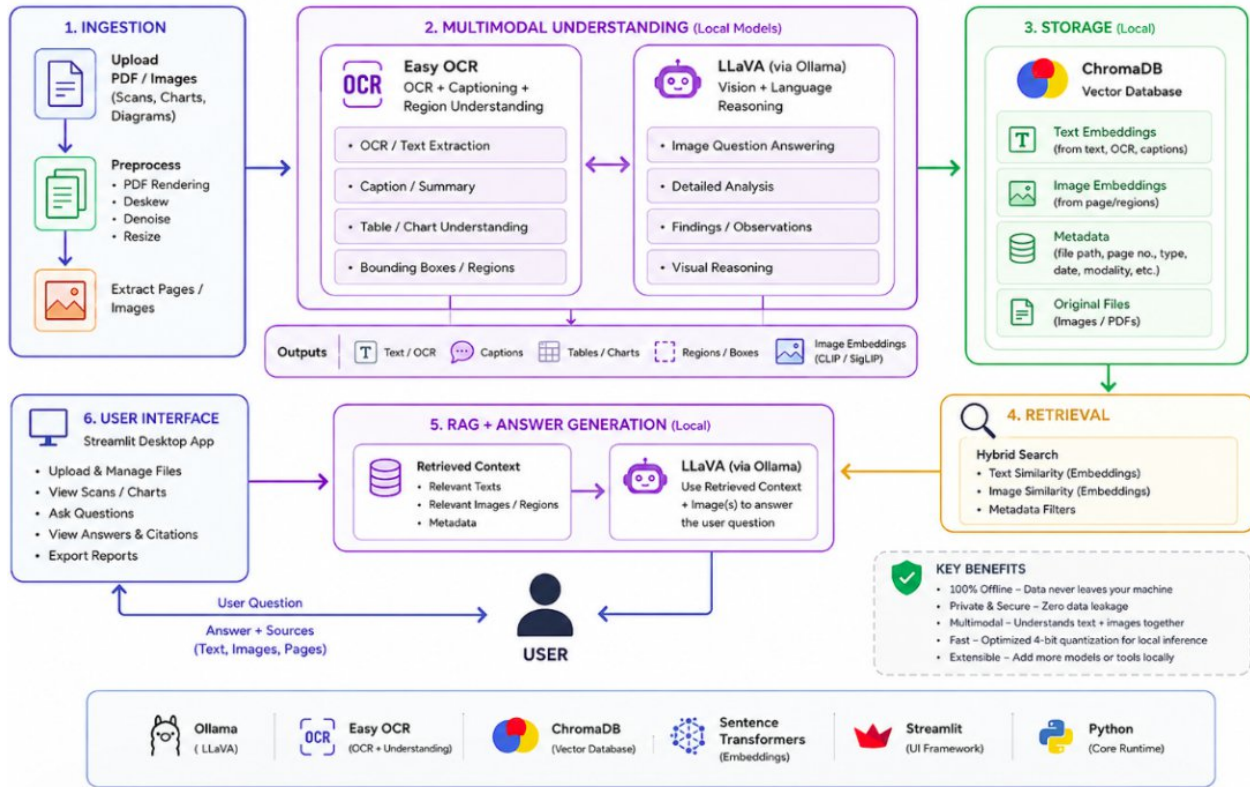


LLaVA + retrieved context answers locally

Architecture:

OFFLINE VISION ANALYZER – ARCHITECTURE

Local · Private · No Internet Required



Minimal project structure:

```
OfflineVisionAnalyzer/
├── app.py                # Streamlit UI
├── pipeline.py           # End-to-end ingestion pipeline
├── config.py             # Paths and model settings
├── requirements.txt
├── ingestion/
│   ├── pdf_processor.py  # PDF → page images (PyMuPDF)
│   └── image_processor.py # Deskew, resize
├── models/
│   ├── ocr.py            # OCR via EasyOCR
│   ├── llava.py          # LLaVA via Ollama
│   └── embeddings.py     # MiniLM text + CLIP image embeddings
├── storage/
│   └── chromadb_client.py # ChromaDB collections
├── retrieval/
│   └── hybrid_search.py  # Text + image similarity search
├── rag/
│   └── answer_generator.py # Retrieval → LLaVA → answer
└── data/                # Created automatically at runtime
    ├── originals/       # Uploaded source files
    ├── images/          # Rendered page images
    └── chromadb/         # Vector database
```