

 Title: Knowledge Assistant – Build a Local Retrieval AI Agent

Assignment Overview:

Your onboarding task is to build a FOXO Knowledge Assistant — an AI Agent that can answer questions using internal company documents (provided as local .pdf, .txt, or .md files).

The agent should be powered by LLMs and designed using frameworks like LangChain, LlamaIndex, or similar. It should be capable of retrieval-augmented generation (RAG) and should support multi-step reasoning, if needed, with a CLI or minimal web interface for interaction.

This challenge is designed to test how well you can:

- Build independently with LLM tools
- Apply vector search & RAG concepts
- Write modular, readable Python code
- Handle prompt design, tool integration, and agent planning
- Create something testable, usable, and explainable

Your Task

Build an AI Agent that can:

1. Ingest local documents (at least 3 sample .pdf or .txt files — you choose the domain/topic)
2. Create vector embeddings and store them in a vector DB (FAISS or ChromaDB)
3. Accept user questions via a CLI or basic UI
4. Retrieve and synthesize answers from the ingested content
5. Return a response with source attribution (filename/snippet if possible)

Challenge Layer:

Add a "Multi-Tool Mode" — where the agent can also:

1. Call an external API (e.g., weather, search, calculator)
2. Route tasks using reasoning (e.g., "summarize PDF" vs "get live data")

Tech Guidelines

Use:

- Python (3.8+)
- LangChain / LlamaIndex / AutoGen
- OpenAI / Claude / Ollama (your choice)
- FAISS or ChromaDB
- Streamlit or CLI (text-only interface is acceptable)

- PyMuPDF, pdfplumber, or similar to parse PDFs

Do not use:

- AI-generated code
- Prebuilt agents from GitHub or templates
- Vector databases with paid plans
- Frontend libraries or any UI/UX framework

What to Submit

1. A **public GitHub repository** with your complete code.
2. A **README.md** that includes:
 - Setup instructions (`pip install -r requirements.txt`, how to run)
 - Tools used + rationale
 - 2–3 sample questions and responses from your agent
 - Any limitations or next steps if you had more time

Repo must be public and accessible.