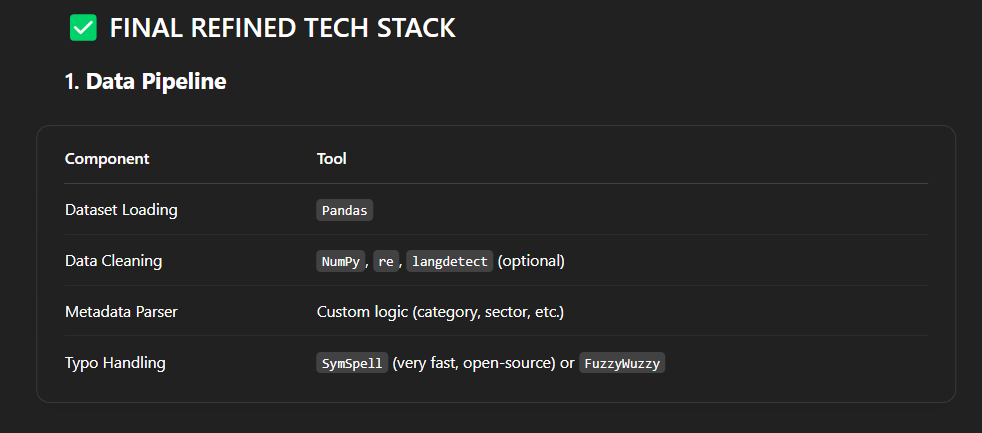
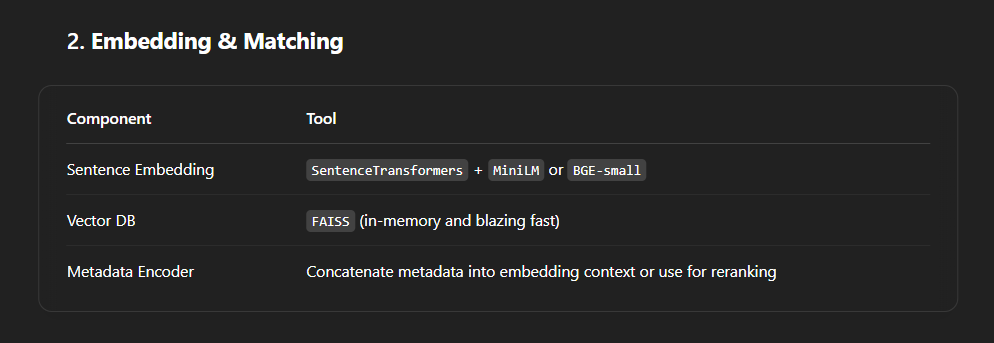
a **production-ready, real working project**, not just a demo. Since you're serious about **end-to-end functionality**, we’ll refine the tech stack and architecture to be:

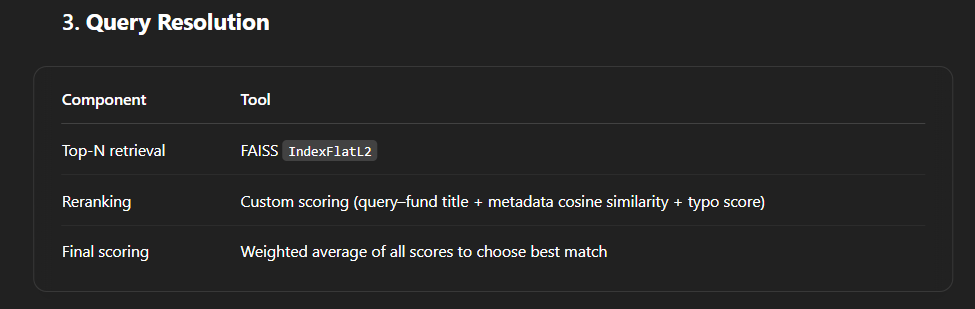
* ✅ Scalable
* ✅ Efficient on student-level resources (no billing!)
* ✅ Accurate and robust
* ✅ Modular (clean project layout)
* ✅ CLI/UI optional, but **backed by real logic**

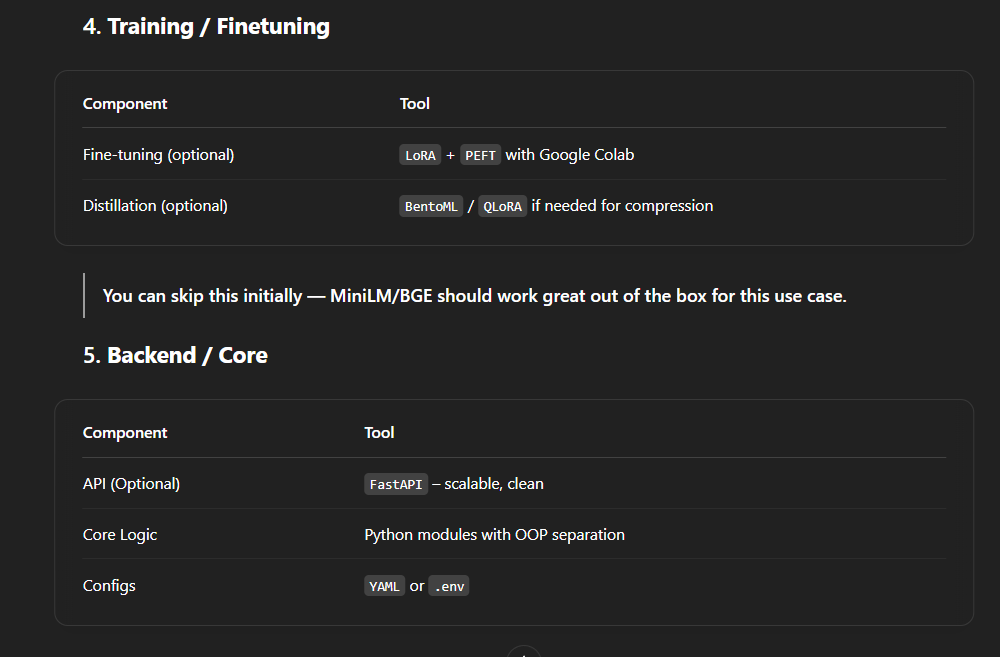
**🧠 Project Goal Recap**

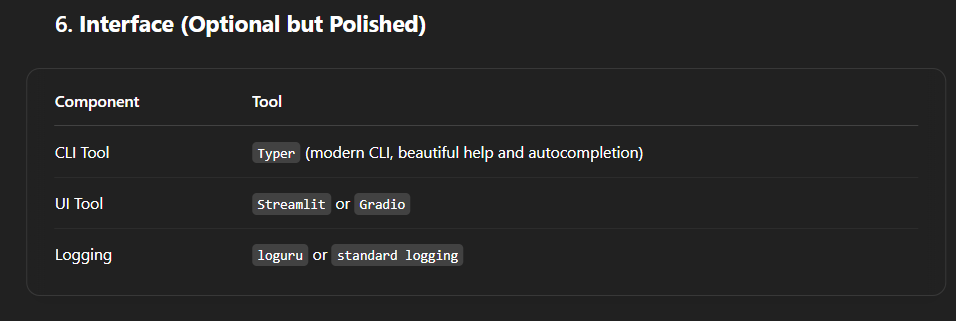
Build a **real-time, accurate, and fuzzy search system** for Indian mutual funds using a **small, deployable language model**, vector similarity, and metadata-enhanced matching.











**🚀 Real Working Architecture Flow**

flowchart TD

A[Query Input: "SBI infra"] --> B[Preprocessing: Lowercase, Remove Stopwords, Typo Correction]

B --> C[Sentence Encoding using MiniLM]

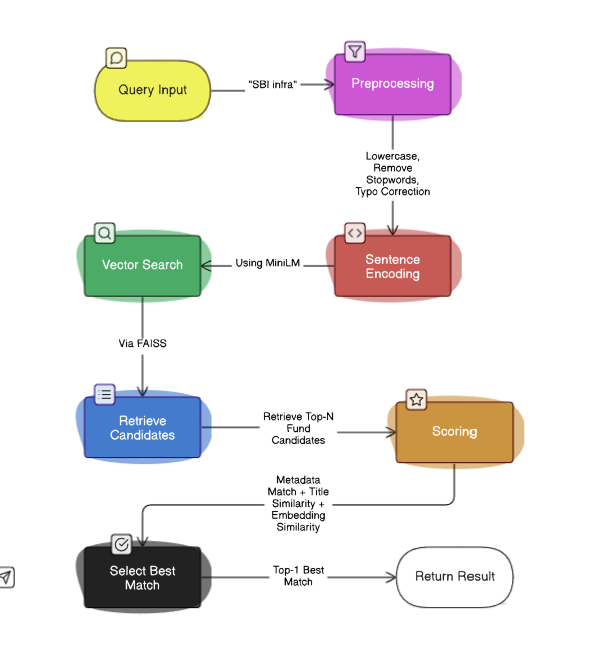
C --> D[Vector Search via FAISS]

D --> E[Retrieve Top-N Fund Candidates]

E --> F[Score using Metadata Match + Title Similarity + Embedding Similarity]

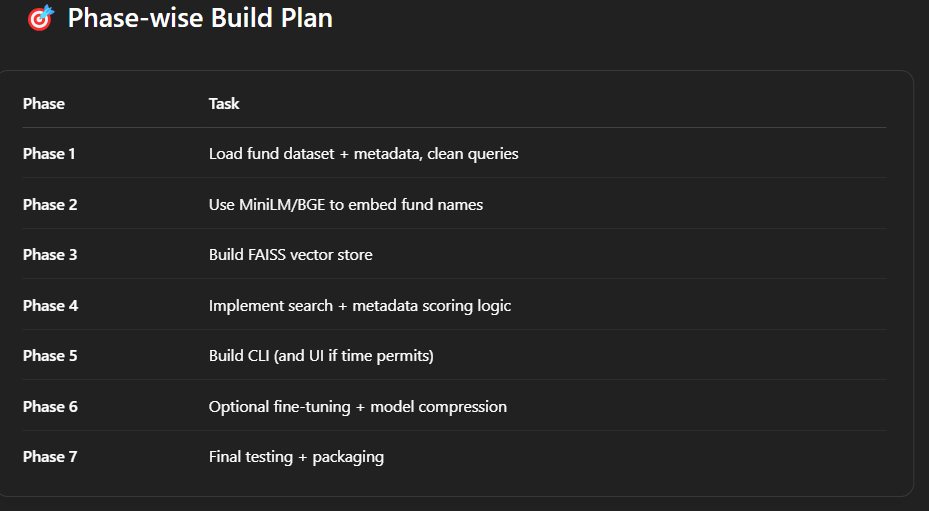
F --> G[Select Top-1 Best Match]

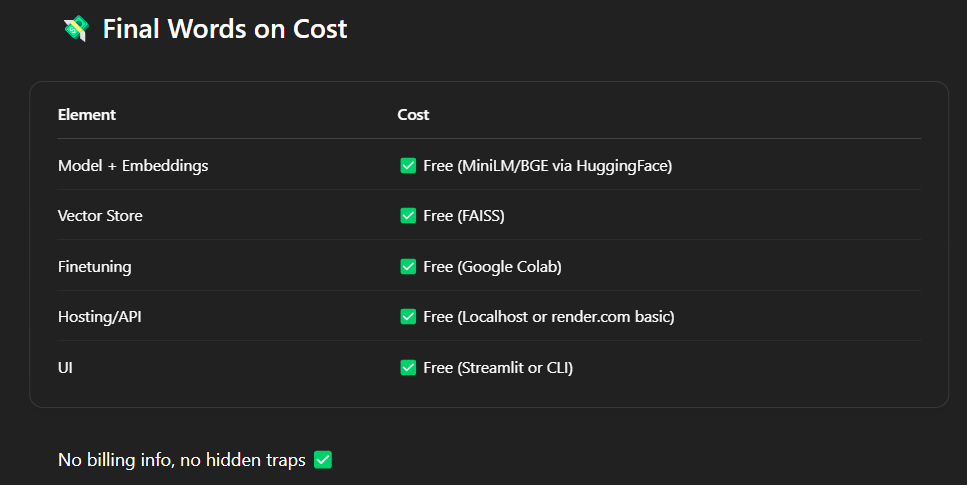
G --> H[Return Fund Name, Match Score, Explanation (Optional)]



***DIRECTORY STRUCTURE { REFERENCE }***





****

**REQUIREMENTS**

* pandas
* numpy
* sentence-transformers
* faiss-cpu
* typer
* streamlit
* fastapi
* uvicorn
* loguru
* python-dotenv
* fuzzywuzzy
* python-Levenshtein