**5. How You’re Doing It Better (in one line):**

**“We’re building an explainable LLM-powered fund search assistant that understands investor intent — even from fuzzy, ambiguous queries — and maps them to the right mutual funds using rich metadata.**

**1. What Are We Solving?**

**Problem:**

Most FinTech apps (Groww, Zerodha, PayTM Money) offer basic or fuzzy search for mutual funds and stocks. They:

* Fail on **typos**, **partial names**, **abbreviations**
* Can’t handle **natural language** like:
  + “Show me tax-saving funds with high return”
  + “Funds from the tech sector with low volatility”
* Don’t **explain why** a result was shown

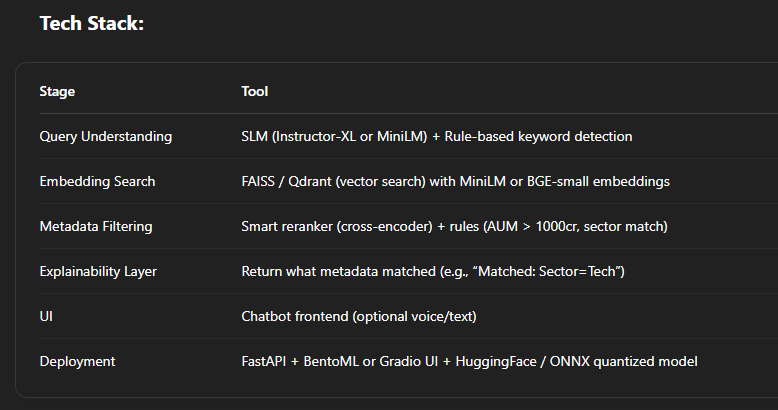
**Our Mission:**

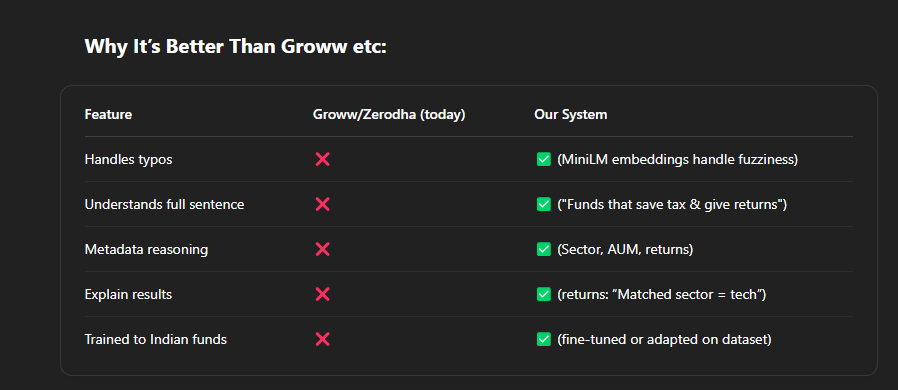
**Build a smart search assistant** (compact LLM-powered system) that:

* Understands user queries *like a financial advisor*
* Links them to relevant funds using **metadata**, not just keywords
* Works in **real-time**
* Explains why the fund was selected
* Supports **advanced use cases** like portfolio suggestions, loan analysis (MyCash)

**2. Our Approach (And Why It’s Better)**

**Overall Flow:**

1. **Understand the Query** (Intent + Entities)
2. **Use Embeddings to Fetch Candidates**
3. **Rerank with Metadata**
4. **Explain why the result was chosen**



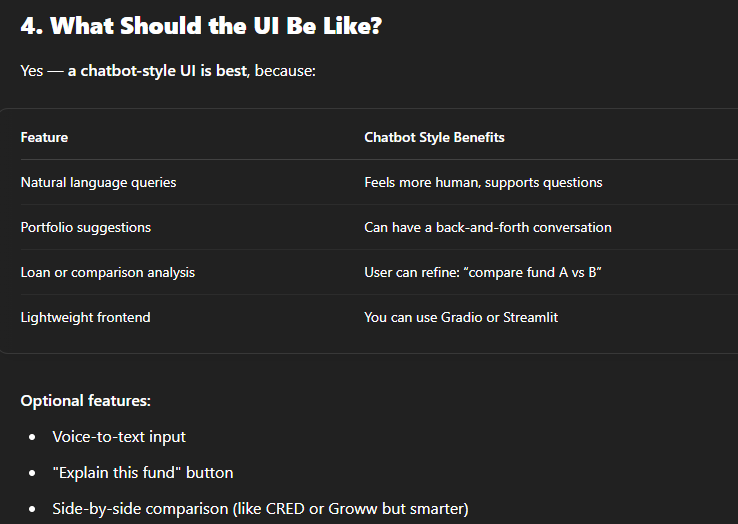
**3. Which LLM / API to Use?**

**Free Models You Can Use:**

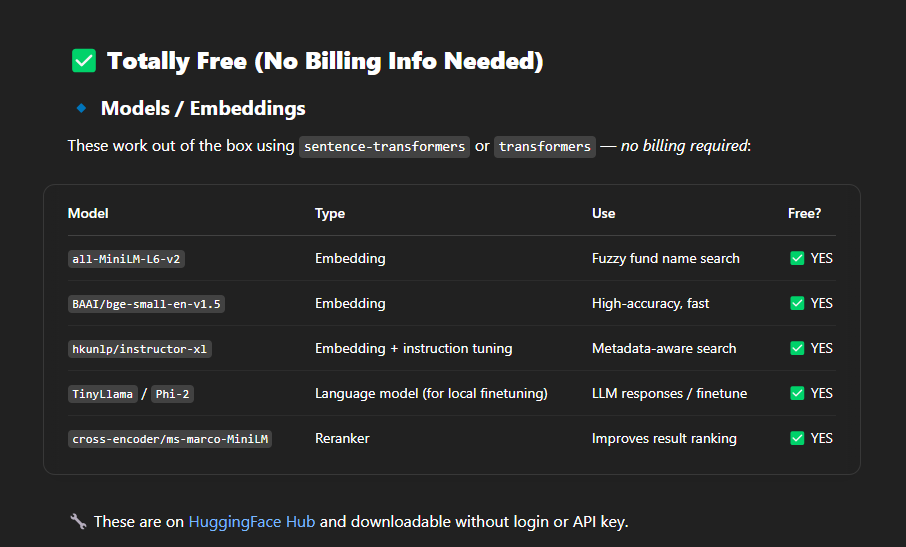
1. **Instructor-XL (best for zero-shot instructions + embeddings)**
   * Free on HuggingFace: hkunlp/instructor-xl
   * Accepts queries like "Retrieve funds for: Tax Saving + High Return"
2. **MiniLM (lightweight, fast)**
   * sentence-transformers/all-MiniLM-L6-v2
3. **TinyLlama / Phi-2** (if you want to fine-tune a SLM)
   * Run on **Google Colab / Kaggle Notebooks**
   * Use **QLoRA** or **LoRA** to train efficiently

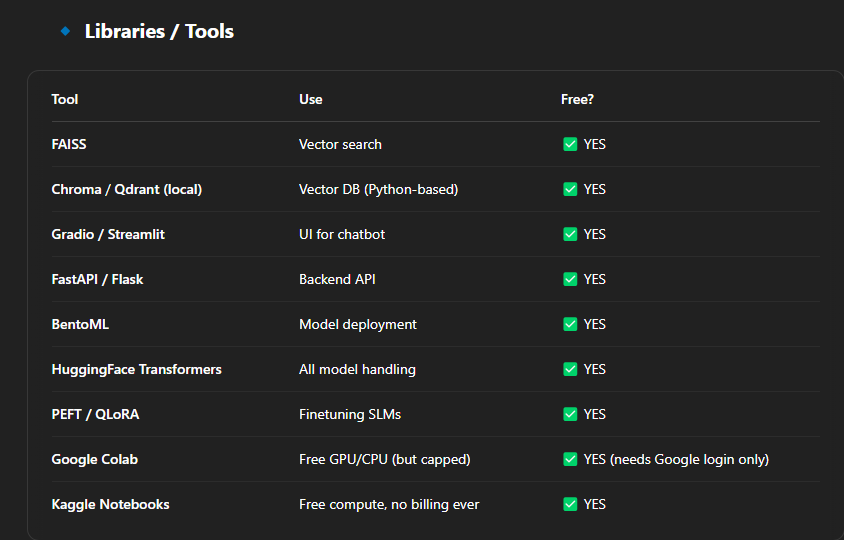
**Are There Free APIs for Funds/Stocks?**

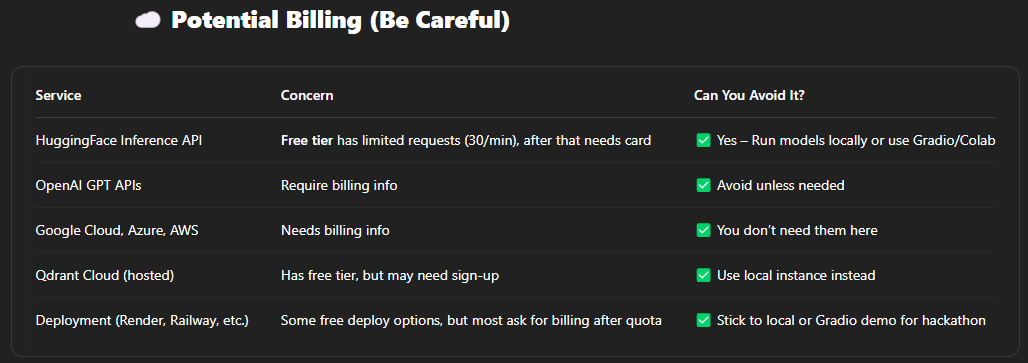
* **Pre-trained models:** No one has an API trained specifically for Indian mutual funds *with semantic understanding*. That's your edge.
* **Financial Data APIs (optional for real-time data):**
  + Groww API *(unofficial)*
  + [Morningstar India](https://www.morningstar.in/) *(scraping possible)*
  + [NSE India API](https://www.nseindia.com/) *(can be scraped)*
  + [AMFI India](https://www.amfiindia.com/) *(mutual fund data dumps)*



**exactly what's free** and **what may need billing**.







**🔥 Best Strategy Without Billing**

Use:

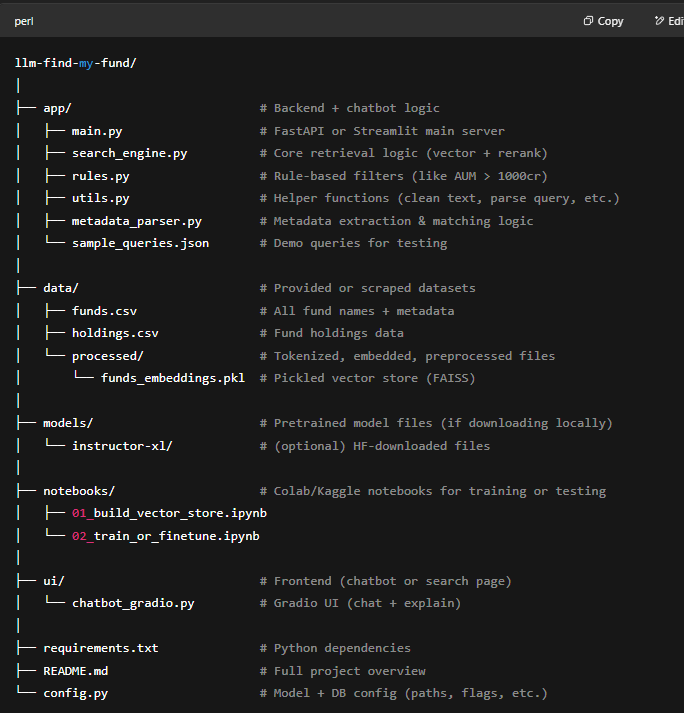
* **MiniLM/BGE/Instructors** with FAISS locally
* **Gradio** as your UI layer
* Run everything on **Kaggle** or **Google Colab**
* Optional: deploy locally via FastAPI or share demo via Gradio link

You’ll be able to:

* Handle embeddings
* Perform retrieval + reranking
* Show a chatbot UI
* Train or fine-tune if needed
* All for **₹0** and **no card added anywhere**

**full project plan**, including:

✅ Directory structure  
✅ Pre-requisites  
✅ File-by-file purpose  
✅ How to run — *fully local & billing-free*



**Pre-requisites**

**🔧 Environment**

* Python 3.10+
* Pip / Conda
* Git

**Python Dependencies (requirements.txt)**

faiss-cpu

sentence-transformers

transformers

torch

scikit-learn

pandas

numpy

gradio

tqdm

fastapi

uvicorn

peft

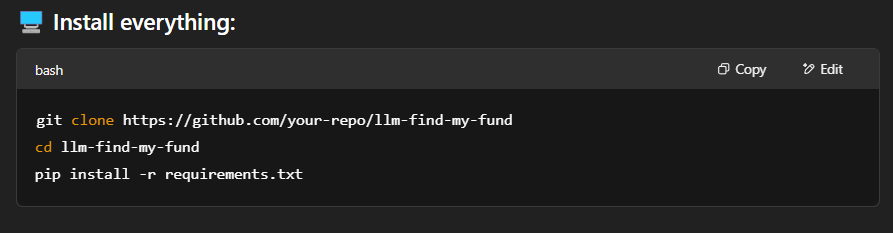
accelerate

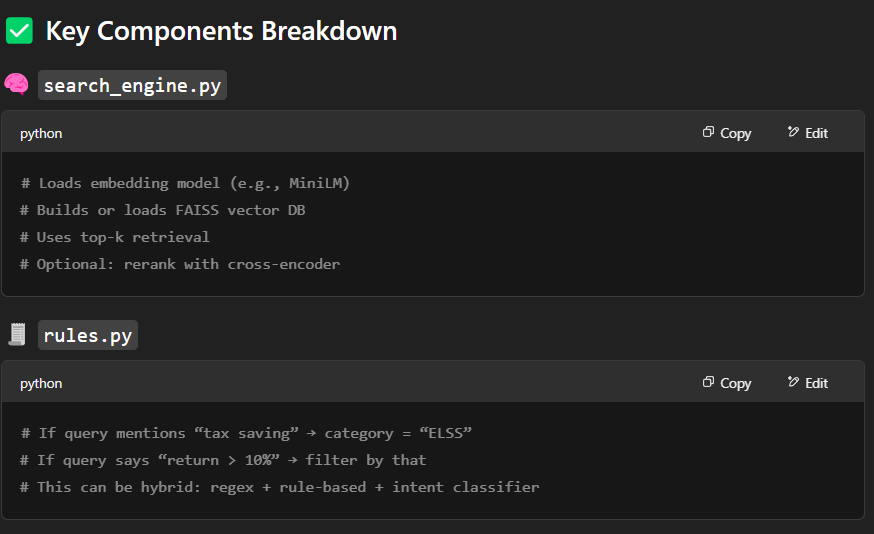
Optional if using reranking:

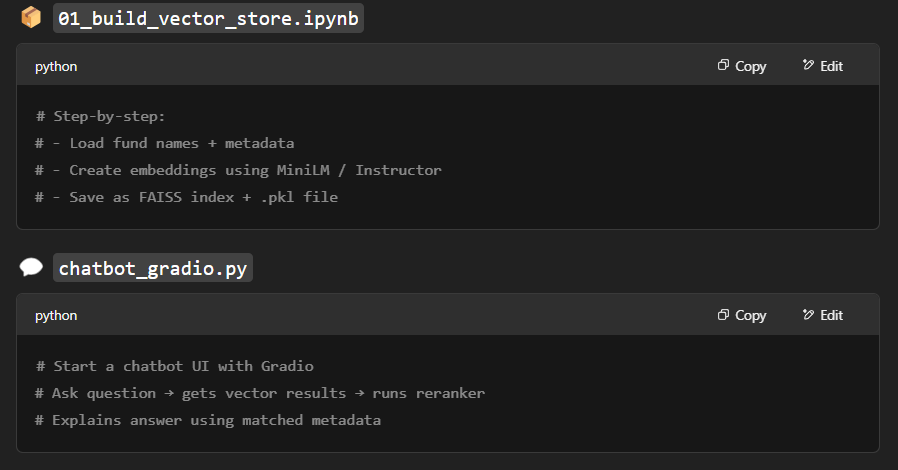
txt

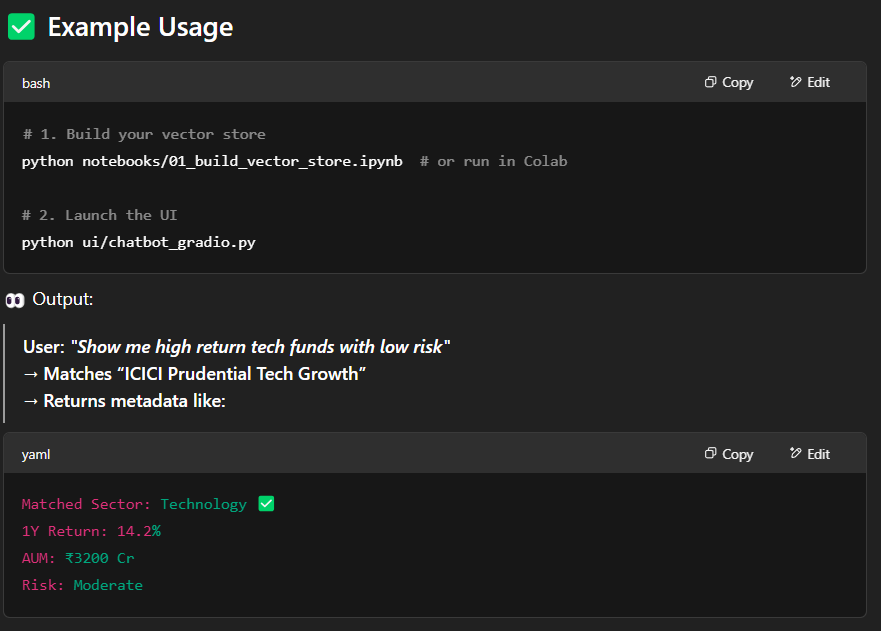
CopyEdit

cross-encoder









**🧪 Optional Bonus**

Want to train your own model?

* Use 02\_train\_or\_finetune.ipynb with LoRA/QLoRA on TinyLlama or Phi-2
* Host on HuggingFace Spaces for free demo (no billing)

