

RAG + LLM Hands-On Assignment

Objective

Build a **Retrieval-Augmented Generation (RAG)** system that answers complex financial and legal questions by retrieving information from **Apple's 2024 10-K** and **Tesla's 2023 10-K**, and generating accurate, well-sourced responses using a **local or open-access LLM**.

Provided Materials

You are given two documents:

- **10-Q4-2024-As-Filed.pdf** – Apple Inc.'s Annual Report on Form 10-K for the fiscal year ended September 28, 2024
- **tsla-20231231-gen.pdf** – Tesla, Inc.'s Annual Report on Form 10-K for the year ended December 31, 2023

These are real SEC filings containing financial statements, risk factors, legal proceedings, and executive compensation details.

Your Task

Step 1: Document Ingestion & Indexing

- Parse both PDFs into clean text.
- Generate embeddings using an **open-source model**.
- Store in a vector database (e.g., FAISS, Chroma).
- Preserve metadata: **document, section, page number**.

Step 2: Build a Retrieval Pipeline

- Implement vector similarity search to retrieve top-5 relevant chunks.
- Add a re-ranker.
- Justify your choice in the design report.

Step 3: LLM Integration

- Use a locally hosted or open-access LLM (e.g., Llama 3, Mistral, Phi-3).
- **Do not use GPT-4, Claude, or any closed API.**
- Design a custom prompt that:

- Uses only retrieved context.
- Cites sources as ["Apple 10-K", "Item 8", "p. 28"].
- Responds with "Not specified in the document." if the answer is not in the provided files.
- Refuses out-of-scope questions with:
"This question cannot be answered based on the provided documents."

Step 4: Answer the Test Questions

Run your system on the 13 questions below and output answers in the specified JSON format.

Evaluation Questions

```
[{"question_id": 1, "question": "What was Apples total revenue for the fiscal year ended September 28, 2024?"}, {"question_id": 2, "question": "How many shares of common stock were issued and outstanding as of October 18, 2024?"}, {"question_id": 3, "question": "What is the total amount of term debt (current + non-current) reported by Apple as of September 28, 2024?"}, {"question_id": 4, "question": "On what date was Apples 10-K report for 2024 signed and filed with the SEC?"}, {"question_id": 5, "question": "Does Apple have any unresolved staff comments from the SEC as of this filing? How do you know?"}, {"question_id": 6, "question": "What was Teslas total revenue for the year ended December 31, 2023?"}, {"question_id": 7, "question": "What percentage of Teslas total revenue in 2023 came from Automotive Sales (excluding Leasing)?"}, {"question_id": 8, "question": "What is the primary reason Tesla states for being highly dependent on Elon Musk?"}, {"question_id": 9, "question": "What types of vehicles does Tesla currently produce and deliver?"}, {"question_id": 10, "question": "What is the purpose of Teslas 'lease pass-through fund arrangements'?"}, {"question_id": 11, "question": "What is Teslas stock price forecast for 2025?"}, {"question_id": 12, "question": "Who is the CFO of Apple as of 2025?"}, {"question_id": 13, "question": "What color is Teslas headquarters painted?"}]
```

Output format for each question

```
[{"question_id": 11, "answer": "This question cannot be answered based on the provided documents.", "sources": []}]
```

Ground-Truth Answers (for evaluation)

Q#	Question	Answer	Source
----	----------	--------	--------

1	Apple's total revenue FY 2024	\$391,036 million	Apple 10-K, Item 8, p. 282
2	Apple's common stock outstanding (Oct 18, 2024)	15,115,823,000 shares	Apple 10-K, first paragraph
3	Apple's total term debt (Sept 28, 2024)	\$96,662 million	Apple 10-K, Item 8, Note 9, p. 394
4	Filing date of Apple's 2024 10-K	November 1, 2024	Apple 10-K, Signature page
5	Unresolved SEC staff comments?	No. Checkmark indicates "No" under Item 1B	Apple 10-K, Item 1B, p. 176
6	Tesla's total revenue FY 2023	\$96,773 million	Tesla 10-K, Item 7
7	Tesla revenue from Automotive Sales	~84% (\$81,924M / \$96,773M)	Tesla 10-K, Item 7
8	Tesla's dependency on Elon Musk	Central to strategy, innovation, leadership; loss could disrupt	Tesla 10-K, Item 1A
9	Tesla's current vehicles	Model S, Model 3, Model X, Model Y, Cybertruck	Tesla 10-K, Item 1
10	Tesla's lease pass-through arrangements	Finance solar systems with investors; customers sign PPAs	Tesla 10-K, Item 7
11	Tesla's stock price forecast 2025	Not answerable	N/A
12	Apple CFO (2025)	Not answerable	N/A
13	Tesla HQ color	Not answerable	N/A

Hosting & Evaluation: Run Your Solution in the Cloud

You must host your complete RAG system in a public Colab or Kaggle notebook, fully runnable end-to-end.

Submission Requirements

Item	Description
GitHub Repository	Public repo with code, requirements, and README.
Concept Note	Short <code>design_report.md</code> (1 page) explaining chunking, LLM choice, out-of-scope handling.
Runnable Notebook	Colab/Kaggle notebook cloning repo, indexing PDFs, and running inference.
Direct Link	Live notebook link in README.

Required Interface: One Function to Rule Them All

```
def answer_question(query: str) -> dict:
    """
    Answers a question using the RAG pipeline.

    Args:
        query (str): The user question about Apple or Tesla 10-K filings.
    """

```

```
Returns:
dict: {
    "answer": "Answer text or 'This question cannot be answered based on the provided
documents.'",
    "sources": ["Apple 10-K", "Item 8", "p. 28"] # Empty list if refused
}
"""
# Your RAG logic here
return {"answer": "...", "sources": [...]}
```