

RAG for Financial Document Summarization

By: Harshit Kumar, Sarthak Khandelwal, Alexander Leon



High-Level System Description

Objective: To create an AI-powered Chatbot that summarizes financial documents for investor insights using Retrieval Augmented Generation (RAG).

Input: Financial reports such as 10-K and 10-Q.

Output: Chatbot interactions that provide summarized financial insights.

TESLA, INC.
FORM 10-Q FOR THE QUARTER ENDED SEPTEMBER 30, 2023
INDEX

	Page
PART I. FINANCIAL INFORMATION	
Item 1.	4
Financial Statements	4
Consolidated Balance Sheets	5
Consolidated Statements of Operations	6
Consolidated Statements of Comprehensive Income	7
Consolidated Statements of Redeemable Noncontrolling Interests and Equity	9
Consolidated Statements of Cash Flows	10
Notes to Consolidated Financial Statements	27
Item 2.	36
Management's Discussion and Analysis of Financial Condition and Results of Operations	36
Item 3.	36
Quantitative and Qualitative Disclosures about Market Risk	36
Item 4.	36
Controls and Procedures	36
PART II. OTHER INFORMATION	
Item 1.	36
Legal Proceedings	37
Item 1A.	37
Risk Factors	37
Item 2.	37
Unregistered Sales of Equity Securities and Use of Proceeds	37
Item 3.	37
Defaults Upon Senior Securities	37
Item 4.	37
Mine Safety Disclosures	37
Item 5.	37
Other Information	38
Item 6.	38
Exhibits	38
Signatures	
	39

CS

You

what's the total assets for Tesla as of September 30, 2023?



ChatGPT

As of September 30, 2023, Tesla's total assets were approximately \$106.62 billion .

CS

You

what's the total assets for Tesla as of September 30, 2023?



ChatGPT

As of September 30, 2023, Tesla's total assets were valued at approximately \$93.94 billion (YCharts).



Technical Problem Formulation

Problem statement: Given a PDF document and a query, retrieve the relevant details and information from the document as per the query, and synthesize this information to generate accurate answers.

Data Ingestion and Processing: Reading PDFs of financial reports and split the documents for efficient text chunking of long documents.

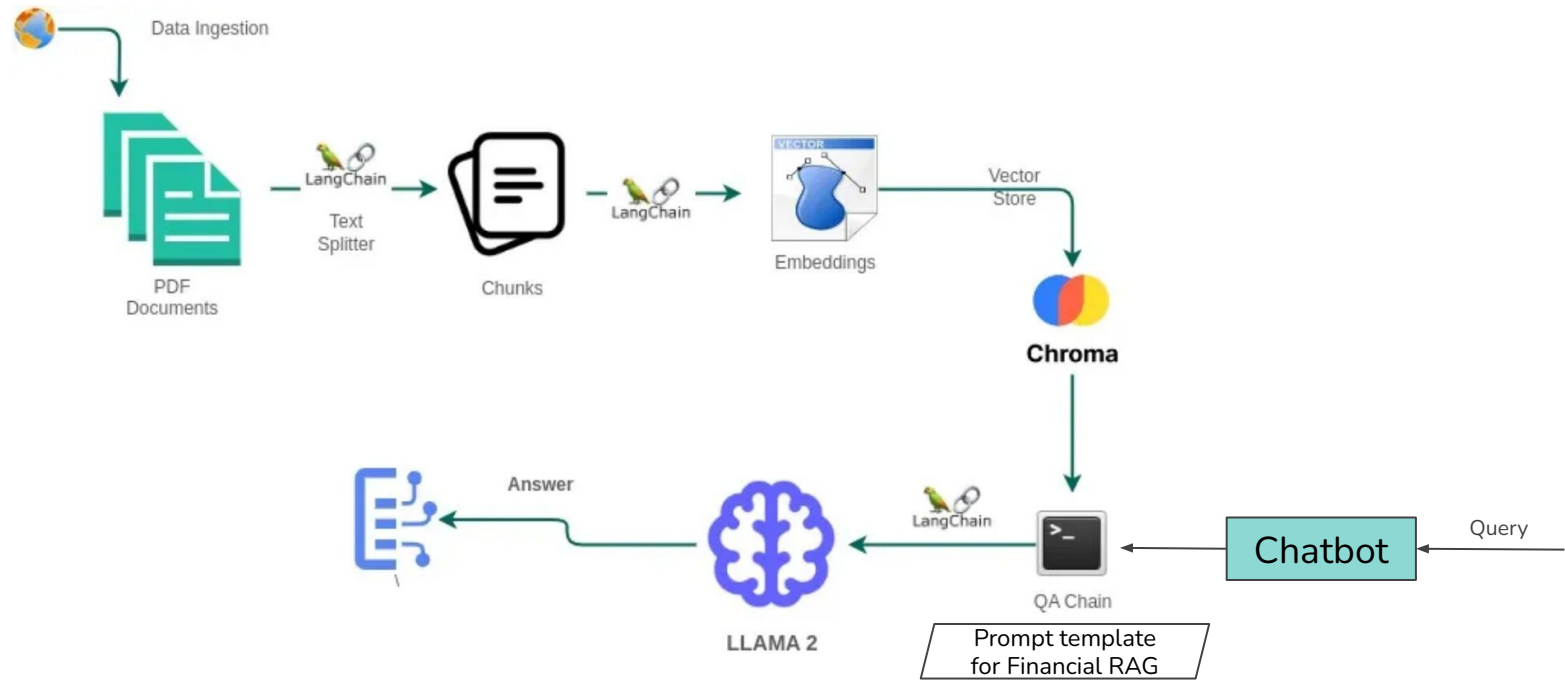
Retrieval-Augmented Generation (RAG): Combination of document retrieval with the generative capabilities of the chosen language models.

Large Language Models: Evaluation of various models, including GPT-3.5-turbo, LLama 2, Gemma 1.1, etc.

Conversation Chain and Prompt Design: Crafting of a prompt template designed for concise two-sentence financial summaries.

User interface: Designing Chatbot like user interface.

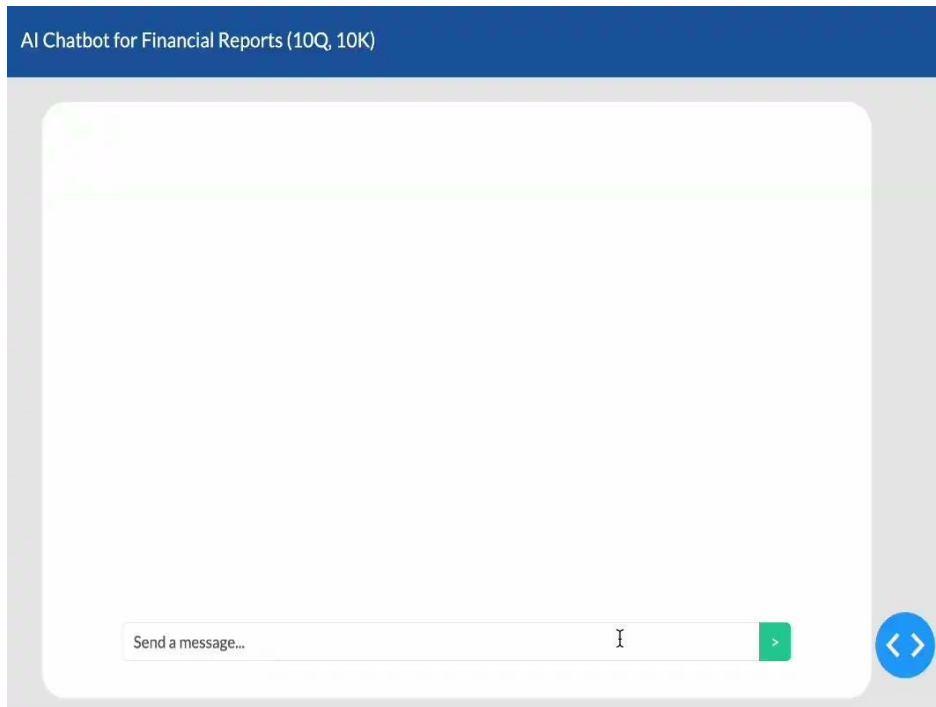
System architecture





Demonstration

Demo for RAG for Tesla 10Q



https://github.com/sarthakforwet/Financial_Document_Summarization_through_RAG

Experiments



Questions	Answers			
	LLama 2 (7b params)	GPT-3.5	Gemma 1.1 (7b params)	Flan-T5
What's the total assets?	The total assets of the company is \$93,941.	The total assets are \$93,941.	The total assets as of December 31, 2022, was \$45,026 million.	\$ 93,941 \$ 82,338
How much cash was provided by or used in operating activities during the quarter?	The amount of cash provided by or used in operating activities during the quarter was \$8.89 billion.	The net cash provided by operating activities during the quarter was \$8,886 [millions].	The provided text does not contain any information regarding the amount of cash provided or used in operating activities during the quarter, so I am unable to answer this question from the given context.	11,446 [millions]
What are the biggest risks for Tesla as a business?	The biggest risks for Tesla as a business are its ability to continue as a going concern and its inability to raise additional capital to fund its operations and growth.	One of the biggest risks for Tesla as a business is legal proceedings, risks, and claims that arise from normal business activities. Additionally, data breaches and leaks of non-public Tesla business and personal information pose significant risks.	Tesla faces significant risks related to manufacturing and supply chain disruptions, as well as competition from established automakers.	legal proceedings, risks and claims that arise from the normal course of business activities



RAG Evaluation

On an evaluation set of 11 questions, our RAG system is able to answer a majority of questions correctly (using simple user queries).

ragas score

generation

faithfulness

how factually accurate is
the generated answer

answer relevancy

how relevant is the generated
answer to the question

retrieval

context precision

the signal to noise ratio of retrieved
context

context recall

can it retrieve all the relevant information
required to answer the question



RAG Evaluation

	question	answer	contexts	ground_truths	ground_truth	context_precision	context_recall	faithfulness	answer_relevancy
0	What's the total assets?	The total assets as of September 30, 2023, wer...	[September 30,\n2023December 31,\n2022\nAssets...	[The total assets of the company is \$93,941.]	The total assets of the company is \$93,941.	1.0	0.0	0.0	0.887874
1	How much cash was provided by or used in opera...	Net cash provided by operating activities duri...	[Net cash provided by operating activities 8,8...	[The amount of cash provided by or used in ope...	The amount of cash provided by or used in oper...	1.0	1.0	1.0	0.960075
2	What are the biggest risks for Tesla as a busi...	The biggest risks for Tesla as a business incl...	[position or brand.\nWe are also subject to va...	[The biggest risks for Tesla as a business are...	The biggest risks for Tesla as a business are ...	0.0	0.0	1.0	0.994410

GPT-3.5 answers evaluation by ragas library



Challenges and Future Work

- **Challenges:** There are many similar terms that occur in the same financial document (cash flow from *operating* activities, cash flow from *investing* activities, cash flow from *financing* activities).
- **Future Work:**
 - Scaling to multiple companies (and cross-company comparisons).
 - Dashboard: Inclusion of a KPI dashboard that generates contextually relevant information specific to the KPIs .