



DS 552: Generative AI - Project Report

Project Title : FinStatAI - A Generative AI-Powered Financial Statement Analysis Tool

Sreejith Madhavankutty

Email: smadhavankutty@wpi.edu;

Date: 4/13/2025

Introduction and Objective

Financial statement analysis is a critical process for investors, analysts, and businesses to evaluate the financial health and performance of companies. Traditional methods of financial analysis often involve manual data extraction, interpretation, and reporting, which can be time-consuming and prone to human error. With the advent of Large Language Models (LLMs), there is an opportunity to automate and enhance the accuracy of financial statement analysis.

Project proposal is to develop a web application - **FinStatAI** that leverages an open source pretrained LLM to automate and enhance financial statement analysis by providing interactive, intelligent, and real-time insights.

The primary objective of **FinStatAI** is to address the following :

- Processes financial statements (balance sheets, income statements, cash flow statements).
- Answers queries in natural language (e.g., "What is the company's profit margin compared to previous year?").
- Generates summaries, trend analyses, and key financial ratios.
- Helps users make data-driven decisions faster.

Reasons for selecting Generative AI to build **FinStatAI** are as follows:

- Natural Language Understanding (NLU) capabilities of LLMs. They can interpret complex financial jargon.
- Retrieval-Augmented Generation(RAG) ensures responses are factually accurate by referencing financial docs.
- Provides opportunities for automation there by reducing manual effort in financial statement analysis.

Selection of Generative AI Model

In order to build **FinStatAI** , project evaluated Retrieval-Augmented Generation (RAG) with Llama 2 , GPT-4 and Claude, all modern LLMs known for its efficiency and performance in natural language understanding tasks. All of them are equally suitable for this project due to its ability to handle complex financial terminology and its relatively low computational resource requirements. Evaluation results for each model configuration is published as part of the report (Appendix).

Features considered and justification why it's suitable are illustrated as follows:-

Features	Why its suitable
Accuracy	RAG retrieves relevant financial data before generating responses, reducing hallucinations
Scalability	Works efficiently with structured (tables) and unstructured (MD&A sections) financial data
Customizability	Can be fine-tuned on financial datasets (e.g., SEC filings, earnings reports)
Cost-Effectiveness	Open-weight models (Llama 2) reduce API costs compared to OpenAI GPT-4

Various model configurations provide capability to combine retrieval mechanisms with generative capabilities, enabling it to provide accurate and contextually relevant answers based on the information in financial statements. The RAG system's relevance to the task, coupled with its feasibility for implementation within the Azure Databricks environment (platform used for the build process), makes it an ideal choice for building and deploying **FinStatAI** application.

At a high-level, [basic RAG](#) has following components:

- ✓ Ingest
- ✓ Retrieval
- ✓ Synthesis

Project Definition and Use Case

Application Concept

FinStatAI web application will function as a financial statement analysis tool that allows users to upload financial documents (e.g., balance sheets, income statements, cash flow statements) and receive a detailed analysis of key financial metrics such as profitability, liquidity, and solvency ratios. The application will also provide a summary of the financial health of the company and highlight potential areas of concern. It can also provide a basic question-answering interface for financial analysts.

In summary, **FinStatAI** handles following:-

AI-Powered Financial Q&A System: Users upload financial statements (PDF/Excel) and ask questions.

Automated Summarization: Extracts key metrics (revenue growth, debt ratios).

Trend Analysis: Compares financial performance across quarters/years.

Use of LLM to build FinStatAI

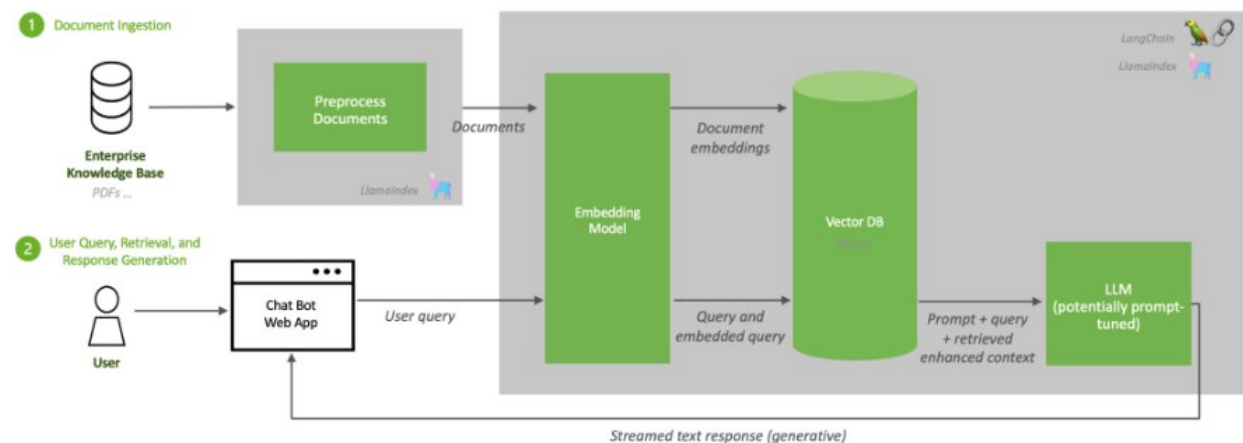
The LLM will be used to:

- ✓ Extract key financial data from uploaded documents.
- ✓ Summarize the financial performance of the company.
- ✓ Generate insights based on the extracted data, such as identifying trends, anomalies, or potential risks.

FinStatAI Integration Plan

- ✓ For efficient financial statement document processing, **FinStatAI** integrates with Azure Document Intelligence which will assist with extraction of text/tables.
- ✓ **FinStatAI** uses Vector Database (Azure AI Search) to Store embeddings.
- ✓ RAG Pipeline will be used to retrieve relevant data from vector store before generating answers.
- ✓ Streamlit UI provides interactive dashboard for queries and visualizations.

Overview of RAG pipeline components: ingest and query flows



Implementation Plan

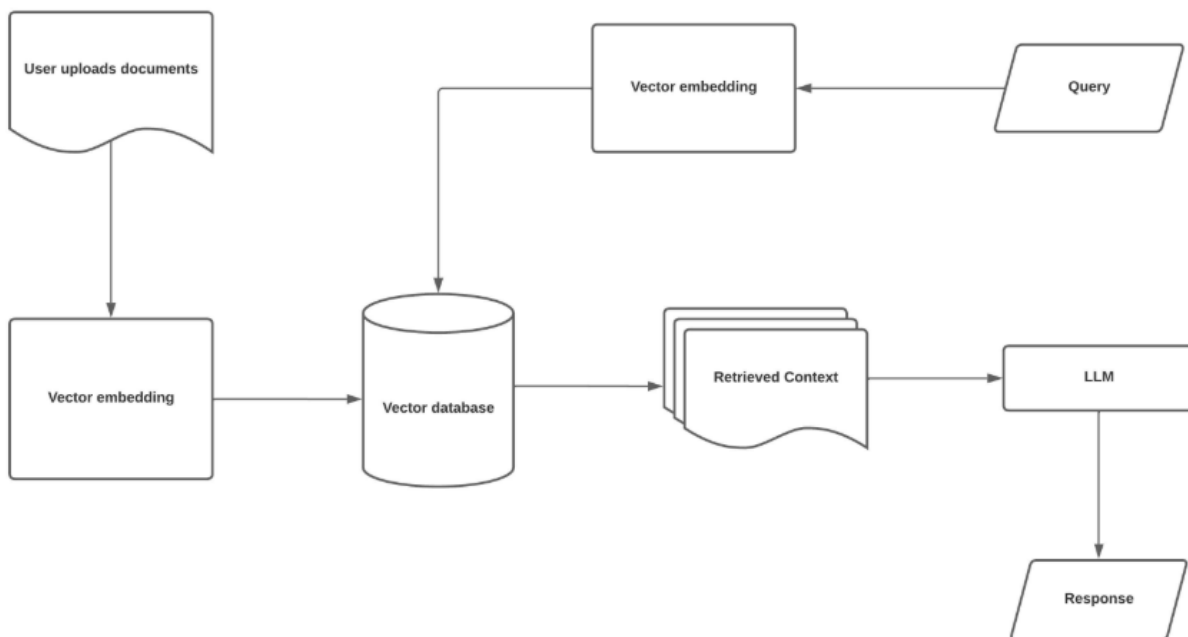
Technology Stack

Component	Technology
Cloud Platform	Azure Databricks (ML workflows), Azure Blob Storage (documents)
AI Model	Llama 2 (via Hugging Face) , GPT-4 (Azure OpenAI)
Framework	LangChain (RAG pipeline)
Web App development	Streamlit (for rapid prototyping) and Azure App Services (deployment)
Vector Database	Azure Search
Document Processing	Azure Document Intelligence

Development Steps

- ✓ Data Ingestion/Preprocessing: Develop a pipeline to parse/preprocess financial documents (PDFs) into a format suitable for the LLM.
- ✓ Embedding generation: Use OpenAI embeddings or Sentence Transformers
- ✓ RAG Setup: Integrate vector DB (Azure AI Search) with LangChain
- ✓ Model Integration: Integrate Llama using Hugging Face's Transformers library and fine-tune it on financial data if necessary.
- ✓ Web Application Development: Build a user-friendly interface using Streamlit that provides following capabilities
 - File upload functionality.
 - Chat interface for queries.
 - Visualization of financial trends.
- ✓ Deployment: Deploy the application on a cloud platform (Azure App Services) for UI access.

RAG Pipeline for FinStatAI



Model Evaluation Criteria

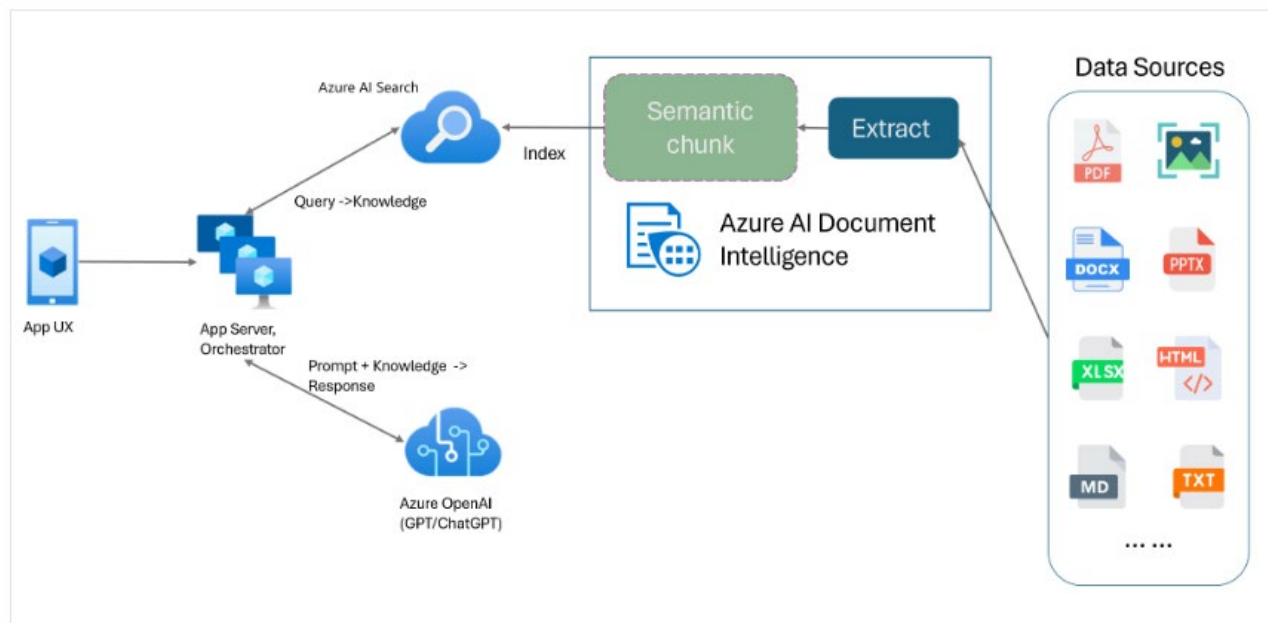
Metrics for Model Evaluation

- ✓ Inference Time: Measure the latency and processing speed of the model when analyzing financial documents.
- ✓ Computational Resource Usage: Monitor CPU/GPU memory consumption during model inference.
- ✓ Model Accuracy: Evaluate the accuracy of the model using metrics such as F1-score for data extraction and ROUGE for summarization tasks.
- ✓ Scalability: Assess the application's ability to handle multiple user requests simultaneously.

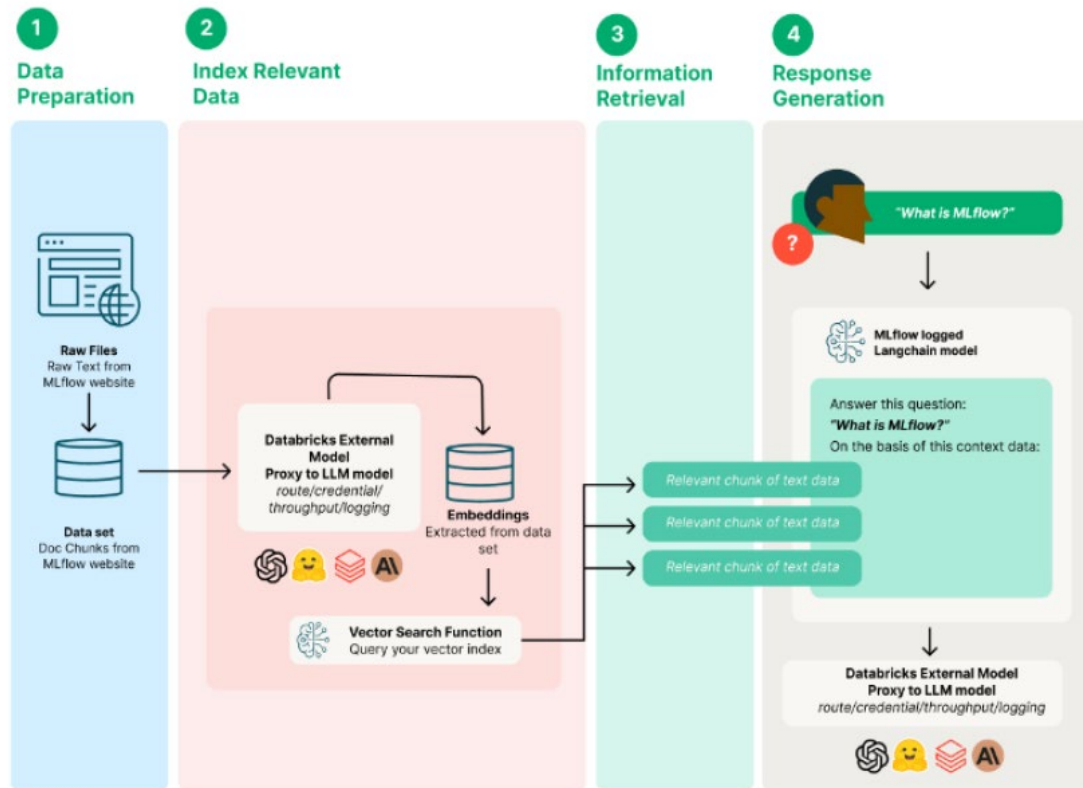
Metric	Target
Latency	< 3 sec response time
Accuracy	> 90% on financial QA benchmarks
Resource Usage	Optimize for Azure GPU instances
User Feedback	Survey usability (1-5 scale)

Deployment Strategy

Data Ingestion/Preprocessing in **FinStatAI** leverages [Retrieval-Augmented Generation with Azure AI Document Intelligence](#)

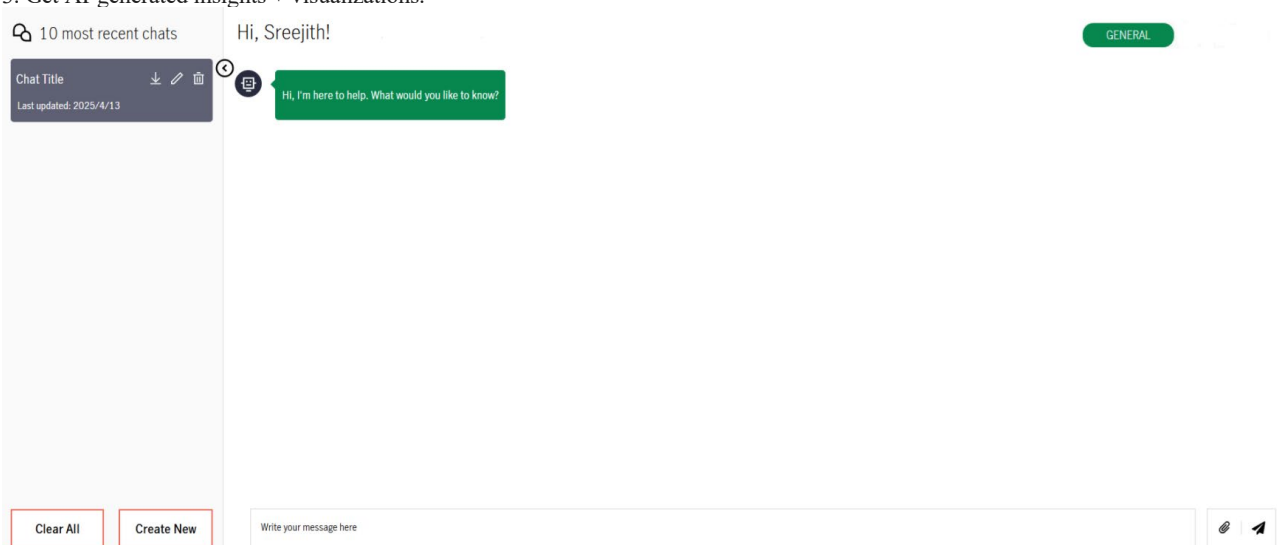


FinStatAI leverages Databricks [Model Serving feature](#) which provides proxy to external models and that allow to streamline the usage and management of various large language model (LLM) providers such as OpenAI.



User Interaction Flow

1. Upload financial documents.
2. Ask questions (e.g., "Show me the trend for 3M stocks in Q4").
3. Get AI-generated insights + visualizations.



Note: More sample outputs and responses accuracy validation information in Appendix Quick Prototyping using [FinStatAI - Streamlit App](#)

Expected Outcomes and Challenges

Expected Outcomes

- ✓ A fully functional **FinStatAI** that automates financial statement analysis.
- ✓ Improved efficiency and accuracy in financial data interpretation.
- ✓ Enhanced user experience with a simple and intuitive interface.

Potential Challenges

- ✓ Data Privacy: Ensuring the security and privacy of financial data uploaded by users.
- ✓ Model Fine-Tuning: Fine-tuning the LLM to accurately interpret financial terminology and context.
- ✓ Scalability: Handling large volumes of financial data and multiple user requests simultaneously.

Possible Solutions

- ✓ Implement robust data encryption and access control mechanisms.
- ✓ Use domain-specific financial datasets for fine-tuning the LLM.
- ✓ Optimize the application for cloud deployment to ensure scalability.

Challenges building FinStatAI	Technology Solution
Scalability	Auto-scale Azure Databricks clusters
Data Privacy	Only use publicly available company statement used with no PII/PHI datasets
Model Hallucinations	Implement RAG for fact-based responses
Document data extraction accuracy	Use Azure Document Intelligence confidence scores (>90%)
Deployment	Azure App Services on separate Azure subscription
Model configurations/tuning	Databricks Model Serving feature
Limited Compute Availability with WPI Student Azure subscription	Leverage additional compute on separate Azure subscription
Unstructured documents/formats and workflows	Leverage Azure Document analysis models Document analysis models
Faster prototyping for UI/UX	Leverage streamlit-community-cloud

Project Resources Required

Tools and Frameworks

Cloud Platform:- Azure Databricks (ML workflows), Azure Blob Storage (documents)

AI Model:- Llama 2 (via Hugging Face) , GPT-4 (Azure OpenAI)

Framework:- LangChain (RAG pipeline)

Web App development:- Streamlit (for rapid prototyping) and Azure App Services (deployment)

Vector Database :- Azure Search

Document Processing:- Azure Document Intelligence

Libraries

- ✓ transformers: This library provides support for loading and working with pre-trained transformer models like BERT.
- ✓ datasets: This library simplifies loading and processing large datasets like Finance Bench.
- ✓ squad_evaluate: This library offers metrics specifically designed for evaluating question answering models.

```

import os
import sys
import json
import pickle
import datetime
import numpy as np
import pandas as pd
from tqdm.notebook import tqdm
from dotenv import load_dotenv

load_dotenv()

#LangChain Stuff
from langchain.document_loaders import PyMuPDFLoader
from langchain.text_splitter import RecursiveCharacterTextSplitter
from langchain.vectorstores import Chroma
from langchain.embeddings import OpenAIEmbeddings
from langchain.chains import RetrievalQA

```

LLM Wrappers/Provides

```

# LangChain Model Wrappers
from langchain.chat_models import ChatOpenAI
from langchain.chat_models import ChatAnthropic
from langchain.llms.replicate import Replicate

# Model Providers
import openai
import anthropic
import replicate
import tiktoken

```

Datasets

Financial statement datasets of public listed companies for fine-tuning and testing the model (e.g., SEC filings, Yahoo Finance data).

Finance benchmark dataset used in this project comprises of following:-

- ✓ Earnings Reports (PDFs) of 40 companies that are publicly traded in the USA and 361 public filings, released between 2015 and 2023
- ✓ 10,231 questions, correct answers, and evidence validated by Subject Matter Experts

Data Summary :-

data/finstatai_document_information.jsonl -- PDFs Meta Data

data/finstatai_open_source.jsonl – Questions / Correct Answers

Data Structure:-

- doc_name (str) Unique Document Identifier. Format: {COMPANY}_{PERIOD}_{TYPE}
- doc_type (str) Type of the Document: {"10K", "10Q", "8K", "EARNINGS", "10K_ANNUAL"}
- doc_period (int) Period of the relevant financial document
- doc_link (str) URL of the relevant document
- company (str) Company
- company_sector_gics (str) Company Sector in terms of GICS standard

Hardware/Cloud Requirements

GPU: NVIDIA GPU for model training and inference.

Cloud Platform: Microsoft Azure for deployment.

Conclusion

FinStatAI aims to revolutionize financial statement analysis by leveraging the power of open-source LLMs.

By automating the extraction, summarization, and interpretation of financial data, it provides users with accurate and actionable insights, saving time and reducing the risk of human error.

While **FinStatAI** has the potential to significantly impact the financial analysis industry, making it more accessible and efficient, it is important to have Human-in-the-loop and remains essential for critical applications, ensuring responsible, trustworthy, and effective use of this powerful technology.

To make sure there is Human In The Loop (HITL), use finance domain experts in reviewing the responses generated by **FinStatAI**. This includes professional roles such as :-

- ✓ Actuaries
- ✓ Finance Portfolio Analysts
- ✓ Finance Risk/Regulatory and Compliance analysts.

The comprehensive approach could result in better risk management, accurate forecasts, and ultimately, more educated financial market decision-making.

References

- ✓ [\[2310.08678\] Can GPT models be Financial Analysts? An Evaluation of ChatGPT and GPT-4 on mock CFA Exams \(arxiv.org\)](#)
- ✓ [\[2310.13001\] Conversational Financial Information Retrieval Model \(ConFIRM\) \(arxiv.org\)](#)
- ✓ [ConvFinQA: Exploring the Chain of Numerical Reasoning in Conversational Finance Question Answering - ACL Anthology](#)
- ✓ [2311.11944 \(arxiv.org\)](#)
- ✓ [Evaluating Open-Domain Question Answering in the Era of Large Language Models - ACL Anthology](#)

Appendix

Model performance of different configurations (when ran against bulk financial statements of 40 companies)

Model Config	Correct Answer	Incorrect Answer	Failed to Answer	Total
GPT4	n/a	9 (6%)	141 (94%)	150
GPT4 - Turbo	14 (9%)	5 (3%)	126 (88%)	150
Llama2	29 (19%)	104 (70%)	17 (11%)	150
Llama3	62 (41%)	81 (54%)	7 (5%)	150
Claude2	114 (76%)	32 (21%)	4 (3%)	150
Claude3	118 (79%)	29 (19%)	3 (2%)	150

Model Test Results – Failed to Answer

Hi, I'm here to help. What would you like to know?

3M_2023Q2_10Q.pdf

Your new document has been successfully uploaded. All responses generated will now pertain to this document.

Does 3M have a reasonably healthy liquidity profile based on its quick ratio for Q2 of FY2023? If the quick ratio is not relevant to measure liquidity, please state that and explain why.

I'm sorry, but I don't have access to specific financial data for 3M for Q2 of FY2023. However, I can tell you that the quick ratio is a measure of a company's ability to meet its short-term obligations with its most liquid assets. A quick ratio of 1 or higher is generally considered healthy, as it indicates that the company has enough liquid assets to cover its short-term liabilities. If the quick ratio is not relevant to measure liquidity, it may be because it doesn't give a full picture of the company's financial health or because other financial metrics are more appropriate for evaluating liquidity. If you have access to 3M's financial statements, I recommend consulting them for a more accurate assessment of the company's liquidity profile.

```
{"financebench_id":"financebench_id_00807","model_name":"llama2","eval_mode":"sharedStore","temp":0.01,"question":"Does 3M have a reasonably healthy liquidity profile based on its quick ratio for Q2 of FY2023? If the quick ratio is not relevant to measure liquidity, please state that and explain why.", "gold_answer":"No. The quick ratio for 3M was 0.96 by Jun'23 close, which needs a bit of an improvement to touch the 1x mark/"}
```

Model Test Results – Correct Answer

```
{"financebench_id":"financebench_id_01858","model_name":"llama3","eval_mode":"sharedStore","temp":0.01,"question":"Does 3M maintain a stable trend of dividend distribution?", "gold_answer":"Yes, not only they distribute the dividends on a routine basis, 3M has also been increasing the per share dividend for consecutive 65 years.", "model_answer":"Yes, 3M has maintained a stable trend of dividend distribution. The company has paid dividends since 1916 and has increased its dividend payout for 65 consecutive years, including a 1% increase in the first quarter of 2023. Additionally, 3M's Board of Directors has consistently declared dividends every year, with the latest declaration being a second-quarter 2023 dividend of $1.50 per share. This suggests that 3M prioritizes sharing its profits with its shareholders and has a long-standing commitment to distributing dividends.", "label":"Correct Answer"}
```

Model Test Results – Incorrect Answer

```
{"financebench_id":"financebench_id_03029","model_name":"llama2","eval_mode":"sharedStore","temp":0.01,"question":"What is the FY2018 capital expenditure amount (in USD millions) for 3M? Give a response to the question by relying on the details shown in the cash flow statement.", "gold_answer":"1577", "model_answer":"The FY2018 capital expenditure amount for 3M is not explicitly stated in the provided cash flow statement. However, we can calculate it based on the information given.\n\nIn the Consolidated Statement of Cash Flows, there is a line item called \"Net cash used in investing activities\" for the year ended December 31, 2018. This line item includes the following items:\n\n* Proceeds from sale of businesses, net of cash sold: $576 million\n* Purchases of property, plant and equipment: $(1,493) million\n* Other investing activities: $(102) million\n\nTherefore, the total amount of capital expenditures for 3M in FY2018 was:\n\n$1,493 million - $576 million + $102 million = $915 million\n\nSo, the FY2018 capital expenditure amount (in USD millions) for 3M is $915 million.", "label":"Incorrect Answer"}
```