

**THE UNIVERSITY OF HONG KONG**  
**DEPARTMENT OF COMPUTER SCIENCE 2024 – 25**



**COMP4801/FITE4801**

**Final Year Project**

Detailed Project Plan

**Group 24079**

Lee Jong Seung (3035555547)

Lee Changjin (3035435840)

Kim Taehyun (3035741330)

**Topic**

LLM-based Real-time Personalized Financial News Notification System

# 1. Project Background

In today's fast-paced financial markets, timely and accurate information is critical for all market participants, such as investors, traders, and financial institutions. However, an enormous volume of daily financial news presents a significant challenge, especially for retail investors: processing vast amounts of data and cherry-picking the most relevant information. While existing financial news platforms offer some level of personalization, they often rely on basic keyword filtering and lack the contextual understanding needed to truly tailor news to an individual's unique portfolio, risk tolerance, or investment goals.

Moreover, the challenge lies in most retail investors' limited time and resources. It is simply insufficient to fully observe the daily changes in financial markets and companies they invest in. According to a web traffic study from FinText (*Who Reads Finance News? Traffic and User Behaviour*, 2023), the average time financial news readers spend on each page of the news platform is about 30 seconds to 1 minute per page, and they visit around 3 to 4 pages daily on average. The study reflects that retail investors trade with an insufficient amount of information, which could mislead them into careless investment decisions. Considering the readership statistics, there is a need for a financial news delivery system that can deliver relevant information more concisely.

Recent advancements in large language models (LLMs) have opened new opportunities for real-time delivery and personalization of text data, as LLMs are capable of processing and understanding complex language structures rapidly. They can be fine-tuned with financial texts to extract relevant information from the news and predict which news may have the most significant impact on a user's portfolio. The news delivery system, with the use of LLMs, enables users to stay ahead of market trends, make informed decisions as quickly as possible, and ask LLMs any questions regarding the news if they have any. The team believes this is a great opportunity to bridge the information gap between retail investors and financial institutions and make a prompt investment decision.

The project 'LLM-based real-time personalized financial news notification system' offers a web application which allows users to input their investment portfolio and then receive a notification about relevant news with positivity & negativity metrics to the user portfolio. Then, the users can ask LLMs any questions regarding the news, just like sending a prompt to other LLM services such as ChatGPT. The application will conduct an embedding search of the news and import real-time data using Refintiv API, which offers various connection methods for reliable news data delivery in real-time. Then, it will process whether the news has any implications for the price of the personal portfolio holdings. LLMs open the opportunity to filter out any noise and let the investors focus on relevant information which affects the price, both positively and negatively.

## 2. Project Objectives

The primary objective of this project is to develop a personalized real-time financial news notification application that leverages LLMs and Natural Language Processing (NLP) to deliver highly relevant and time-sensitive information to investors. This application aims to enable retail investors to make informed decisions quickly and effectively by delivering summarized news updates and insights.

The breakdown of the main objective includes:

### 1. Delivery of personalized notifications

One of the objectives is to provide investors with personalized notifications on financial news that directly affects their portfolio holdings. By leveraging LLMs and NLP, the system will filter and deliver news content tailored to their personalized preferences based on users' inputs regarding their portfolios. This ensures that users receive only the most relevant updates, enabling them to focus on news that matters for their financial decisions.

### 2. Summarization of key information

The application aims to minimize the time retail investors spend reading and analyzing news articles by providing concise summaries. For each relevant article delivered through the notification, LLM-based content summarization will help users quickly grasp key insights and suggest them which articles are worth reading in full. This feature allows investors to make faster decisions without reviewing entire news.

### 3. Analysis of the news and prediction of stock price

After identifying relevant news, LLMs will analyze the content of news articles to predict their potential impact on stock prices. This analysis will serve as a guide for retail investors in their decision-making. It will be conducted in real-time and delivered alongside each news notification, helping users make more informed investment decisions.

### 4. Interaction with LLM for further clarification

To provide further analysis and insight on the selected news delivered to users, our system allows users to gain deeper insights by providing a chat function with LLM. This interactive feature aims to help users better understand the impact of news with more detailed information about certain news.

### 5. Real-time acquisition of data and delivery of notifications

To ensure investors are updated immediately, the application aims to deliver real-time notifications using financial news APIs. The system is planned to poll relevant news sources continuously every few seconds. The combination of rapid news collection, LLM-based filtering, and instant notifications via email or social media messages

ensures that stockholders receive timely updates, allowing them to act promptly on global market events.

By meeting these objectives, the application will assist individual investors in staying up-to-date on financial developments, making timely decisions, and reducing the risks associated with delays in processing financial news.

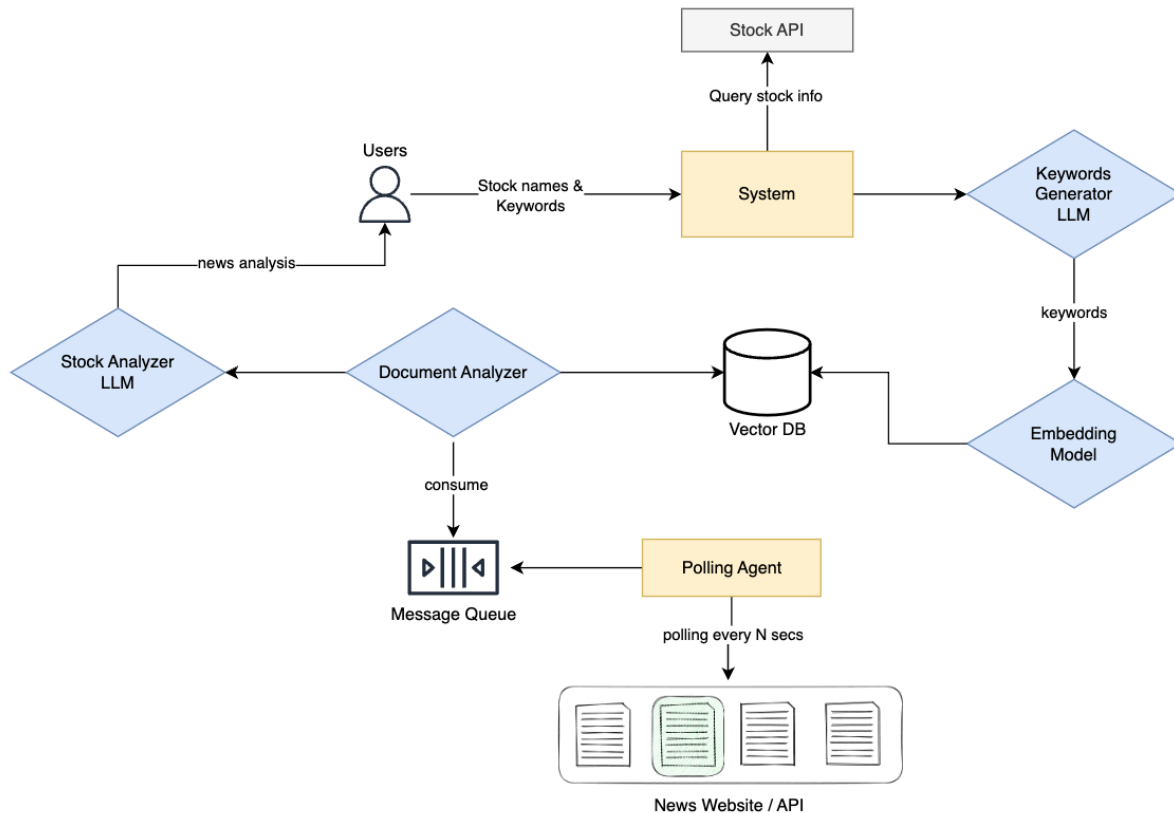
### 3. Project Methodology

This section elaborates on the system architecture, data source, and key system components.

#### 3.1 Data Sources & APIs

The project primarily uses Refinitiv API to retrieve stock information using a stock ticker. Also, Refinitiv provides real-time financial news articles with the full content. The polling agent continuously sends requests to Refinitiv API to extract the financial news articles. Moreover, the custom web scraper will poll extra data sources such as Yahoo Finance website as a supplementary data source if the news data from Refinitiv API is insufficient..

#### 3.2 System Architecture



**Figure 1: System Architecture**

The project is composed of a number of system components to deliver a personalized and real-time financial news notification system (Figure 1).

Upon user registration, users can provide stock names or keywords to the system for personalized notification. If stock names are provided, the system queries more detailed stock information, such as stock industry and related keywords. Then, the fine-tuned large language model generates example keywords likely to impact the provided stock. For example, if a user enters 'Coca Cola', the generated example keywords may include 'beverage industry trends', 'sugar tax impact', or 'global supply chain challenges'. These keywords are then converted into embeddings and stored in the Vector DB, which will be used for embedding search with the financial news article to extract only the relevant ones.

To ensure real-time delivery, a polling agent continuously polls the financial news website with a short interval (about 10 seconds). If there are any new articles, the agent will place them on the asynchronous message queue like RabbitMQ. Next, the Document Analyzer, continuously consuming the message queue, performs an embedding search between the article and each user's stored keyword embeddings to calculate the relevance score.

If relevant, then the article is passed to the Stock Analyzer LLM which is fine-tuned to generate the news analysis including a brief summary and how the article might impact the user's stock portfolio, which will be delivered to the user via email or SNS channels.

### 3.3 Development Environment

The backend of the project will be primarily developed using Java and Spring Boot, ensuring robust Object-Oriented software. The user interface will be built with Javascript and React. Python and FastAPI will power the servers for various large language models. Also, AWS Simple Queue Service(SQS) is chosen as the asynchronous message queue used in the project. The infrastructure of the project will be hosted primarily on AWS for its scalability and reliability.

## 4. Project Schedule & Milestones

Phase 1: Inception (1 Oct, 2024 - 17 Jan, 2025)

### 1 Oct - 31 Oct, 2024

- Complete system architecture design and review on a full-scale
- Test Refinitiv API and validate any alternative data source if required
- Test embedding search model

### 1 Nov - 30 Nov, 2024

- Complete UI Design with Figma

- Database Schema Design
- Backend implementation of user authentication and stock keyword input functionality
- Infrastructure setup with AWS

### **1 Dec - 31 Dec, 2024**

- Backend Implementation and integration with stock API
- Implementation of the Keywords Generator LLM
- Implementation of the embedding generator model
- Vector database infrastructure set up
- Front-end baseline setup

### **1 Jan - 31 Jan, 2025**

- Frontend development
- Backend development
- Implementation of polling agent and integration with the message queue
- Implementation of the embedding search model

Phase 2: Elaboration (18 Jan, 2025 - 20 Apr, 2025)

### **1 Feb - 28 Feb, 2025**

- Frontend development
- Backend development
- Implementation of prompt engineering on LLM summarization on news articles
- Implementation of Document Analyzer

### **1 Mar - 31 Mar, 2025**

- Implementation of the backend of the LLM chat function
- Test and experiment with the Stock Analyzer LLM (summary generation LLM & stock price impact analysis NLP model)

### **1 Apr - 20 Apr, 2025**

- Implementation of the frontend of the LLM chat function
- Integration test of the overall system
- Continuous testing and enhancement of the language models of the system

Phase 3: Construction (21 Apr, 2025 -)

### **21 Apr - 30 Apr, 2025**

- Preparation of final presentation and project exhibition

## 5. Citation

*Who Reads Finance News? Traffic and User Behaviour*. FinText. (2023, February 21).  
<https://www.fintext.io/case-studies/benchmarking/who-reads-financial-news-web-traffic-and-user-behaviour/>