

**Team name:**

Finance Bull 

**Name & USC-ID:**

Chu-Huan Huang, 1934208430

Chung-Sheng Chang, 1676324151

Jiaying Wang, 3018182967

**Domain we decided on:**

Stock trading, finance, and economics

**Shortlisted options for kinds of data to incorporate:**

**Stock Market Data:** Historical and real-time stock prices, including metrics such as open, close, high, low, and volume, are essential for analyzing market trends, price movements, and trading patterns. This data will enable the chatbot to provide users with insights into stock performance, technical analysis indicators, and market sentiment over different timeframes. By incorporating stock market data, the chatbot can help users make informed decisions about portfolio management and investment strategies.

**Macroeconomic Data:** Economic indicators such as GDP, inflation rates, interest rates, and employment statistics are crucial for understanding the broader economic environment that impacts financial markets. These macroeconomic factors influence investor confidence, business cycles, and stock market volatility. By integrating this data, the chatbot can offer insights into economic trends, helping users assess market conditions and make strategic investment decisions based on economic forecasts and policy changes.

**Financial News and Reports:** Earnings reports from publicly traded companies and financial news from reputable sources such as Bloomberg and Reuters provide valuable insights into corporate performance, industry trends, and market-moving events. These reports help investors stay updated on key developments, such as mergers and acquisitions, regulatory changes, and financial forecasts. Incorporating financial news will allow the chatbot to deliver timely and relevant information to users, helping them stay informed and respond to market changes effectively.

**Investor Sentiment and Opinion:** Sentiment analysis from social media platforms like Twitter and Reddit, along with financial blogs and discussion forums, provides insights into market psychology and investor behavior. This data helps capture trends driven by retail and institutional investors, offering a

unique perspective beyond traditional financial metrics. By incorporating sentiment analysis, the chatbot can identify market hype, potential bubbles, or undervalued opportunities, allowing users to gauge market sentiment and adjust their strategies accordingly.

## **Publicly available dataset links**

### **1. CNBC Markets (<https://www.cnbc.com/markets/>)**

This source provides real-time and historical data on major stock indexes such as S&P 500, NASDAQ, DJIA, and FTSE. It includes stock prices, percentage changes, market movers (top and bottom performers), and financial quotes like bonds and cryptocurrencies. The data is useful for tracking short-term and long-term market trends.

#### Sample Excerpt

S&P 500: 6,086.37 (+37.13, +0.61%)

NASDAQ: 20,009.34 (+252.56, +1.28%)

DJIA: 44,156.73 (+130.92, +0.30%)

FTSE: 8,545.13 (-3.16, -0.04%)

### **2. Yahoo Finance (<https://finance.yahoo.com/>)**

Yahoo Finance offers comprehensive financial news, stock market data, economic reports, and portfolio tracking tools. The platform provides live stock quotes, historical data, company earnings reports, and the latest news on market trends, making it a valuable resource for financial education and analysis.

#### Sample Excerpt

Stock Market Today:

- Asian shares are mixed after China rolls out market-boosting measures.
- Microsoft's stock surges 4.13% following news of strategic partnerships.
- Netflix hits all-time closing high after near flawless earnings report.

### **3. FRED (Federal Reserve Economic Data) (<https://fred.stlouisfed.org/>)**

This platform offers macroeconomic data, including GDP, employment statistics, inflation rates, and interest rates. The data is updated regularly and provides valuable insights into economic indicators that influence financial markets.

#### Sample Excerpt

Gross Domestic Product (GDP)

Q3 2024: 29,374.914 (Billions of Dollars, Seasonally Adjusted Annual Rate)

Updated: Dec 19, 2024

### **4. SEC EDGAR Filings (<https://www.sec.gov/search-filings>)**

The SEC's EDGAR system provides public access to financial filings from companies listed on U.S. stock exchanges. These include 10-K annual reports,

10-Q quarterly reports, and 8-K current reports, offering critical financial and operational insights about publicly traded companies.

#### Sample Excerpt

Tesla, Inc.

Filing: January 2, 2025 - 8-K: Current Report

Earnings Release: Revenue of \$25.1 billion, up 17% year-over-year

Net Income: \$3.2 billion

### **5. Reddit - r/algorithmtrading (<https://www.reddit.com/r/algorithmtrading/>)**

This subreddit contains discussions, recommendations, and sentiment analysis related to algorithmic trading. The data includes user-generated content, trading strategies, and market opinions, which can be useful for understanding investor sentiment and emerging trading trends.

#### Sample Excerpt

User: gotchab003

Post: "Books you'd recommend to someone getting started in algorithmic trading? I'm a software developer and want to learn the basics of trading... Any suggestions?"

### **6. Alpha Vantage (<https://www.alphavantage.co/>)**

Alpha Vantage offers real-time and historical stock market data APIs covering a wide range of financial instruments, including stocks, options, forex, and cryptocurrencies. The platform provides access to over 60 technical and economic indicators, market news APIs, and sentiment analysis tools with global coverage.

#### Sample Excerpt

Symbol	Date	Open	High	Low	Close	Volume
TSLA	2025-01-22	640.50	650.00	635.00	645.20	5,000,000
AAPL	2025-01-22	180.00	182.50	178.90	181.75	2,345,678

### **Script Description**

The python script collects and processes data from three different sources: stock market data from Alpha Vantage, financial news from Yahoo Finance, and financial reports from SEC EDGAR filings. It uses libraries such as requests, pandas, BeautifulSoup, and pdfplumber to retrieve, clean, and store the data in structured formats.

First, the script fetches Tesla's stock market data via the Alpha Vantage API, retrieving key metrics such as open, high, low, close prices, and volume. The final dataset is stored as TSLA\_stock\_data.csv for further analysis. Basic operations such as viewing the first few records, checking dimensions, and identifying missing or duplicate data are performed.

Next, the script scrapes financial news from Yahoo Finance using BeautifulSoup. It extracts headlines, links, and publication timestamps, storing the information in `yahoo_finance_news.csv`.

Finally, the script processes a financial report in PDF format downloaded from the SEC EDGAR database. Using pdfplumber, it extracts text from each page and saves it to `extracted_text.txt`. The script then analyzes the extracted content by counting pages, total characters, empty pages, and calculating the average number of words per page.

## **Enhancing Chatbots**

Existing financial chatbots face several limitations that hinder their ability. One of the key challenges is their reliance on static or delayed data, which limits their ability to deliver real-time market insights. Our dataset addresses this issue by integrating stock market data from Alpha Vantage, allowing the chatbot to access up-to-date information on stock prices, trading volumes, and market trends. This real-time data retrieval enables the chatbot to provide users with timely and relevant market insights, helping them make informed decisions based on the latest financial data.

Another limitation of current chatbots is their inability to extract and analyze relevant financial news from web sources in an automated manner. Our solution overcomes this challenge by utilizing web scraping techniques to collect financial news articles from Yahoo Finance. By extracting key details such as headlines, publication sources, and timestamps, our chatbot can keep users informed about recent market developments and trends. This capability enhances the chatbot's ability to provide context-aware recommendations by incorporating the latest news alongside stock market data.

Additionally, financial chatbots often struggle with processing complex financial documents, such as earnings reports and regulatory filings, which contain valuable insights crucial for investment decisions. Our approach leverages PDF extraction techniques to analyze reports from the SEC's EDGAR database. By extracting textual data from these reports, we can identify key financial metrics, risk factors, and strategic insights that would otherwise require manual analysis. This automated document processing improves the chatbot's ability to offer deeper insights into a company's financial health and performance.

Name: Jiaying Wang

USCID: 3018182967