Alexey Akopov  
CAP 5619  
3/26/2025

LLM 8-k Extraction Report

The objective of this assignment was to develop a Python script capable of automatically extracting new product or service announcements from SEC filings (initially 8-K reports). The extracted information specifically includes the product name and a corresponding product description. To achieve this, I employed web scraping techniques combined with natural language processing capabilities provided by Large Language Models (LLMs).

Initially, I experimented with a local LLM—DeepSeek—to process and interpret the filings. However, after multiple attempts, I found DeepSeek’s performance insufficient, as it often produced inaccurate or irrelevant results. Consequently, I transitioned to using OpenAI's ChatGPT API, which significantly improved the accuracy and reliability of the outputs.

The implemented script performs the following core functions:

**1. Data Retrieval:**

- Reads a CSV file containing company tickers and CIK numbers.

- Fetches the most recent 8-K filings from the SEC website.

- Specifically identifies and retrieves documents labeled EX-99.1, which typically contain press releases or announcements.

**2. Natural Language Processing:**

- Uses ChatGPT to analyze the fetched documents.

- Instructs the LLM via a tailored prompt to identify explicitly mentioned new products or services and their descriptions.

- Ensures output consistency by specifying a structured JSON response from the model.

**3. Data Processing and Storage:**

- Parses the JSON responses from ChatGPT to extract structured information.

- Compiles the results into a coherent CSV file for further analysis.

The final implementation successfully extracted new product announcements from various companies, confirming the effectiveness of switching from a local LLM to OpenAI’s API. The final dataset (CSV) clearly lists each company, ticker symbol, new product identified, and the product description, serving as a valuable tool for financial analysis and market research.