Financial Services Innovation Lab, Georgia Tech
Programming Task for Summer Research

This task is split into multiple parts or steps to make it easier for you to execute in step by step fashion. They are not separate tasks. It is due May 7th.

Our hope is that even if you are not selected, the project would be useful for you to showcase of your skills.

Submission Details

- Submit through GitHub. Upload all your work in a public GitHub repository and once you are done with your work share the GitHub repository URL.
- Feel free to use any programming language and tech stack of your choice.
- You have to submit ONLY
  - The program, results, visualization, and URL to deployed dashboard
  - **IMPORTANT**: Writing the correct code is necessary but not sufficient to complete the task. We will also look at how you document the code (README file, docstrings, comments, etc.). The README file should include full details of the tech-stack and a 1-2 line explanation for why you chose the particular component of the stack. Follow best coding practices.
- Note that you do not need to submit any datasets for this task
- **IMPORTANT**: Please note that this task requires some self-study and research; it is imperative that students begin working on this task early

Task 1.1: Download Data from the SEC-EDGAR

The first part of the task is to download some SEC 10-K filings of public firms. These documents will be analyzed in the next part of the task.

- First, select 2-3 companies or tickers of your choice.
- Second, for a given company ticker from a user, you need to download a 10-K filing for each year from the SEC website for the period 1995 through 2023 (resulting in approximately 29 10-K documents, it can be less depending on when the company IPOed).

- Note that the task is to write a script or program to do this automatically, and not to download all files manually.
- You can use the sec-edgar-downloader (`https://sec-edgar-downloader.readthedocs.io/en/latest/`) package in Python for ease or other similar packages.

Task 1.2: Text Analysis

- Note: You have to submit ONLY
  - The program is used to merge data, clean data, perform analysis, and generate visualization.

The main task for this part is to conduct text analysis using LLM API available for free.

- There are a few services that provide LLM inference API along with some free credits to use. You don't need to use any paid service.
- Use LLM inference API to gather some information or generate some insights from the 10K filings for a given company.
- note: we are not giving instructions on a specific "insight". The judgment calls you need to make to decide on a "good insight" is part of the assignment scoring. The assignment is deliberately vague as we want you to make some good judgment calls.
- give 1-2 line explanation on why a user would care about this insight.
- Construct a visualization from the generated insight/information.

Task 2: Construct and Deploy Simple App

- Using the Task 1 code as the backend, create a simple app that takes the company ticker as input and displays some visualization.
- Feel free to use any tech stack for this task. But as mentioned before, document in the README file and give a 1-2 line explanation on the rationale for the choice.
- We would like to see the app deployed on some platform if not, please record a short video of a local demo of the app.