**Q: Executive Summary (Group)**: A brief overview of what has been accomplished this week.

A:

* Data：
  + Collected annual reports for 10 target companies in the past 5 years (together)
* Research
  + Research and finalized GPU infra (Yi Lu)
  + Research and finalized Vector Database (yanni)
  + Research and finalized LLM (alisa)
  + Research and developed Langchain (jasper)
  + Research and figured out KPIs to measure (michell)
  + Study on FAISS nearest neighbor algorithms (alisa)
  + Github account creation and structure (michell)
* Code
  + Q&A model
  + Conversational model
  + (Yi Lu, jasper combined and wrapped up parts by yanni, alisa, michell & all reviewed together)

**Q: Challenges and Roadblocks (Group)**: Enumerate any obstacles faced and how they were managed or are planned to be managed.

A:

1. Current structure (Loader/LLM) is not good at getting information from tables and graphs, and thus sometimes fails to find the correct answers to the questions.

Solution：Separate the tables (numerical data) and text in the document using OCR and related techniques, feed different types of data into the model respectively.

1. FAISS, as one of the most advanced similarity search techniques is a library but not a traditional vector database, which we are not familiar with and not sure whether it can perform well on our use-case.

Solution：List the basic qualification of our use-case and compare the performance of FAISS with traditional vector databases that we have hands-on experience with.

**Q: Next Steps (Group)**: Outline the plan for the upcoming week.

A:

1. Selection

* - Add more data to the database (sustainability reports etc.)
* - Comparative analysis on vector databases, retriever algorithms, and LLM models
* - Visualization in UI

1. Development

* - Build a summarization model and prepare a demo to showcase

1. Optimization

* - Improve model itself by RLHF (Reinforcement Learning with Human Feedback)
* - Prompt optimization - see if we can get better answers
* - Table/graph interpretation - split data (text/numerical/table)

**Q: Executive Summary (Individual)**: A brief overview (one or two sentences) of what has been accomplished this week. Group coordinators collect and report the progress made by individual team members.

A:

**Yi Lu (yl5118)**: collected 2 of 10 companies annual report; researched and finalized GPU infrastructure, researched and finalized UI; combined and wrapped up parts by yanni, alisa, michelle & all reviewed together

**Yanni Chen (yc4179)**: collected 2 of 10 companies annual report; researched and finalized on Vector Database; Wrote code on the vector database part and reviewed code with the team

**Xiaolin Sima (xs2483)**: collected 2 of 10 companies annual report; researched and finalized the LLM, Studied on FAISS nearest neighbor algorithms. Wrote code on the Large Language Model part and reviewed code with the team

**Junyuan Huang (jh4608)**: collected 2 of 10 companies annual report; researched and developed on Langchain to process documents; generated the framework of the bots, combined and wrapped up parts by yanni, alisa, michelle & all reviewed together

**Michelle Sun (ms6514)**: collected 2 of 10 companies annual report; researched and figured out KPIs to measure company on the prompt part and reviewed code with the team