[**Form link**](https://docs.google.com/forms/d/e/1FAIpQLScqG543NuDmitKEqycYNpbydC5h_IlBmrskmS7gKlk5xCFP1g/viewform)

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

A:

**Data**

Collected quarterly reports for target companies (All)

**Visualization**

Draw flow graph of summarization model (Alisa & Yanni)

**Code**

1. Summarization: used bigger model (flan-t5-xxl) and did prompt engineering to improve output quality (Alisa & Yanni)
2. Performed comparative analysis on pdf loader and document splitter; optimized text and image query (Yi)
3. Used GPT-3.5 to convert documents into Q&A pairs (Jasper)

**Research**

Researched on Reinforcement Learning with Human Feedback

**Report**

Work on writing the 1st progress report (All)

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

A:

**Problem**

Summarization output is still not ideal, but it is because of the structure of the reports.

**Solution**

Our mentors suggested to first find the major KPIs using the Q&A model, and then input the results into the summarization model to create a summary on company performance

**Problem**

Although we found the solution to transform unstructured images in pdf to structured text, the problem of table translation remains. Other techniques is necessary to further process the data.

**Solution**

One solution is to separately process structured data (pdf table and text table resulted from image) using some external python packages that translate tables to pandas df.

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

A:

**Summarization**

Try to use KPI results from Q&A model to generate summaries and check its performances.

**Optimization**

Try different methods to have a better representation of complex data structures in pdf.

**Research**

1. Apply semantic index in FAISS to improve its efficiency
2. When the pdf preprocessing workflow is confirmed, compare different models on Q&A and conversation tasks.
3. Improve model itself by RLHF (Reinforcement Learning with Human Feedback)

**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 quarterly reports for target companies; performed comparative analysis on pdf loader and document splitter; optimized text and image query; worked on writing the 1st progress report

**Yanni Chen (yc4179)**: Collected quarterly reports for target companies; draw flow graph of summarization model; used bigger model (flan-t5-xxl) and did prompt engineering to improve output quality; worked on writing the 1st progress report

**Xiaolin Sima (xs2483)**: Collected quarterly reports for target companies; draw flow graph of summarization model; used bigger model (flan-t5-xxl) and did prompt engineering to improve output quality; worked on writing the 1st progress report

**Junyuan Huang (jh4608)**: Collected quarterly reports for target companies; used GPT-3.5 to convert documents into Q&A pairs; worked on writing the 1st progress report

**Michelle Sun (ms6514)**: Collected quarterly reports for target companies; researched on Reinforcement Learning with Human Feedback; worked on writing the 1st progress report