**Project Proposal**

**Ahreum Cho, Hoyoung Jang, Liam Monninger**

1. **Role Assignments**
   1. Data Wrangling : ~~
   2. Building model & ~~
2. **Data Source**
   1. Additional data we have searched for
      1. [Google Cloud console](https://console.cloud.google.com/marketplace/product/global-patents/labeled-patents?organizationId=731690703799)
      2. <https://evasharma.github.io/bigpatent/>
      3. [big\_patent](https://www.kaggle.com/datasets/jackchungchiehyu/big-patent)
3. **Project Plan**
   1. Explain what you intend to study with your project.
      1. We intend to study ~~
   2. What is the ultimate objective?
      1. We aim to assist individuals who are planning to file a patent or conducting patent research by recommending appropriate research papers or relevant patents through our big data analysis project.
   3. What types of models are you considering?
      1. Before building a model, we would identify common themes or topics, and use natural language processing to extract key information from data in arXiv papers, as well as patent descriptions from additional patent data.
      2. To build the recommendation model, we are considering creating a knowledge graph and implementing a BFS (breadth-first search) greedy search algorithm.
      3. We may build simple interactive recommendation system where individuals can put keywords in
4. Why is this project interesting?
   1. ~~
5. What challenges and obstacles might you anticipate with this project?
   1. ~~