



*Figure 1: Pipeline*

BackTrack and IDI are international tracking and campaigning organizations that track the companies that private commercial banks loan to in order to trace the financial accountability of projects with negative societal effects. To facilitate the process of tracking banks, they asked the University of Chicago's Data Science clinic to develop an automated method of collecting loan information from financial documents. This is important to them as it saves them valuable time which could be put towards advocacy and campaigning.

The team developed a pipeline to ingest financial documents and train models on these documents. The trained models in this pipeline can then be used to create a database that contains companies' loan information. In the process of training these models, the team ran into issues with model performance, potentially related to limited data for fine tuning these models.

The initial pipeline for ingesting documents and training models was completed by the end of the quarter. Future work could focus on methods for improving model performance such as creating more training data.