# CS 670: Artificial Intelligence

# Finetuning Language Model – Project Report

**Submitted By:**

Fouzia Jafarullakhan (fj9)

## Introduction:

In this project, Harvard USPTO Patent dataset is used to develop required classifier using a language model that is trained to do sentiment analysis and provide a patentability score for each patent application submitted in Jan2016.

## Code:

* First import all the required libraries to run the code and start the streamlit application Streamlit, datasets, transformers, pandas, and torch
  + Next step is to load the HUPD dataset from hugging face datasets and not the complete dataset as it is huge. The requirement of the project was just to be use Jan 2016 patent applications. So, I am loading just the sample and loaded all of it into training batch.
  + I am creating a dataframe from the loaded dataset so I can filter on columns that I just need, and the data is in a format I desire so it is easy for me to work with.
  + I load just the patent number column values into PAN variable so I can use that to display as a drop-down option on streamlit app.
  + Next piece of code is to use streamlit app commands to display required information on the streamlit app, starting with title for the app, then a form to display patent number as a choice for user to select from.
  + Once, user selects the patent number and clicks on submit button.
  + The selected patent number and its corresponding decision is fed into a sentiment analysis classifier.
  + I tried to use the abstract and claim fields as well but saw that decision field gave more probability range in terms of the results observed and was closer to the required results.
  + Have used the distilbert model for sentiment analysis.
  + Converting the decision field to string and feeding it to the desired classifier gives the probability and have used a for loop to display the value as the score of the patent application selected.
  + Next two blocks of code is to display the abstract and claims of the patent application selected.