**Project Title:** Enhancing Legal Document Summarization with LLM

**Problem Setting:**

Given a collection of legal documents D=d1,d2,...,dND=d1​,d2​,...,dN​, where each document didi​ is a comprehensive text composed of legal terminology and complex sentence structures, our goal is to develop a Large Language Model (LLM) based system f(D)f(D) that generates concise, accurate summaries S=s1,s2,...,sNS=s1​,s2​,...,sN​ for each document. The challenge involves not only extracting key information but also preserving the legal context and nuances crucial for understanding. Our objective is to facilitate faster and more efficient review processes for legal professionals, enhancing accessibility to vital information.

Dataset: <https://github.com/rezuanul/awesome-legal-nlp>

**Evaluation Protocol:**

* We will split our dataset into a training set (80%), a validation set (10%), and a test set (10%). This distribution ensures that the model is trained on a large corpus while still allowing for robust evaluation and testing.
* The model's performance will be evaluated using ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores to measure the quality of the summaries against human-generated reference summaries. Specifically, we will focus on ROUGE-N (for n-gram overlap) and ROUGE-L (for longest common subsequence) to assess both precision and recall.

**Machine Learning Model & Architectures:**

* We will utilize a pre-trained Large Language Model (such as GPT-3 or a suitable variant) as our base, leveraging its extensive knowledge base and natural language understanding capabilities.
* Fine-tuning will be performed on a legal document dataset to adapt the model to the specific nuances and vocabulary of legal texts.
* A comparison will be made against a baseline model, such as a simpler extractive summarization technique using TF-IDF scores to evaluate the enhancement provided by the LLM.

**Baseline for Comparison:**

A "simple" baseline model will involve extracting key sentences from each document based on TF-IDF scoring, serving as a comparative measure to showcase the advanced understanding and summarization capabilities of our LLM-based approach.

**Tuning & Experimentation:**

* Hyperparameters of the LLM, such as learning rate and number of fine-tuning epochs, will be optimized based on validation set performance.
* Exploration of different summarization strategies (e.g., extractive vs. abstractive summarization) within the LLM framework to identify the most effective approach for legal documents.