**Project Report: Fine-Tuning Llama-3.1 with LoRA for Legal Title Generation**

**Aryan Sai Gati DSE-B REG NO. 220968104**

**Problem Description:**

The goal is to use the NVIDIA NeMo framework to fine-tune the Llama-3.1 model with LoRA adapters for generating meaningful and legally relevant titles from detailed legal questions in the Law StackExchange repository. The model aims to transform complex legal queries into concise, readable titles that increase engagement and maintain critical legal information.

**Dataset:**

* **Field:**
  + **Question**: Detailed legal questions asked in the forum.
  + **Additional Metadata**: Tags or timelines are available but will not be used directly in this project.
* **Determine if there is a model in terms of length or complexity**: Analyse the length of questions and their complexity.
* **Relationship Between Question and Title**: Investigate if there is a correlation between specific words or phrases in the question and their occurrence in the title.

**Example:**

* **Input Question**:

In order to be sued in a particular jurisdiction, say New York, a company must have a minimal business presence in the jurisdiction. What constitutes such a presence? Suppose the company engaged a New York-based Plaintiff, and its representatives signed the contract with the Plaintiff in New York City. Does this satisfy the minimum presence rule? Suppose, instead, the plaintiff and contract signing were in New Jersey, but the company hired a law firm with offices in New York City. Does this qualify?

* **Output Title**:

What constitutes "doing business" in New York?

**Preprocessing:**

* **Text Cleaning**:
  + Remove special characters, HTML tags, and improper whitespace from the question.
  + Convert the text into tokens (sub words) compatible with the Llama-3.1 tokenizer.
* **Example Transformation**:

QUESTION: In order to be sued in a particular jurisdiction, say New York, a company must have a minimal business presence in the jurisdiction. What constitutes such a presence? Suppose the company engaged a New York-based Plaintiff, and its representatives signed the contract with the Plaintiff in New York City. Does this satisfy the minimum presence rule? Suppose, instead, the plaintiff and contract signing were in New Jersey, but the company hired a law firm with offices in New York City. Does this qualify?

**Output**:

TITLE: What constitutes "doing business" in New York?

**Dataset Split:**

* **Training Set**: 70% of the data.
* **Validation Set**: 15% of the data.
* **Test Set**: 15% of the data.

**LoRA Custom Adapter:**

* **Method**: Utilize LoRA’s PEFT (Parameter-Efficient Fine-Tuning) to adapt the Llama-3.1 model to generate legal titles efficiently.

**Performance Metrics:**

* **ROUGE Scores**:
  + **ROUGE-1**: Measures unigram overlap between generated and reference titles.
  + **ROUGE-2**: Measures bigram overlap.
  + **ROUGE-L**: Measures the longest common subsequence overlap.
  + **Example**:
    - **Generated Title**: "What constitutes 'doing business' in New York?"
    - **Reference Title**: "Legal criteria for 'doing business' in New York"
    - ROUGE-1 Score: 45.6
    - ROUGE-2 Score: 22.1
    - ROUGE-L Score: 40.8
* **Precision, Recall, and F1 Scores**: Evaluate the accuracy of generated titles in capturing legal terms.
* **Human Evaluation**: Legal experts assess the generated titles for clarity, accuracy, and relevance.

**Define Project Goals:**

* **Fine-tuning the Model**: Achieve improved performance in generating accurate and relevant legal titles using the Llama-3.1 model fine-tuned with LoRA on the Law StackExchange dataset.
* **Deployment**: Provide a fine-tuned LoRA model using NVIDIA NIM for scalable and efficient real-world deployment.
* **Usability**: Ensure the model effectively generalizes across various legal topics, assisting forum users in generating appropriate titles for their legal questions.