**Week 10: Lab Programs**

**Question 1 -3: Deep Learning**

**Q1. Word Prediction Using GPT (Next Word Prediction)**

**Objective:** Implement a next-word prediction model using GPT-2 or GPT-Neo pretrained models. Consider the Dataset: **Use Hugging Face WikiText-2**

**Task:**

* Load the wikitext dataset and preprocess it by tokenizing with a GPT-compatible tokenizer (e.g., GPT2Tokenizer from Hugging Face).
* Use a pretrained GPT2LMHeadModel to fine-tune on a small portion of WikiText-2. Train for 1-2 epochs and generate predictions.
* Write a function that takes an input sentence and predicts the next 5 words.
* Evaluate model perplexity on a small test set.

**Q2. Language Translation Using Transformer Encoder-Decoder (T5 or MarianMT)**

**Objective:** Translate English to French or German using a pretrained translation model. **Dataset:** Use Hugging Face OPUS Books (or any parallel text dataset), Alternatively: ManyThings (eng-fra)

**Task:**

* Load a pretrained model such as:
  + Helsinki-NLP/opus-mt-en-fr or
  + t5-small (for general tasks like "translate English to French: ...")
* Preprocess dataset into sentence pairs and tokenize both source and target.
* Perform batch translation and calculate BLEU score on the test set.
* Take user input and generate translation output.

**Q3. Masked Word Prediction Using BERT**

**Objective:** Use BERT for masked word prediction (fill-in-the-blank task).

Consider the **Dataset:**

* Use a small custom text dataset or Hugging Face wikitext again.

**Task:**

* Load bert-base-uncased model and BertTokenizer.
* Create masked sentences, e.g.,
* Use BERT to predict the [MASK] token and list top-5 predictions with probabilities.
* Test accuracy on a few custom test sentences where the ground truth is known.