**Natural Language Processing (NLP)**

Tokenization (WordPiece, BPE), word embeddings (word2vec, GloVe, FastText), sequence models (BiLSTM, GRU), attention mechanisms, Transformers (BERT, RoBERTa, DistilBERT), fine-tuning for classification, QA, and NER, HuggingFace ecosystem (datasets, tokenizers, transformers), evaluation metrics (BLEU, F1, perplexity), retrieval-augmented generation (RAG), intro to vector databases (FAISS, Chroma)

**Task 1 (due July 25 EOD):**

*Course:*

"Natural Language Processing Specialization" by deeplearning.ai

Also, learn everything you can about RAGs

*Model Development:*

Develop one Transformer-based model (e.g., using BERT or RoBERTa) for Named Entity Recognition (NER) on the CoNLL-2003 dataset. Submit the model code along with a summary of the results.

*Papers:*

* "Attention Is All You Need" by Ashish Vaswani et al. (2017)
* "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" by Jacob Devlin et al. (2018)
* "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks" by Patrick Lewis et al. (2020)

Provide a brief report on each paper, summarizing its key contributions, methodologies, and findings.

**Task 2:**

This task focuses on core research thinking and independent academic exploration.

### *Task:*

Dive deep into a specific research area within your chosen specialization (CV, NLP, RL, or Gen-AI). Your objective is to:

* Study Research Papers related to your domain (As many as needed)
* Identify a research gap or a novel question
* Propose how you plan to proceed in that direction

### *Deliverables:*

Prepare a well-structured research report including:

1. Title and Research Area: Clearly define the domain/topic you're focusing on.
2. Background: Provide a brief introduction and motivation for the problem.
3. Related Work: Summarize at least 3–4 relevant research papers you've studied.
4. Identified Gap: Highlight the limitations or gaps in current literature.
5. Proposed Direction: Describe your proposed idea, method, or approach.
6. Next Steps: How you plan to proceed (implementation, experimentation, etc.)

## Submit your report as a PDF, docx, latex or markdown. Latex is preferred.