

KATHMANDU UNIVERSITY
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

Subject: Large Language Models

Course Code: AICL 434

Credit: 3

F.M.: 100

Type: Elective II [Theory +Practical]

Course Description:

This course delves into advanced topics in Natural Language Processing (NLP) and Large Language Models (LLMs), building upon foundational NLP knowledge. Students will explore cutting-edge techniques, including transformer architectures, Retrieval Augmented Generation (RAG), agent handling, prompt engineering, multimodal learning, and advanced applications of LLMs. The course emphasizes both theoretical understanding and practical implementation, with a focus on current research trends.

Course Objectives:

1. Develop understanding of NLP processing fundamentals
2. Understand transformer architectures and their variants for LLMs.
3. Develop and fine-tune large language models for complex NLP tasks.
4. Understand RAG, agent handling, prompt engineering for advanced NLP and LLMs.
5. Work with multimodal data, including text and images.
6. Apply LLMs to advanced applications with the ethical considerations and societal impact

Course Evaluation:

Internal Examination: 50%

Final Examination: 50%

Chapter 1: Fundamentals of Natural Language Processing (NLP) [8 hrs]

- 1.1 Introduction to NLP and its applications
- 1.2 Tokenization, Lemmatization, and Stemming
- 1.3 Part-of-Speech (POS) Tagging and Named Entity Recognition (NER)
- 1.4 Word Embeddings: Tf-idf, Word2Vec, GloVe
- 1.5 Sequence-to-Sequence Language Learning: RNNs
- 1.6 Applications of Encoder-Decoder Models in NLP

Chapter 2: Language Processing via Transformer [10 hrs]

- 2.1 Transformer Model: Attention mechanism, encoder-decoder structure.
- 2.2 Pre-trained Language Models: BERT, GPT
- 2.5 Fine-tuning Pre-trained Models for Downstream Tasks
- 2.6 Training LLM from scratch, Scaling loss
- 2.7 Applications using transformers
- 2.8 Optimization Methods for Generative Models
- 2.9 Evaluation Metrics for Generative Models

Chapter 3: Advanced Transformer for LLM [10 hrs]

- 3.1 Advanced Attention Mechanisms: Sparse attention, linear attention.
- 3.2 Transformer Variants: BERT, RoBERTa, ALBERT, ELECTRA
- 3.3 Retrieval Augmented Generation (RAG)
- 3.4 Agent and multi-agent handling
- 3.5 Transfer learning and Fine-tuning of Large Language Models
- 3.6 Parameter-Efficient Fine-tuning: Adapter methods, prefix-tuning.

Chapter 4: LLM Prompt Engineering and Applications [10 hrs]

- 4.1 Prompt Engineering: Crafting effective prompts for LLMs.
- 4.2 Few-shot Learning and Zero-shot Learning.
- 4.3 Prompt Tuning and Optimization.
- 4.4. Prompting Techniques: Chain-of-Thought, Graph-of-thought etc.
- 4.5 Building Applications with LLMs: Chatbots, content creation, code generation.
- 4.6 Ethical Considerations in NLP: Bias, fairness, and privacy.

Chapter 5: Multimodal Learning with LLMs [7 hrs]

- 5.1 Multimodal LLMs: Combining Text, Images, and Audio
- 5.2 Multimodal Transformers: Visual-linguistic models, speech-text models.
- 5.3 Cross-modal Attention and Fusion Techniques.
- 5.4 Applications of Multimodal LLMs: Image captioning, visual question answering, multimodal dialogue.

Reference Books:

1. Large Language Models: A Deep Dive: Bridging Theory and Practice by Uday Kamath et. al., Springer publication.
2. Hands-On Large Language Models-Jay Alammar and Maarten Grootendorst, O'Reilly Media, Inc.
3. GPT-3, Building Innovative NLP Products Using Large Language Models-Sandra Kublik and Shubham Saboo, O'Reilly Media, Inc.
4. Natural Language Processing with PyTorch Build Intelligent Language Applications Using Deep Learning - Delip Rao and Brian McMahan, O'Reilly Media, Inc.