Applied NLP with Large Language Models

Code: CS31125CS	Course Type: Elective	Semester: I
Evolution Scheme (TA-MSE-ESE): 20-30-100	Total Marks: 150	L-T-P (Credits): 3-0-0 (3)

Pre-Requisites: Python Programming, Machine Learning Fundamentals, Linear Algebra and Probability, Basic Deep Learning.

Course Objective: To develop understanding of NLP fundamentals, transformer-based large language models, and practical AI applications for intelligent language processing systems.

Course Outcomes: At the end of the course, the student will be able to

CO-1	Understand fundamental concepts of NLP, linguistic structures, and traditional language modeling techniques.
CO-2	Apply neural network architectures (RNN, LSTM, Transformers) and decoding strategies for solving NLP tasks.
CO-3	Analyze and implement Large Language Models (LLMs), including pre-training, fine-tuning, and parameter-efficient techniques (LoRA, QLoRA).
CO-4	Evaluate real-world applications of NLP and LLMs in domains such as healthcare, chatbots, and knowledge discovery, along with future trends.

Course Articulation Matrix:

PO	PEO-1	PEO-2	PEO-3	PEO-4	PEO-5
CO-1	2	2	-	-	3
CO-2	3	3	1	-	2
CO-3	3	3	1	-	3
CO-4	2	2	2	2	3

1 - Slightly; 2 - Moderately; 3 – Substantially

Syllabus:

Unit 1: Classical NLP & Distributed Semantics

Introduction to NLP, Understanding linguistic fundamentals: syntax, semantics, morphology, and phonetics. Tokenization, stemming, and lemmatization, Introduction to Statistical Language Models, Feature extraction methods for representing text data, including bag-of-words and TF-IDF, Word Embeddings (Word2Vec, fastText, GloVe), Tokenization Strategies, N-gram models.

Unit 2: The Deep Learning & Transformer Revolution

Introduction to neural networks in the context of NLP, Sequence-to-Sequence Models(Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks), Greedy Decoding, Beam search, Other Decoding Strategies: Nucleus Sampling,

Temperature Sampling, Introduction to Transformers architecture: self and Multi-Head Attention, Positional Encoding and Layer Normalization.

Unit 3: Large language Models

Introduction to LLMs, LLM use cases and tasks, Popular LLM families: GPT/Llama/Gemini/Claude, Application of LLMs (GPT, BERT, etc.,) in NLP tasks, Pre-training large language models (GPT, BERT, BART, and T5), Fine-tuned language models, Computational challenges of training LLMs, Instruction fine-tuning, Parameter efficient fine-tuning(PEFT) technique, Introduction to Transfer Learning in NLP, LLM Optimization & Compression, Advanced Retrieval-Augmented Generation (RAG).

Unit 4: Application areas of NLP and Future trends

NLP in healthcare and biomedical engineering, Case studies on text-based analysis of medical literature, electronic health records, and patient data. Exploration of literature mining and knowledge discovery in biomedicine. Overview of advanced NLP applications in AI, including chatbots, summarization, and question-answering systems, Exploration of emerging trends in NLP and Large Language Models. Discussion on the potential impact of future advancements in AI and biomedical engineering.

Learning Resources:

Textbooks:

1. Pam baker, Generative AI For Dummies, Wiley, 2025
2. Tanmoy Chakraborty, Introduction to Large Language Models Generative AI for Text, Wiley, 2024.

Reference Books:

1. Dan Jurafsky and James H. Martin, Speech and Language Processing, 2nd edition, Pearson Press, 2008.
2. Jacob Eisenstein, Natural Language Processing, First edition, The MIT Press, 2019.