Subject Code	Subject Name (Lab oriented Theory Courses)	Category	L	T	P	C
AD23633	Generative AI	PC	2	0	2	3

Objectives:

1	To provide a fundamental understanding of the transition from discriminative to generative modelling.
2	To explore the architecture and mechanics of Transformers and Large Language Models (LLMs).
3	To master the art and science of Prompt Engineering for optimizing model outputs.
4	To understand the integration of external data through Vector Databases and Retrieval-Augmented Generation (RAG).
5	To equip students with the skills to build, evaluate, and deploy real-world Generative AI applications.

UNIT I	Foundations of Generative AI	6

Concepts of Generative AI – Discriminative vs Generative AI, Statistical Models: GMM, HMM. Key NLP tasks (text classification, sentiment analysis, summarization), Tokenization, Text Representation – Word embeddings. Ethics and Challenges of Gen AI – Bias in AI models.

UNIT II	Core Architectures of LLMs	6

Generative Models: Variational Autoencoders (VAE) - Generative Adversarial Networks (GAN). Introduction to Transformers: – Encoder-decoder architecture – Self-attention mechanism - Text Generation: Greedy Search Decoding, Beam Search Decoding, Sampling methods (Top-k, Nucleus Sampling).

UNIT III	Prompt Engineering	6

Fundamental of prompt engineering- Prompt Elements - Techniques: Zero-shot, one-shot, few-shot and Chain-of-thought (CoT) – Graph Prompt – Automatic Prompt Engineer (APE). Best practices of Prompt Design – Handling ambiguous or complex queries.

UNIT IV	Retrieval-Augmented Generation	6

Vector Databases - Integrating vector databases with LLMs - FAISS – Pinecone – Deep Lake – Parameter Efficient Fine Tuning - Retrieval Augmented Generation (RAG) – Pipeline – Core Components.

UNIT V	Real World Applications using Gen AI	6

Integrating Gen AI into workflows –Automating tasks using AI. Building Custom Apps – Chatbots and Virtual Assistants, Content generation tools (blog writing, marketing copy), Code generation and debugging tools. Evaluating and Optimizing AI outputs – Metrics for evaluating AI performance, Iterative improvement of prompts and models

Contact Hours	:	30

List of Experiments

1	Develop a predictive text system using N-Gram and HMM (Hidden Markov Models) to predict the next word based on previous context. Dataset: Brown Corpus or Wikipedia dataset.
2	Develop a program to perform Bias Audit on pre-trained word embeddings to identify gender or racial stereotypes using the Word Embedding Association Test (WEAT). Dataset: Common Crawl or Google News Embeddings.

3	Implement a Generative Adversarial Network (GAN) to generate synthetic images or tabular data. Dataset: MNIST (for images) or Synthetic Financial Transactions.			
4	Analyze and compare the text generation capabilities of different LLM architectures and access methods (OpenAI API, Hugging Face).			
5	Analyze the impact of Zero-shot, Few-shot, and Chain-of-Thought (CoT) prompting on the reasoning capabilities of Large Language Models (LLMs) and identify patterns in logical fallacies or "hallucinations." Dataset: SQuAD dataset or GSM8K (Math Word Problems) dataset.			
6	Implement a program to retrieve similar products or movies based on high-dimensional vector embeddings using FAISS or Pinecone. Dataset: MovieLens or Amazon Product dataset.			
7	Integrate a Vector Database with an LLM to build a Retrieval-Augmented Generation (RAG) system that answers questions based on external PDF documents. Dataset: LinkedIn Profile or Research Paper PDFs.			
8	Deploy a fine-tuned language model on Hugging Face Spaces or Streamlit for customer support, and evaluate it using ROUGE and BLEU metrics. Dataset: Customer service dataset or Bank FAQ dataset.			
		Contact Hours	**:**	**30**
		Total Contact Hours	**:**	**60**

Course Outcomes: **The students will be able to**	
1	Apply statistical models and text representation techniques to perform foundational NLP tasks and identify bias in AI models. (K3)
2	Apply generative architectures and evaluate transformer-based components to implement text generation using various decoding and sampling strategies.(K4)
3	Analyze the effectiveness of various prompting techniques by deconstructing model reasoning paths to identify logic errors and optimize response accuracy. (K4)
4	Construct a Retrieval-Augmented Generation (RAG) pipeline by integrating Large Language Models with vector databases and parameter-efficient tuning. (K3)
5	Design an end-to-end Generative AI application and evaluate the model performance using standardized metrics. (K6)
Text Book (s):	
1	Tanmoy Chakraborty, "Introduction to Large Language Models – Generative AI for Text", Wiley, 1st Edition, 2024.
2	Thimira Amaratunga, "Understanding Large Language Models: Learning Their Underlying Concepts and Technologies", Apress, 1st Edition, 2023.
Reference Books(s) / Web links:	
1	Valentiana Alto, "Modern Generative AI with ChatGPT and OpenAI Models: Leverage the capabilities of OpenAI's LLM for productivity and innovation with GPT3 and GPT4", Packt Publishing, 1st Edition, 2023
2	Omar Sanseviero, Pedro Cuenca, Apolinario Passos, Jonathan Whitaker, "Hands-On Generative AI with Transformers and Diffusion Models", O Reilly Media, 1st Edition, 2024
3	Michael Lanham, M, "*AI Agents in Action",* Manning Publications., 2025

CO – PO – PSO matrices of course

CO No.	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	PSO3
CO1	3	2	1	-	2	2	-	3	-	-	-	3	1	3
CO2	3	2	2	1	3	-	-	-	-	-	-	3	2	1
CO3	2	3	2	2	3	-	-	-	1	2	-	2	3	1
CO4	2	2	3	2	3	-	-	-	2	-	1	3	3	1
CO5	2	2	3	2	3	1	1	2	3	2	3	3	3	2
Average Mapping	2.4	2.2	2.2	1.4	2.8	0.6	0.2	1	1.2	0.8	0.8	2.8	2.4	1.6

1: Slight (Low)2: Moderate (Medium)3: Substantial (High)
No correlation: "-"