Prompt Engineering and Context Engineering

Code: CS33324CS	Course Type: Core	Semester: III
Evolution Scheme (TA-MSE-ESE): 20-30-100	Total Marks: 150	L-T-P (Credits): 3-0-0 (3)

Pre-Requisites: Programming in Python, Data Structures and Algorithms, Database Management Systems, Basic Machine Learning, Introductory Natural Language Processing / Deep Learning fundamentals, Familiarity with APIs, JSON, and basic software development workflows.

Course Objective:

1. Understand the fundamentals of large language models, prompt engineering, and context management techniques used in generative AI systems.
2. Design and optimize prompts for diverse AI tasks such as text generation, summarization, information extraction, reasoning, and structured output generation. Develop context-aware LLM applications using retrieval, embeddings, vector databases, memory, tool calling, and agentic workflows.
3. Evaluate and deploy LLM-based systems responsibly by addressing performance, robustness, privacy, bias, fairness, and ethical issues.

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

CO-1	Explain the working principles of LLMs, prompt engineering methods, and context engineering concepts for generative AI applications.
CO-2	Design effective prompts and interaction strategies for solving domain-specific tasks using large language models.
CO-3	Build context-aware intelligent applications using retrieval-augmented generation, tool use, memory, and agent-based workflows.
CO-4	Evaluate LLM applications and apply responsible AI practices to ensure reliable, ethical, and efficient deployment.

Course Articulation Matrix:

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

1 - Slightly; 2 - Moderately; 3 – Substantially

Syllabus:

Unit-1: Foundations of LLMs and Prompt Engineering

Evolution of NLP to foundation models and LLMs, Tokens, embeddings, transformers, attention, context window, In-context learning, zero-shot, one-shot, few-shot prompting,Anatomy of prompts: role, instruction, context, examples, constraints, output format, Prompt patterns for summarization, classification, extraction, rewriting, code assistance, Structured outputs, prompt templates, and prompt debugging

Unit-2: Advanced Prompting and Context Engineering

Prompt optimization and iterative refinement, Reasoning-oriented prompting, decomposition, self-critique, verification, Context engineering concepts: static context, retrieved context, memory, tool-generated context, Context window management and information budgeting, Prompt injection, jailbreaks, adversarial instructions, Grounding and citation-aware response design

Unit-3: Retrieval, RAG, Tools, and Agents

Embeddings and semantic similarity, Chunking strategies and metadata-aware indexing, Vector databases and semantic retrieval, Retrieval-Augmented Generation (RAG): architecture, benefits, failure modes, Tool calling / function calling, Memory-enabled systems, Agentic workflows: planning, routing, tool use, multi-step orchestration, Frameworks such as LangChain / LangGraph / DSPy / agent stacks

Unit-4: Evaluation, Responsible AI, and Applications

Evaluation metrics: relevance, correctness, faithfulness, toxicity, latency, cost, Prompt evaluation, retrieval evaluation, end-to-end evaluation, Benchmarking, rubric-based assessment, A/B testing, Bias, fairness, privacy, explainability, legal and societal concerns, Hallucination, data leakage, secure deployment, governance basics, Domain applications: education, healthcare, enterprise assistants, coding copilots, analytics assistants, Mini project / case study presentation

Learning Resources:

Textbooks:

1. Lewis Tunstall, Leandro von Werra, and Thomas Wolf, Natural Language Processing with Transformers, O'Reilly.
2. Sebastian Raschka, Build a Large Language Model (From Scratch), Manning.
3. Hands-On Large Language Models by Jay Alammar and Maarten Grootendorst.

Reference Books:

1. Daniel Jurafsky and James H. Martin, Speech and Language Processing
2. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press.
3. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, MIT Press.
4. Charu C. Aggarwal, Neural Networks and Deep Learning, Springer.