

<b>Semester: 3</b>				
<b>Programme: Data Science</b>				
<b>Course: Data Engineering - I</b>				
<b>Paper code:</b>			<b>Credits: 3</b>	
<b>Hours/week: 4</b>				
<b>Category: Core/MDC/SEC/VAC: Skill</b>				
<b>Theory / Practical / Composite: Practical</b>				
<b>No of Module: 2</b>				
<b>Course Outcome:</b>				
1. <b>Remember</b> the fundamental syntax of R programming, basic data storage types such as dataframes and tibbles, and the core terminology of Generative AI prompt types including zero-shot and few-shot prompting.				
2. <b>Understand</b> the logic behind the tidyverse and graphics packages for data manipulation, the conceptual foundations of Large Language Models (LLMs), and the structural anatomy of a prompt.				
3. <b>Apply</b> R programming functions, conditional statements, and core prompting techniques to perform data extraction, text summarization, and exploratory data analysis (EDA).				
4. <b>Analyze</b> real-life datasets using inbuilt R mathematical and statistical tools and examine the effectiveness of reasoning-enhancement strategies like chain-of-thought and role-based prompting.				
5. <b>Evaluate</b> the accuracy, relevance, and ethical implications of AI-generated responses to iteratively refine prompts and minimize bias in data science applications.				
6. <b>Create</b> integrated, reproducible data science reports and automated workflows using R Markdown and sophisticated prompting strategies to solve complex real-world problems.				
<b>SYLLABUS</b>				
<b>UNIT/Module</b>	<b>CONTENT</b>	<b>HOURS or NUMBE R OF CLASSE S</b>	<b>CO Mappi ng</b>	<b>COGN ITIVE LEVE L</b>
<b>Module 1:</b>	<b>Introduction to R Programming:</b>  Installing and getting started with R studio. Using R as calculator. Reading numeric, character & logical data. Use of inbuilt mathematical & statistical functions. Data storage using dataframe, tibble, matrix, list. Use of tidyverse package, graphics package. Conditional statements. Implicit and explicit loops. User defined functions. Importing data from external files. Analysing real life data using R functions and report writing using R markdown.	26	CO1, CO2, CO3, CO4, CO6	K1,K2, K3, K4, K6

<b>Module 2: Introduction to Prompt Engineering</b>				
	<b>Unit 1: Foundations of Prompt Engineering</b>  Introduction to Generative AI & LLMs, Concept of Prompt Engineering, Types of Prompts (Zero-shot, Few-shot, Chain-of-thought), Anatomy of a Prompt (Instructions, Context, Output Format), Evaluating AI Responses (Accuracy, Relevance, Bias)	6	CO1, CO2, CO5	K1, K2, K5
	<b>Unit 2: Core Prompting Techniques</b>  Zero-shot vs Few-shot prompting in practice, Chain-of-thought prompting & reasoning enhancement, Role prompting & persona-based instructions, Instruction tuning & context setting, Delimiters, formatting, and structured outputs (JSON, tables), Prompting for summarization, translation, and Q&A, Error handling & iterative refinement of prompts	8	CO3, CO4, CO5	K3, K4, K5
	<b>Unit 3: Advanced Prompting Strategies</b>  Self-consistency & ensemble prompting, Multi-turn conversations & memory-aware prompts, Prompting for data extraction and cleaning, Prompting for text classification & labeling tasks, Prompting for knowledge retrieval & augmentation (RAG overview)	6	CO3, CO6	K3, K6
	<b>Unit 4: Prompt Engineering in Data Science Applications</b>  Prompting for exploratory data analysis (EDA) assistance, Prompting for feature engineering ideas, Prompting for visualization & storytelling with data.	4	CO3, CO4	K3, K4
	<b>Unit 5: Ethics, Tools &amp; Project</b>  Ethics, Bias, and Responsible Prompting, Case Study.	2	CO5, CO6	K5, K6
<b>Text Books</b>				
1. <b>LLM Prompt Engineering for Developers</b> , Jason Wei, O'Reilly Media				
2. <b>Prompt Design Patterns</b> , James Phoenix & Mike Taylor, O'Reilly Media.				
<b>Evaluation</b>	Continuous Assessment			

### Course outcomes (COs) and Cognitive Level Mapping

COs	CO Description	Cognitive levels
CO1	<b>Remember</b> the fundamental syntax of R programming, basic data storage types such as dataframes and tibbles, and the core terminology of Generative AI prompt types including zero-shot and few-shot prompting.	K1
CO2	<b>Understand</b> the logic behind the tidyverse and graphics packages for data manipulation, the conceptual foundations of Large Language Models (LLMs), and the structural anatomy of a prompt.	K2
CO3	<b>Apply</b> R programming functions, conditional statements, and core prompting techniques to perform data extraction, text summarization, and exploratory data analysis (EDA).	K3
CO4	<b>Analyze</b> real-life datasets using inbuilt R mathematical and statistical tools and examine the effectiveness of reasoning-enhancement strategies like chain-of-thought and role-based prompting.	K4
CO5	<b>Evaluate</b> the accuracy, relevance, and ethical implications of AI-generated responses to iteratively refine prompts and minimize bias in data science applications.	K5
CO6	<b>Create</b> integrated, reproducible data science reports and automated workflows using R Markdown and sophisticated prompting strategies to solve complex real-world problems.	K6