

Assessment - AI and Generative AI

Total Marks – 15 Marks

Total Time – 30 Minutes

Q.1 What would be the expected output of the code snippet attached below?

```
import numpy as np
from sklearn.cluster import KMeans

X = np.array([[1, 2], [1, 4], [1, 0], [10, 2], [10, 4],
              [10, 0]])

kmeans = KMeans(n_clusters=2)
kmeans.fit(X)
```

Q.2 What is meant by model's 'learning rate'?

We should keep a high learning rate to achieve reliable accuracy – comment on this statement.

Q.3 Differentiate between Fine-Tuning and Prompt Tuning in Large Language Models (LLMs). When would you prefer one over the other?

Q.4 Explain the role of embeddings in Generative AI. How are they used in retrieval-based approaches like RAG?

Q.5 Fine-tuning a generative model typically requires:

- A. Zero labeled data
- B. Domain-specific training data
- C. Internet access during inference
- D. Only pre-trained embeddings

Q.6 What does "Prompt Engineering" primarily involve?

- A. Training new machine learning models from scratch
- B. Designing questions to trick the model
- C. Crafting inputs to guide model behavior and outputs
- D. Engineering hardware for faster AI inference

Q.7 You are provided with a synthetic dataset - 'fraud_dataset.csv' that mimics real-world credit card or digital transaction records. Your goal is to build a Machine Learning model to detect fraudulent transactions.

Background: Fraudulent transactions are rare but high-impact events. This makes fraud detection a classic imbalanced classification problem. The dataset contains 5,000 transactions with 20 features like transaction amount, account balance, location risk, 2FA status, and more. Only ~3% of the data represents fraud cases.

Task Instructions:

As an AI Scientist, your goal is to perform end-to-end analysis and develop a fraud detection model. You must:

1. Understand the dataset using summary statistics and visualizations.
 2. Perform Exploratory Data Analysis (EDA): Compare fraud vs. non-fraud cases.
 3. Preprocess the data.
 4. Train a machine learning model.
 5. Evaluate model performance with confusion matrix and obtain accuracy score.
-