| Roll No. | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|

# PRESIDENCY UNIVERSITY

## BENGALURU

**End - Term Examinations – December 2025**

**Date:** 11 – 12- 2025 **Time:** 01:00pm – 04:00pm

| | | |
|---|---|---|
| **School:** SOCSE | **Program:** B.Tech [CSE(AIML)] | |
| **Course Code:** CAI3405 | **Course Name:** Explainable AI | |
| **Semester**: VII | **Max Marks**: 100 | **Weightage**: 50% |

| **CO - Levels** | **CO1** | **CO2** | **CO3** | **CO4** | **CO5** |
|---|---|---|---|---|---|
| **Marks** | **22** | **22** | **26** | **26** | **NA** |

**Instructions:**

*(i) Read all questions carefully and answer accordingly.*

*(ii) Do not write anything on the question paper other than roll number.*

## Part A

**Answer ALL the Questions. Each question carries 2marks.** **10Q x 2M=20M**

| | | | | |
|---|---|---|---|---|
| **1.** | Define the term "black-box model" in the context of AI and explain why it presents a challenge for trust. | **2 Marks** | **L1** | **CO1** |
| **2.** | Identify two primary reasons for the growing Motivation for Explainable AI (XAI). | **2 Marks** | **L1** | **CO1** |
| **3.** | Differentiate between Global Explanations and Local Explanations. | **2 Marks** | **L1** | **CO2** |
| **4.** | Define Permutation Feature Importance (PFI) and state what it measures. | **2 Marks** | **L1** | **CO2** |
| **5.** | What is the primary purpose of a Saliency Map in explaining image models? | **2 Marks** | **L1** | **CO3** |
| **6.** | Define Attention Visualization in the context of text data explainability. | **2 Marks** | **L1** | **CO3** |
| **7.** | Name two examples of Deep Explainers and state their common goal. | **2 Marks** | **L1** | **CO3** |
| **8.** | Define the term "Explainability by Design". | **2 Marks** | **L1** | **CO4** |
| **9.** | List two major Risks of not including explainability in a high-stakes AI system. | **2 Marks** | **L1** | **CO4** |
| **10.** | What is a Counterfactual Explanation? | **2 Marks** | **L1** | **CO4** |

## Part B

**Answer the Questions.** **Total Marks 80M**

| | | | | | |
|---|---|---|---|---|---|
| **11.** | **a.** | Discuss the critical need for Trust and Accountability in AI systems. Explain how Transparency and Interpretability address this need, providing a clear example of an accountable AI decision. | **10 Marks** | **L2** | **CO1** |
| | **b.** | Analyse the concept of the Trade-off between accuracy and interpretability. Provide scenarios where maximizing accuracy is paramount, versus scenarios where maximizing interpretability is essential. | **10 Marks** | **L4** | **CO1** |
| **Or** | | | | | |
| **12.** | **a.** | Elaborate on the various Goals and Objectives of XAI for different Audience Segmentation (e.g., end-users, developers, regulators). | **10 Marks** | **L2** | **CO1** |
| | **b.** | Analyse the trade-off between interpretable and black-box models based on the required level of Trust and Accountability. Evaluate two contrasting application scenarios—one where model opacity creates unacceptable systemic risk and one where the opacity is negligible—to justify the essential versus unnecessary role of Explainable AI (XAI). | **10 Marks** | **L4** | **CO1** |

| | | | | | |
|---|---|---|---|---|---|
| **13.** | **a.** | Describe the mechanism of Partial Dependency Plots (PDP) and Accumulated Local Effects (ALE) plots. Explain how ALE plots provide an improvement over PDPs. | **10 Marks** | **L2** | **CO2** |
| | **b.** | Given a complex black-box model's predictions on a dataset, apply the methodology of Surrogate Models by detailing the steps to fit a global, simple Decision Tree to these predictions. Interpret how the resulting tree provides an approximation of the black-box model's overall decision logic and its key feature interactions. | **10 Marks** | **L3** | **CO2** |
| **Or** | | | | | |
| **14.** | **a.** | Elaborate on the method of LIME (Local Interpretable Model-agnostic Explanations). Discuss how LIME creates a local, faithful explanation for a single prediction of any black-box model. | **10 Marks** | **L2** | **CO2** |
| | **b.** | Given a regulatory requirement for absolute transparency, apply the characteristics of a Transparent Model (e.g., linear regression) and a Model-Agnostic Technique (e.g., RuleFit). Evaluate the trade-off between the high model fidelity offered by RuleFit versus the guaranteed interpretability of the transparent model when choosing a solution. | **10 Marks** | **L3** | **CO2** |

| | | | | | |
|---|---|---|---|---|---|
| **15.** | **a.** | Describe the methodologies of Saliency Maps and Grad-CAM for explaining image model predictions. Explain the key difference in the information highlighted by each method. | **10 Marks** | **L2** | **CO3** |
| | **b.** | Discuss the challenges of Explaining structured tabular data compared to image or text data. Explain how feature importance methods address these challenges. | **10 Marks** | **L4** | **CO3** |
| **Or** | | | | | |
| **16.** | **a.** | Elaborate on the techniques used for Text data explainability, focusing on Attention Visualization and Interpretable Embeddings. How do these methods provide insight into model reasoning? | **10 Marks** | **L2** | **CO3** |
| | **b.** | Analyze the specific requirements and challenges of Explainability in LLMs and foundation models. Why are traditional XAI methods often insufficient for these large models? | **10 Marks** | **L4** | **CO3** |

| | | | | | |
|---|---|---|---|---|---|
| **17.** | **a.** | Explain the process of integrating XAI into AI pipelines (Explainability by design). Describe the key checkpoints where explainability must be considered throughout the model development lifecycle. | **10 Marks** | **L2** | **CO4** |
| | **b.** | Analyze the challenges of satisfying heterogeneous user needs through User-Centered Design (UCD) for XAI. Evaluate how the user's role (e.g., domain expert vs. layperson) and resulting cognitive load necessitate distinct interface design choices for explanation complexity and delivery. | **10 Marks** | **L4** | **CO4** |
| **Or** | | | | | |
| **18.** | **a.** | **Analyze** the **Industry-specific requirements for XAI** in two different sectors (e.g., finance, healthcare). How do these requirements influence the choice of XAI technique? | **10 Marks** | **L4** | **CO4** |
| | **b.** | **Describe** the necessity and application of **Counterfactual explanations** in sensitive systems like **Job Screening Systems**. Explain how they provide actionable recourse for individuals. | **10 Marks** | **L2** | **CO4** |