RV College of Engineering®
(Autonomous Institution affiliated to VTU)
Bengaluru-560 059

COURSE CODE: AI374TFB	**SEM**: VII
COURSE TITLE: Explainable Artificial Intelligence	
Duration of Paper: 03 Hrs.	

Instructions to Candidates:
1. Answer all questions from Part A
2. Any 5 Full questions from Part B choosing one from each side. (Question No.2 is compulsory)

Part A

SL. No	Questions	Marks	BTL	CO
1.1	Why do we need explanations for AI Models?	2	1	1
1.2	Differentiate interpretability and explainability.	2	1	1
1.3	Model predicts House Price (in ₹ thousands) using: Feature 1: Income (I) Feature 2: Credit Score (C) Model outputs: None: {∅}:100, {I}: 130, {C}: 120, {I,C}:160 Compute the Shapley Values for I and C.	2	2	1
1.4	Identify the hidden rule to transform the given input string into the output. **Input String** **Output** CAT3 D15 DOG4 E20 PEN2 Q10	2	1	1
1.5	What are Saliency Maps? Give their purpose in AI.	2	1	2
1.6	How does Integrated Gradients ensure that the contributions of input features are fairly attributed to the model's predictions?	2	1	2
1.7	In what scenarios would you consider using LIME over other explainability techniques?	2	2	2
1.8	What is Layer Integrated Gradients, and how does it differ from standard Integrated Gradients in terms of visualizing feature importance in neural networks?	2	2	2
1.9	Mention any two characteristics of a good explanation.	2	1	1
1.10	Mention one major advantage of Grad-CAM over pixel-level gradient methods.	2	1	2

Part B

SL. No	Questions	Marks	BTL	CO
2a	i. Define the Premodeling Explainability technique and Comment on the advantages and limitations of this technique. ii. Identify which type of explainability is suitable for the following cases and why? Loan Approval, Medical diagnostic Support, Detecting cracks in ships, Student grade prediction, Predictive maintenance of equipment's	4 + 4	3	1
2b	Give the illustrative examples to prove the following statements. 1. Pure explanations are often unsatisfactory 2. Counterfactual explanation gives the context or alternative. 3. Causal Explanations are Logical but not always true 4. Interpretability depends on the model architecture	4 X 2	3	1
3a	Discuss the concept of Permutation feature importance used in explaining the tabular data by considering a house-price prediction as an illustrative example.	8	3	2
3b	Examine the Python tools used for visualizing local feature attributions in tabular data. In your discussion, analyze the implementation details and technical	8	3	2

	capabilities of Waterfall visualizations and Force plots. Additionally, evaluate how these visualizations integrate with SHAP values and assess their implications for enhancing model interpretability and analyzing feature importance.			
	OR			
4a	Write a note on the process used for explaining the Decision Tree-based models by considering a California housing dataset as an illustrative example.	8	3	2
4b	Given a black-box model that predicts car prices using a Tabular dataset, explain how you would use XAI techniques to identify the most influential features.	8	3	2
5a	Given a trained image classification model that predicts "cat" for an input image, describe how you would use Integrated Gradients to visualize the important pixels contributing to this prediction.	8	3	3
5b	Define XRAI in the context of Explainable AI. What is its primary goal when applied to deep neural networks for visual explanations?	8	2	3
	OR			
6a	Describe how LIME approximates complex models locally with simpler interpretable models.	8	2	3
6b	Given an image classification model that predicts "zebra," describe how you would apply Grad-CAM to visualize the most relevant image regions influencing the prediction.	8	3	3
7a	Suppose you are developing a text classifier for email spam detection. Describe how you would use tokenization and embeddings as part of your preprocessing and feature extraction pipeline.	8	3	3
7b	Define Layer Integrated Gradients (Layer IG) and explain its primary purpose in neural network interpretability.	8	2	2
	OR			
8a	Suppose a question-answering model incorrectly answers a query. Describe how analyzing its attention weights can help identify whether the error was due to poor focus or contextual misunderstanding.	8	3	3
8b	Explain how LRP distributes a model's output score backward through the network to individual input features.	8	2	2
9a	Explain the need for evaluation metrics in assessing the quality of explanations produced by AI models.	8	2	4
9b	Illustrate with an example how explainability metrics can be used to compare the performance of two XAI models in terms of user understanding and satisfaction.	8	3	4
	OR			
10a	Summarize the importance of integrating causality and fairness in the design of explainable AI models.	8	2	4
10b	Apply fairness-aware explanation techniques to a real-world case (e.g., credit scoring or healthcare diagnostics) and discuss how causal relationships can improve transparency.	8	3	4

Name of the Scrutinizer Name of the BoE Chairperson

Signature of Scrutinizer Name of the BoE Chairperson