RV College of Engineering®
Mysore Road, RV Vidyaniketan Post,
Bengaluru - 560059 Karnataka, India

Go. change the world

Department of Artificial Intelligence and Machine Learning

Date: 16/12/2025	Test – II	Max. Marks: 50 + 10
Semester: VII	UG	Duration: 2 hours
Course Title: Explainable Artificial Intelligence		Course Code: AI374TFB

Part – A

Q. No.	Question	Marks	BT	CO
1.1	IG is better suited to low-contrast images or images acquired in non-natural environments, such as X-rays. Give a reason.	2	L1	CO1
1.2	Integrated Gradients determine the salient inputs by gradually varying the network input from a _____ to the original input and _____ the gradients along the path.	2	L1	CO1
1.3	How XRAI is different from IG?	2	L2	CO1
1.4	Why Grad-CAM is computationally efficient?	2	L2	CO1
1.5	LIME is called a perturbation-based explainability technique. Mention the reason.	2	L1	CO2

Part – B

Q. No.	Question	Marks	BT	CO
2.	Discuss the following with respect to Integrated Gradients (IGs). a. Significance of Linear Interpolation between two points x and y b. Impact of baseline with noise. c. Using IG to determine whether a CNN is focusing on the main object or irrelevant background patterns in an image.	3+3+4	L3	CO3
3.	Discuss the following with respect to XRAI. a. Purpose and working of Felzenszwalb's algorithm. b. In a medical imaging task (e.g., Tumor detection), illustrate how XRAI can be applied to justify clinical decision support outcomes.	5 + 5	L3	CO3
4.	Discuss the following with respect to Grad-CAM. a. Steps to derive $$L^{c}_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_{k} \alpha^{c}_{k} A^{k}\right)$$ b. You are designing an explainability tool for production deployment, where retraining the model is not possible. Which method, CAM or Grad-CAM, is more appropriate, and why?	5 + 5	L3	CO2
5	Answer the following with respect to LIME. a. Pros and Cons of LIME. b. Process of identifying the important regions using Superpixels.	5 + 5	L2	CO2
6	a. Discuss how the processing method differs between RNNs and transformers. b. A toxic comment detection model is deployed to classify online user comments as *toxic* or *non-toxic*. Apply LIME to determine how local changes affect the probability of toxicity.	5 + 5	L3	CO3