Semester: VII						
EXPLAINABLE ARTIFICIAL INTELLIGENCE **Category: Professional Core Elective – IV (Group F)** **(Theory)**						
Course Code	**:**	**AI374TFB**		**CIE**	**:**	**100 Marks**
Credits: L: T: P	**:**	**3:0:0**		**SEE**	**:**	**100 Marks**
Total Hours	**:**	**45L**		**SEE Duration**	**:**	**3.00 Hours**

UNIT-I	**8 Hrs**
Introduction Explainable AI: Introduction to Explainable AI, How Has Explainability Been Used?- How LinkedIn Uses Explainable AI, PwC Uses Explainable AI for Auto Insurance Claims, Accenture Labs Explains Loan Decisions, DARPA Uses Explainable AI to Build "Third-Wave AI" **An Overview of Explainability:** What Are Explanations? What Are Explanations? Interpretability and Explainability, Explainability Consumers, Types of Explanations, Themes Throughout Explainability	
UNIT-II	**9 Hrs**
Explainability for Tabular Data: Permutation Feature Importance, Shapley Values, Explaining Tree-Based Models, Partial Dependence Plots and Related Plots	
UNIT-III	**9 Hrs**
Explainability for Image Data: Integrated Gradients (IG), XRAI -How XRAI Works, Implementing, XRAI, Grad-CAM, LIME, Guided Back propagation and Guided Grad-CAM	
UNIT-IV	**9 Hrs**
Explainability for Text Data: Overview of Building Models with Text, LIME, Gradient x Input, Layer Integrated Gradients, Layer-Wise Relevance Propagation (LRP).	
UNIT-V	**10 Hrs**
Research Directions and Future Trends: Evaluation Metrics for Explanations, Human Trust and Cognitive Aspects in XAI, Integration with Causality and Fairness, Current Research Trends and Open Challenges, Future of Explainable and Responsible AI **Applications and Case Studies in XAI:** Explainability in Healthcare, Human-in-the-loop AI, Case Studies and Toolkits	

Course Outcomes: After completing the course, the students will be able to:-	
CO1	Understand the fundamental concepts and need for explainable artificial intelligence, including its role in ethical, legal, and social contexts.
CO2	Apply various XAI methods such as SHAP, LIME, Grad-CAM, and model-specific tools to interpret and explain complex machine learning models.
CO3	Evaluate and compare the performance of different explainability techniques across domains like healthcare, cybersecurity, and autonomous systems.
CO4	Design user-centric, trustworthy AI systems that incorporate transparency and interpretability principles, aligning with sustainable and ethical development goals.

Reference Books	
1.	Explainable AI for Practitioners,Michael Munn & David Pitman Foreword by Ankur Taly ISBN: 978-1-098-11913-3 , 2022 edition
2.	Explainable AI: Foundations, Methodologies, and Applications ,Mayuri Mehta, Vasile Palade, Indranath Chatterjee Intelligent Systems Reference Library, ISBN 978-3-031-12806-6, Volume 232, 2019 edition
3.	Responsible Artificial Intelligence, Virginia Dignum, Artificial Intelligence: Foundations, Theory, and Algorithms, ISBN 978-3-030-30370-9, springer 2019
4.	Explainable Artificial Intelligence: A Practical Guide, Parikshit Narendra Mahalle & Yashwant Sudhakar Ingle, ISBN 978-87-7004-715-9 edition 2024 River Publisher
5.	Christoph Molnar “Interpretable Machine Learning”. Lulu, 1st Edition, 2019.

RUBRIC FOR THE CONTINUOUS INTERNAL EVALUATION (THEORY)		
#	**COMPONENTS**	**MARKS**
1.	**QUIZZES:** Quizzes will be conducted in online/offline mode. TWO QUIZZES will be conducted & Each Quiz will be evaluated for 10 Marks. **THE SUM OF TWO QUIZZES WILL BE THE FINAL QUIZ MARKS.**	**20**
2.	**TESTS:** Students will be evaluated in test, descriptive questions with different complexity levels (Revised Bloom’s Taxonomy Levels: Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating). TWO tests will be conducted. Each test will be evaluated for 50 Marks, adding upto 100 Marks. **FINAL TEST MARKS WILL BE REDUCED TO 40 MARKS.**	**40**
3.	**EXPERIENTIAL LEARNING:** Students will be evaluated for their creativity and practical implementation of the problem. Case study-based teaching learning (10), Program specific requirements (10), Video based seminar/presentation/demonstration (20) **ADDING UPTO 40 MARKS**.	**40**
	MAXIMUM MARKS FOR THE CIE THEORY	**100**

RUBRIC FOR SEMESTER END EXAMINATION (THEORY)		
Q. NO.	**CONTENTS**	**MARKS**
	PART A	
1	Objective type questions covering entire syllabus	20
	PART B (Maximum of TWO Sub-divisions only)	
2	Unit 1 : (Compulsory)	16
3 & 4	Unit 2 : Question 3 or 4	16
5 & 6	Unit 3 : Question 5 or 6	16
7 & 8	Unit 4 : Question 7 or 8	16
9 & 10	Unit 5: Question 9 or 10	16
	TOTAL	**100**