WQF7009 Explainable AI

Alternative Assessment – Marking Rubric

Course Learning Outcomes (CLOs)

CLO1: Categorize the concepts of Explainable AI and the current techniques for generating explanations from black-box ML models.

CLO2: Evaluate XAI methods through critical comparison of model types (local vs global, post-hoc vs intrinsic, model-specific vs agnostic) to propose innovative, XAI-driven solutions for emerging domain problems.

Rubric Structure

Part	Task	Marks	CLO Alignment
Part 1	Task 1: XAI-Driven Solution Proposal	20%	CLO2
Part 2	Task 2: Model Building & Evaluation	5%	CLO1
	Task 3: XAI Method Selection, Categorisation & Application	10%	CLO1
	Task 4: Knowledge Assessment	5%	CLO1
TOTAL		**40%**	

PART 1 — TASK 1: XAI-Driven Solution Proposal (20 marks)

(Why XAI matters, stakeholders, strategy, impacts)

CLO2

Performance Level	Descriptor	Marks
Excellent (A)	Demonstrates *comprehensive, domain-deep* understanding. Clearly explains **why XAI is necessary**, identifies stakeholders accurately, and proposes an **innovative and well-justified XAI strategy**. Impact analysis (trust, risk, fairness, compliance, decision-making) is **insightful, specific, and realistic**. Includes diagrams, mock dashboards, or workflow illustrations.	**17–20**
Good (B)	Shows strong understanding of domain needs. Stakeholders correctly identified. Provides a well-reasoned XAI strategy with convincing justification. Covers most impacts clearly but lacks minor depth or specificity.	**13–16**
Satisfactory (C)	Basic explanation of why XAI matters. Stakeholders partially identified. XAI strategy is relevant but **simple**, with limited justification. Impacts described in general terms.	**9–12**
Weak (D)	Minimal explanation of domain relevance. Stakeholders unclear. XAI strategy is generic or mismatched. Impacts superficial or inaccurate.	**5–8**
Fail (F)	No meaningful justification. Strategy missing or incorrect. No evidence of understanding XAI in domain context.	**0–4**

PART 2 — TASK 2: Model Building & Evaluation (5 marks)

(Dataset, model, metrics)
CLO1

Performance Level	Descriptor	Marks
Excellent (A)	Dataset well-described; data features understood. Correct model selection with appropriate training. Evaluation metrics are complete and interpreted (e.g., accuracy, AUC, F1, RMSE).	**5**
Good (B)	Clear dataset summary; model trained correctly. Metrics provided but interpretation limited.	**4**
Satisfactory (C)	Basic dataset/model description. Only simple metrics reported.	**3**
Weak (D)	Incomplete model or dataset description. Incorrect or missing metrics.	**2**
Fail (F)	No model built, no metrics, or misleading results.	**0–1**

PART 2 — TASK 3: XAI Method Selection, Categorisation & Application (10 marks)

(Feature Importance + Counterfactuals, categorisation, justification, explanation output)
CLO1

Component	Excellent (A)	Good (B)	Satisfactory (C)	Weak (D)	Fail (F)
A. Explainability Challenges of Model	Deep understanding, clearly linked to architecture (e.g., CNN spatial hierarchy, RF feature interactions, Transformers' attention complexity).	Good but less detailed.	Basic description only.	Partial or incorrect.	Missing.
B. Categorisation of Two XAI Methods	Accurate classification across **all axes**: Local/Global, Post-hoc/Intrinsic, Model-specific/Agnostic.	Minor errors.	Partially correct.	Many errors.	Missing.
C. Justification of Method Suitability	Strong argument connecting method → model → data modality. Justification is technical and domain-appropriate.	Mostly correct justification.	Some relevance but superficial.	Weak reasoning.	No justificati

Component	Excellent (A)	Good (B)	Satisfactory (C)	Weak (D)	Fail (F)
D. Application of Feature Importance + Counterfactuals	Correct implementation with clear, high-quality visual outputs (e.g., SHAP plots, counterfactual tables). Interpretation is insightful.	Good implementation and visuals but less detailed interpretation.	Outputs present but minimal analysis.	Incorrect or unclear outputs.	No applicati
E. Written Explanation Quality	Clear, accurate, and shows strong understanding of model behavior via XAI.	Mostly clear.	Adequate.	Weak.	Missing.

Total = **10 marks**

PART 2 — TASK 4: Knowledge Assessment (5 marks)

(Concept-based vs Counterfactual; pros/cons of interpretable complex models)
CL01

Performance Level	Descriptor	Marks
Excellent (A)	Clear, accurate differentiation of **concept-based vs counterfactual** explanations using examples from their own model. Deep analysis of interpretable complex models (e.g., GAMs, GNNExplainer) with correct advantages and limitations.	**5**
Good (B)	Correct differentiation with example. Good coverage of advantages/disadvantages.	**4**
Satisfactory (C)	Basic differentiation; examples acceptable. Some advantages/disadvantages mentioned.	**3**
Weak (D)	Vague or inaccurate explanation. Examples missing.	**2**
Fail (F)	Incorrect or missing answers.	**0–1**