INTERNATIONAL QUALIFICATIONS

AND ASSESSMENT CENTRE (IQAC)

«Approved»

2025 y. «__» _____

Programme	Level 7 Diploma in Artificial Intelligence		
Unit Number/ Unit Title	UNIT 2 EXPLAINABLE AI AND CAUSAL INFERENCE		
Cohort Code:	L07EAIC-U2		
Unit Level	7		
Total GLH	Total qualification time 200/ Total Guided learning hours 90/ Self-guided learning hours 110		
Credits	20 CATS/ 10 ECTS		
Lecturer			
Start Date		**End Date**	

Unit Aims	This unit equips learners with the knowledge and skills to address the black-box nature of modern AI systems by exploring methods for explainability and transparency. It also introduces causal inference as a critical tool for understanding cause-effect relationships in data beyond mere correlation. Learners will analyse techniques for model interpretability, apply frameworks like SHAP and LIME, and integrate causal reasoning into AI models to promote accountability, fairness, and trust in AI systems.
Differentiation Strategies *(e.g. planned activities or support for individual learners according to their needs)*	The total number of students to be in the lesson is approximately 20. This is a multicultural group of students predominantly between the ages of 24 – 45, with numerous ethnic, gender, and creed background. These are UK academic level 5 students; hence it is assumed that they have practical, theoretical, or technological knowledge and understanding of a subject or field of work to find ways forward in broadly defined, complex contexts. These students must be able to generate information, evaluate, synthesise the use information from a variety of sources. Various approaches to addressing the various identified students needs will be adopted throughout the lesson. Such will include:-

	1. Progressive tasks 2. Digital resources 3. Verbal support 4. Variable outcomes 5. Collaborative learning 6. Ongoing assessment 7. Flexible-pace learning
Equality & Diversity	Variety of teaching techniques will be employed to ensure that the needs of each individual learner are met.
Safeguarding & Prevent	Safeguarding policies and the Prevent duty are strictly observed to ensure the safety, well-being, and inclusivity of all students and staff.
Health & Safety	SIRM H&S policies will be maintained.
Learning Resources	**Teaching and Learning Materials** • Molnar, C. (2022). Interpretable Machine Learning. Leanpub. • Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal Inference in Statistics: A Primer. Wiley. • Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. ACM KDD. • Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. NeurIPS. • Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. arXiv preprint.

Learning Outcome	Assessment Criteria
LO1. 1. Analyse the importance and challenges of explainability in AI systems.	1.1 Discuss ethical and practical reasons for AI interpretability. 1.2 Evaluate challenges in explaining complex AI models. (Written Report)
LO2. 2. Apply methods and tools for interpreting machine learning models.	2.1 Use tools such as SHAP, LIME, or Partial Dependence Plots. 2.2 Visualise and interpret model predictions effectively. (Practical Lab Work)
LO3. 3. Understand fundamental principles of causal inference in AI.	3.1 Distinguish between correlation and causation. 3.2 Explain the role of counterfactual reasoning and structural causal models. (Case Study Analysis)
LO4. 4. Design AI models incorporating causal reasoning frameworks.	4.1 Build models using tools such as DoWhy or CausalNex. 4.2 Evaluate performance and robustness of causal models. (Project-Based Assignment)
LO5. 5. Critically assess the implications of explainable AI in real-world applications.	5.1 Discuss XAI implications in domains such as healthcare, law, or finance. 5.2 Recommend strategies for improving stakeholder trust in AI systems. (Presentation)

No	Learning Outcome / Topic	Learning and Teaching Activities	Which assessment criteria does the session relate to?	Day/month/year/ signature
1.	**Introduction to Explainable AI (XAI)**	**Introduction to Explainable AI (XAI)** Ethical, legal, and practical motivations (GDPR, AI Act)	LO1: Importance & Challenges of Explainability	
2.	**The Black Box Problem**	**The Black Box Problem** Case studies: Biased hiring algorithms, medical misdiagnoses	LO1: Importance & Challenges of Explainability	
3.	**Trade-offs: Accuracy vs. Interpretability**	**Trade-offs: Accuracy vs. Interpretability** Linear models vs. deep learning, complexity challenges	LO1: Importance & Challenges of Explainability	
4.	**Human-Centered XAI**	**Human-Centered XAI** User trust, cognitive load, actionable explanations	LO1: Importance & Challenges of Explainability	
5.	**Regulatory Landscape**	**Regulatory Landscape** EU's "Right to Explanation," FDA guidelines for AI in healthcare	LO1: Importance & Challenges of Explainability	
6.	**Feature Importance Techniques**	**Feature Importance Techniques** Permutation importance, Gini importance (Random Forests	LO2: Model Interpretation Methods	
7.	**Local Interpretability with LIME**	**Local Interpretability with LIME** How LIME works, pros/cons, Python implementation	LO2: Model Interpretation Methods	
8.	Half-Term Exam	- Review of LO1 topics - Practice questions and mock assessment - **Half-term assessment** based on LO1 (theory)	LO1 LO2	
9.	**SHAP (SHapley Additive Explanations)**	**SHAP (SHapley Additive Explanations)** Game theory basis, global vs. local explanations	LO2: Model Interpretation Methods	

10.	**Partial Dependence Plots (PDPs)**	**Partial Dependence Plots (PDPs)** Visualizing feature interactions, ICE plots	LO2: Model Interpretation Methods	
11.	**Counterfactual Explanations**	**Counterfactual Explanations** "What-if" scenarios (e.g., "Why was my loan denied?")	LO2: Model Interpretation Methods	
12.	**Correlation vs. Causation**	**Correlation vs. Causation** Simpson's paradox, confounding variables	LO3: Causal Inference Fundamentals	
13.	**Causal Graphs & DAGs**	**Causal Graphs & DAGs** Directed Acyclic Graphs (DAGs), backdoor paths	LO3: Causal Inference Fundamentals	
14.	Final Exam Preparation & Review	- Comprehensive review of all learning outcomes - Practice questions and revision of key topics		
15.	Final Exam	- **Final-term assessment** covering all learning outcomes (theory and practical elements)		
16.	Feedback & Reflection	- Review of final exam - Individual feedback on performance - Reflective discussion on key learning points		
17.	**Structural Causal Models (SCMs)**	**Structural Causal Models (SCMs)** Potential outcomes framework (Rubin Causal Model)	LO3: Causal Inference Fundamentals	
18.	**Counterfactual Reasoning**	**Counterfactual Reasoning** Neyman-Rubin model, "but-for" analysis	LO3: Causal Inference Fundamentals	
19.	**Instrumental Variables (IVs)**	**Instrumental Variables (IVs)** Natural experiments, validity checks	LO3: Causal Inference Fundamentals	
20.	**Causal Discovery Algorithms** PC	**Causal Discovery Algorithms** PC algorithm, LiNGAM for inferring causality from data	LO4: Causal AI Modelling	
21.	**DoWhy Library**	**DoWhy Library** Causal effect estimation, refutation tests	LO4: Causal AI Modelling	

22.	**CausalNex & Bayesian Networks**	**CausalNex & Bayesian Networks** Probabilistic causal reasoning, sensitivity analysis	LO4: Causal AI Modelling	
23.	Half-Term Exam	**Capstone Project** Build and explain a causal model for a societal problem		
24.	**Causal Reinforcement Learning Evaluating Causal Models**	**Causal Reinforcement Learning** Off-policy evaluation, avoiding spurious rewards **Evaluating Causal Models** ATE (Average Treatment Effect), robustness checks	LO4: Causal AI Modelling	
25.	**XAI in Healthcare**	**XAI in Healthcare** Explainable diagnosis models (e.g., IBM Watson)	LO5: Real-World XAI Applications	
26.	**XAI in Finance**	**XAI in Finance** Credit scoring, fraud detection (e.g., FICO Score)	LO5: Real-World XAI Applications	
27.	**XAI in Criminal Justice**	**XAI in Criminal Justice** Risk assessment tools (e.g., COMPAS controversy)	LO5: Real-World XAI Applications	
28.	**Stakeholder Communication**	**Stakeholder Communication** Tailoring explanations for clinicians vs. policymakers	LO5: Real-World XAI Applications	
29.	Final Exam Preparation & Review	LO1, LO2, LO3, LO4, LO5	LO1, LO2, LO3, LO4, LO5	
30.	Final Exam		LO1, LO2, LO3, LO4, LO5	