# Human Evaluation of Explainability Methods for Regression Tasks on a Tabular Datasets

Ngozi Nneke
*Department of Computing and Mathematics*
*Manchester Metropolitan University*
Manchester, United Kingdom
ngozi.h.nneke@stu.mmu.ac.uk

Luciano Gerber
*Department of Computing and Mathematics*
*Manchester Metropolitan University*
Manchester, United Kingdom
luciano.gerber@mmu.ac.uk

Huw Lloyd
*Department of Computing and Mathematics*
*Manchester Metropolitan University*
Manchester, United Kingdom
huw.lloyd@mmu.ac.uk

Keeley Crockett
*Department of Computing and Mathematics*
*Manchester Metropolitan University*
Manchester, United Kingdom
keeley.crockett@mmu.ac.uk

Artificial Intelligence (AI) systems are increasingly pervasive, spanning from trivial to non-trivial domains and catalysing decision-making processes. Despite their prevalence, AI systems' decision-making processes preceding output tend to be inaccessible to most stakeholders. Consequently, stakeholders and governments including the UK government, have tightened their policies regarding usage of Artificial Intelligence systems.

In response to these concerns, several explainability methods have been explored, such as interpretable models and post-hoc explanations. Interpretable models are well known for high explainability but lower predictive performance, whilst post hoc explanations models are sophisticated black-boxes that lack interpretability. Our research aims to bridge the gap by exploring a sweet spot somewhere in between, relying on more sophisticated models for regression that generate model comprehensible prediction patterns to increase explainability.

However, most research in this area has focused on building models targeting classification tasks; some explanations do not consider humans in the loop and most importantly, lack empirical evidence of their efficacy. Therefore, this study aims to fill a gap by developing explainable regression models using tabular datasets, which will be evaluated objectively and subjectively. Objective metrics from the literature namely, monotonicity, simulatability, smoothness, chunking and compositionality, will be utilised to assess the model. Also, the explanations generated will be evaluated subjectively with lay humans in a human-centred evaluation to justify the degree and quality of the explanations created.

In conclusion, this study will contribute to literature and practice by developing explainable regression models and different evaluation approaches for explainability.

**Keywords** --- Explainability, Interpretability, Explainable Artificial Intelligence, Artificial Intelligence, Machine Learning, Regression, XAI, Human-centred Evaluation, Tabular Datasets