Structured PhD Position in Machine Learning (4-years full-time)

Project title: Developing and Evaluating Interpretable Approaches for Human-Centered Machine Learning.

Project supervisor: Dr. Eoin Delaney (Trinity College Dublin).

Project locations: Discipline of Statistics and Information Systems, School of Computer Science and Statistics, Trinity College Dublin.

Application deadline: 12th May 2025

Start date: Anticipated start date is September 2025

PhD structure: The funding for the project includes a tax-free stipend of €25,000 per annum. In addition to stipend, EU fees will be covered for four years.

PhD topic: Approaches in interpretable machine learning offer promise in understanding the predictions of opaque models that are widely deployed in high-stakes decision making scenarios. Of particular interest to this project are example-based explanation methods that use individual data points or examples to provide insights into the decision making process of complicated models. This family of explanations has both rich statistical and psychological grounding and the project will look to investigate the utility and the robustness of different explanation strategies. A large focus of this project will be on leveraging novel and interpretable approaches in applied domains such as algorithmic fairness and time series prediction for sustainable practices. As explanations are ultimately for end users, it is crucial to consider how they can impact trust in automated decision-making, creating a need for human-centred and psychologically informed methods that could encourage responsible behaviours. Designing such approaches for opaque generative models, such as Large Language Models (LLMs), is critical and is also of growing interest to industry.

The Institution: The School of Computer Science and Statistics at Trinity College Dublin is a collegiate, friendly, and research-intensive centre for academic study and research excellence. The School has been ranked 1st in Ireland, top 25 in Europe, and top 100 Worldwide (QS Subject Rankings 2018, 2019, 2020, 2021, 2023).

Requirements: Applicants should have (or expect to attain prior to project start) at least a 2.1 honours degree or equivalent in the areas of computer science, mathematics, applied mathematics or statistics. Applicants must demonstrate proficiency in machine learning or statistical modelling and have some experience with computing through Python, R or C. Demonstration of open source project work (e.g., GitHub repositories), and familiarity with machine learning and deep learning frameworks (scikit-learn, PyTorch, LLM API's) is a plus. Applicants should demonstrate an interest in human-centered machine learning, interpretability and applied machine learning. Applicants for whom English is a second language will be required to demonstrate their competence in the English language in line with Trinity College Dublin requirements as appropriate.

Application: Applicants should email Dr. Eoin Delaney (eodelane@tcd.ie) to apply. The application should include a comprehensive CV (2-pages max), academic transcripts of the degree/ degrees, and a short cover letter/statement of purpose (2-pages max) indicating how their skills align with the project and their motivation for applying. Please submit these documents as a single pdf. Please include "PhD Application (Interpretable Machine Learning)" followed by your name in the subject line. The application CV should, at minimum, include the applicant's name, educational institution, qualification stating overall grade/percentage (predicted grades are acceptable for those still studying) and contact details of two academic referees. Informal queries can be made to: eodelane@tcd.ie. Please include "PhD Query (Interpretable Machine Learning)" followed by your name in the subject line.