VERIFAI: A Scalable Verification Framework for Transparent AI Models

Supervisor Information

Lead Supervisor: Dr. Alex Chan, Newcastle University, School of Engineering – Electrical and Electronics Engineering
Email: alex.chan@newcastle.ac.uk

Co-supervisor: Dr. Ekin Can Erkus, Newcastle University, School of Engineering – Electrical and Electronics Engineering
Email: ekin.erkus@newcastle.ac.uk

Executive Summary

This project investigates how high-performance computing (HPC) supports verification and explainability of machine learning (ML) models in complex decision-making systems. The student will explore methods on analysing how these ML models arrive at specific outcomes, and develop prototype tools to extract explainable representations of these decision processes. Here, HPC enables efficient state exploration involving combined large inputs and model behaviour, which the project will evaluate to improve transparency and trust in AI systems. The outcome includes a framework demonstrating how scalable computing supports verification and analysis of ML models, with potential applications in areas like finance and other data-intensive domains.

Project Description

Relevant Context and Overarching Goals

ML models are being increasingly deployed in high-stakes domains such as finance, healthcare, and critical infrastructure. Therefore, ensuring that these models are trustworthy, verifiable, and explainable is critical for reliable decision-making. The goal of this project is to develop prototype tools and methods that allow users to systematically analyse the ML model's decisions, understand how outputs are produced, and verify that these models behave correctly across a wide range of scenarios. This internship will give the student practical experience in AI verification, explainability, and research software development, while also contributing to a scalable verification framework.

Computational Aspects

The project involves systematic evaluation of ML models to identify their decision pathways and potential failure points, where the student will develop crucial software tools to inspect model behaviour, implement verification experiments, and process datasets in a reproducible workflow. This work requires algorithm design, data manipulation, and prototype software development, while providing hands-on experience in research software engineering and scalable AI analysis.

HPC Justification

Verifying realistic ML models can involve millions of input combinations, which is computationally intensive. For example, performing these analyses sequentially on a standard workstation would be prohibitively slow or even impossible. HPC instead enables parallel evaluation of models, where many scenarios can be explored simultaneously and scale current analyses to larger, yet high-dimensional, models. Therefore, using HPC with this framework will not only demonstrate AI models that are verifiable and explainable, but also handle real-world AI systems.

Planned Outcomes from Internship

By the end of the internship, the student will have developed a framework to verify ML model decisions and implemented methods to extract explainable representations of model behaviour. They will demonstrate scalable evaluation of AI models using HPC-enabled analysis, while gaining hands-on experience in research software engineering, AI verification, and data analysis. The student will also produce a documented prototype framework with example analyses, which will provide a foundation for further research or even potential application in industry.