Supervision team:

Supervisor: Dr Rehmat Ullah, Email: rehmat.ullah@newcastle.ac.uk, www.rehmatkhan.com

Edge AI: Building Efficient, Trustworthy and Distributed Intelligence

Edge artificial intelligence (Edge AI) enables deploying AI algorithms and models directly on edge devices. However, AI workloads demand high performance processing, large scale data handling, and specialized hardware accelerators, which are often limited and costly. This project explores the challenges of deploying AI at the edge within the context of federated learning (FL). Topics of interest include, but are not limited to, the following:

Communication efficient learning protocols: FL, especially split FL, involves additional transfer of intermediate activations and gradients between devices and the server, which may lead to increased completion time, training loss, energy consumption and reliability. To tackle communication overhead in FL and SFL, we will focus on designing protocols that minimize data exchange between devices and servers such as gradient compression, asynchronous training, reduced synchronization frequency, semantic communication, and design of new application and transport layer protocols.

Data management challenges: Data management across the edge cloud continuum means tracking the origin, ownership, and transformations of data: where it comes from (source), where it is stored, who controls access and usage, what changes are made (e.g., labelling), and how it is used (e.g., which models train on it, which versions, for what purpose).

Middleware solutions: In this topic, we will explore middleware solutions and architectures that support efficient, secure, and scalable machine learning operations (MLOps) across resource-constrained environments for Edge AI.

Ethical, and responsible FL for healthcare: In this topic, we aim to assess both real and perceived ethical concerns, ensuring AI in healthcare is ethical, accountable, and socially beneficial. We will focus on developing frameworks such as federated unlearning and explainable AI to ensure fairness, accountability, and trustworthiness in clinical settings.

Robust aggregation techniques for safeguarding security and privacy in FL: In this topic, our goal is to design and implement robust aggregation techniques that show robustness against attacks and maintain model integrity under a variety of adversarial attacks (e.g., data poisoning, label flipping, and DDoS).

Supervision environment

Student will be supervised by researchers at the EPSRC National Edge AI Hub, UK. The Hub has exclusive research network of 75+ industry partners and 12 leading UK universities and provide access to state-of-the-art technologies and resources, enabling students to test and implement their solutions in high-impact, real-world scenarios.

Applicant skills/background

The ideal candidate should have strong programming skills, particularly in Python, and experience with machine learning algorithms suited for cloud-edge, mobile, or IoT environments.

References

1. D.Wu, R.Ullah,P.Rodgers,P.Kilpatrick,I.SpenceandB.Varghese,"EcoFed: Efficient Communication for DNN Partitioning-Based Federated Learning," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 35, no. 3, pp. 377- 390, March 2024
2. R.Ullah,D.Wu,P.Harvey,P.Kilpatrick,I.SpenceandB.Varghese,"FedFly: Toward Migration in Edge-Based Distributed Federated Learning," in *IEEE Communications Magazine,* vol. 60, no. 11, pp. 42-48, November 2022
3. D.Wu, R.Ullah,P.Harvey,P.Kilpatrick,I.SpenceandB.Varghese,"FedAdapt: Adaptive Offloading for IoT Devices in Federated Learning," in *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 20889-20901, 1 Nov.1, 2022