National Edge AI Hub

UK's National Edge AI Hub

Rajiv Ranjan, FIEEE, MAE, FAAIA

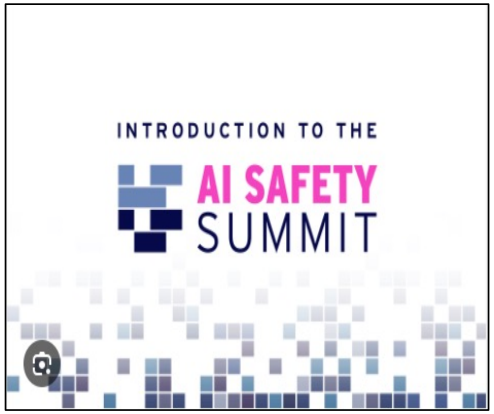
Professor in Computing Science and Internet of Things
School of Computing, Newcastle University

Date: 5th March 2025
Location: One Moorgate Place, London

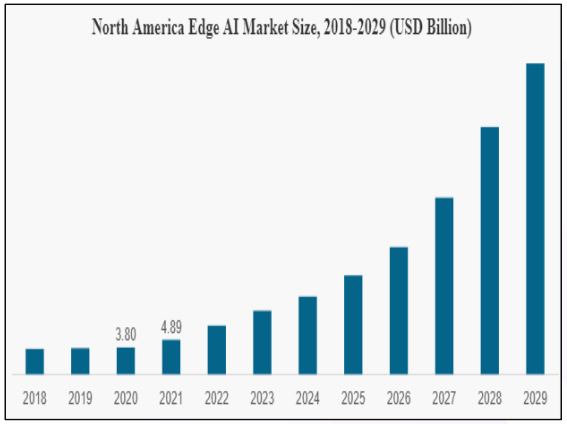
Problem statement (socio-economic)

Within the UKRI funding portfolio, ours is the only Centre of Excellence dedicated to overcoming the key <u>Socio-Economic Question</u>: How UK companies can use edgeAI in a safe and secure way?

Lack of critical workforce having technical skills in edgeAI [UK Government (2022)]



Source: UK Government



Source: Fortune Business Insights

The Emergence of Edge AI
A Game changer for Industries (Gartner 2023)

Edge AI will be the <u>nucleus</u> of AI Innovation over the next 10 years (Gartner 2023)

>55 Industry Partners

Supported by 55 organisations in 7 sectors, including 20 new partners after submission.

National Edge AI Hub

Hardware

Edge /IoT

Innovation Agencies

Healthcare

Cyber Security

Applied AI

Professional Services

Problem motivation (Why Edge AI?)

National Edge AI Hub

Edge AI Benefits

- **Ultra-low latency**
  - Seamless connectivity, communication, and transparency
- **Adapting to environment**
  - Learning and inferring about new conditions
- **Privacy**
  - Confidential and private data
  - Not suitable for public cloud processing
- **Real-time actuation**
  - Automatically taking actions

Smart Transport

Healthcare

Manufacturing

Energy

Problem statement (Cyber-disturbances & Edge AI)

National Edge AI Hub

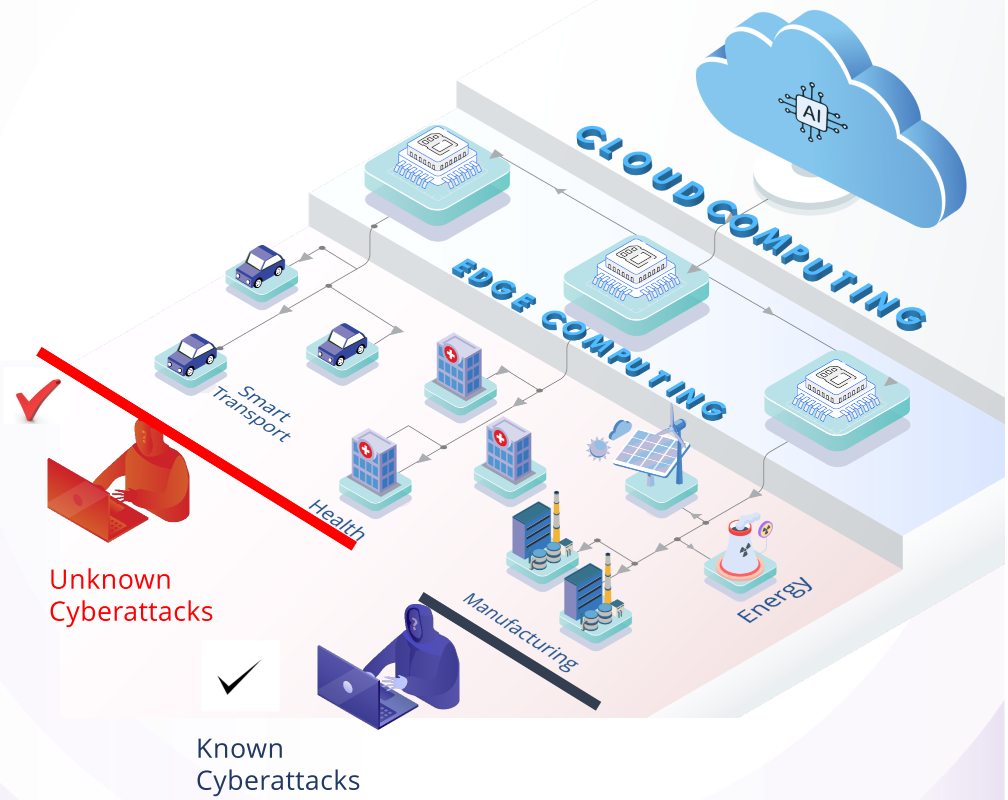
Ground-breaking Unsolved Research Question: How to ensure the Safety and Security of AI and Data from known and unknown "Cyber-Disturbances" at the Edge in Real-time?

Real-time learning Quality

- Model Reconstruction Attack
- Data Leakage
- Adversarial Data Injection

Real-time data Quality

- Missing Data
- Duplicate Data
- Poisoned Data
- Noisy Data



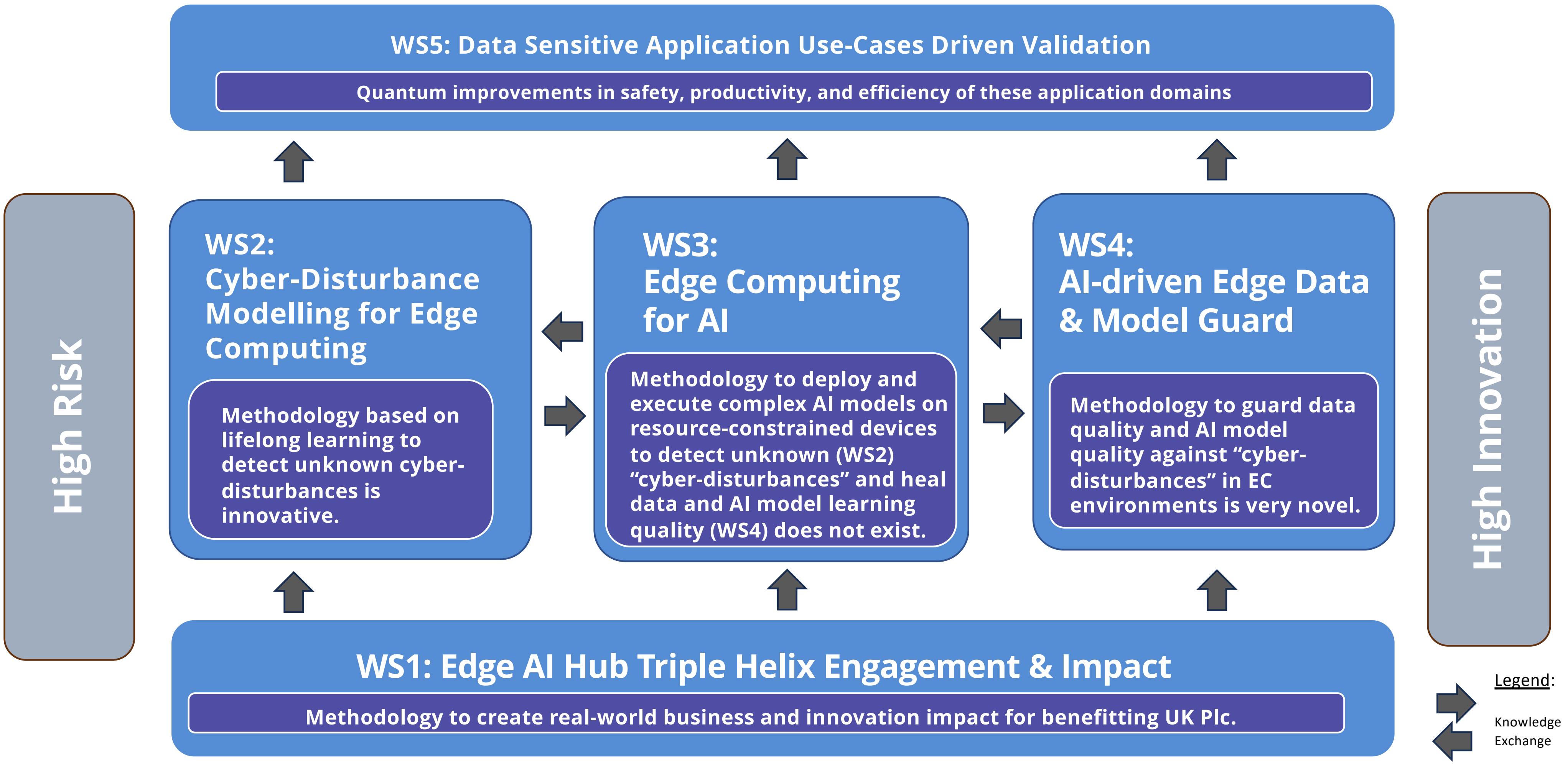
Hub Vision

To deliver world-class fundamental research, co-created with stakeholders from other disciplines and regions, to protect the quality of data and quality of learning associated with AI algorithms when they are subjected to known and unknown cyber-disturbances in the EC environments

Optimum Next Generation edgeAI technologies for handling cyber-disturbances require the ambitious vision!

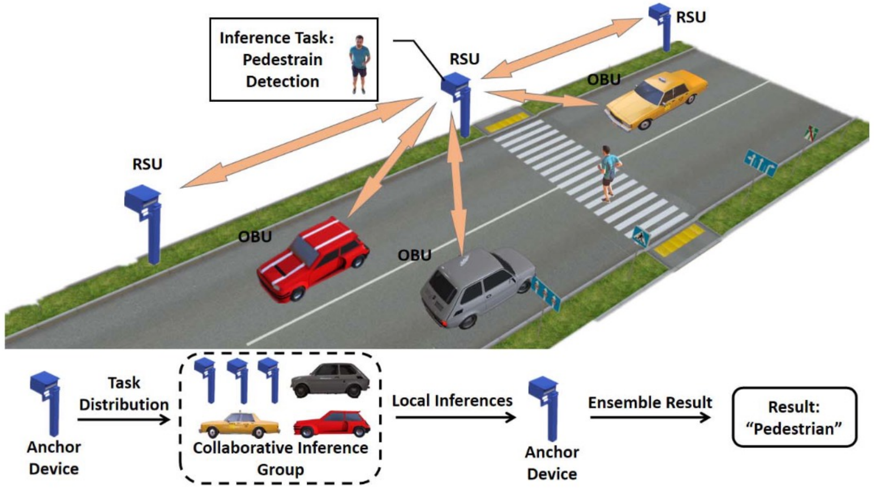
National Edge AI Hub

High Risk/High Gain Innovation through Workstreams



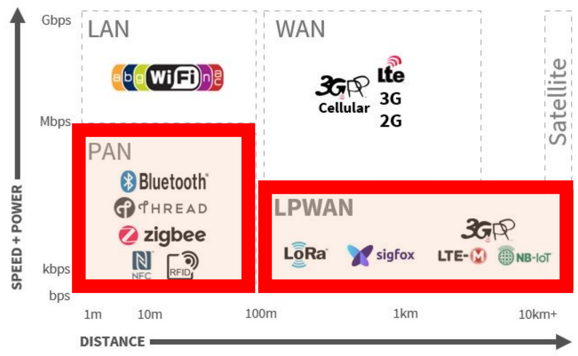
WS2: Cyber-Disturbance Modelling

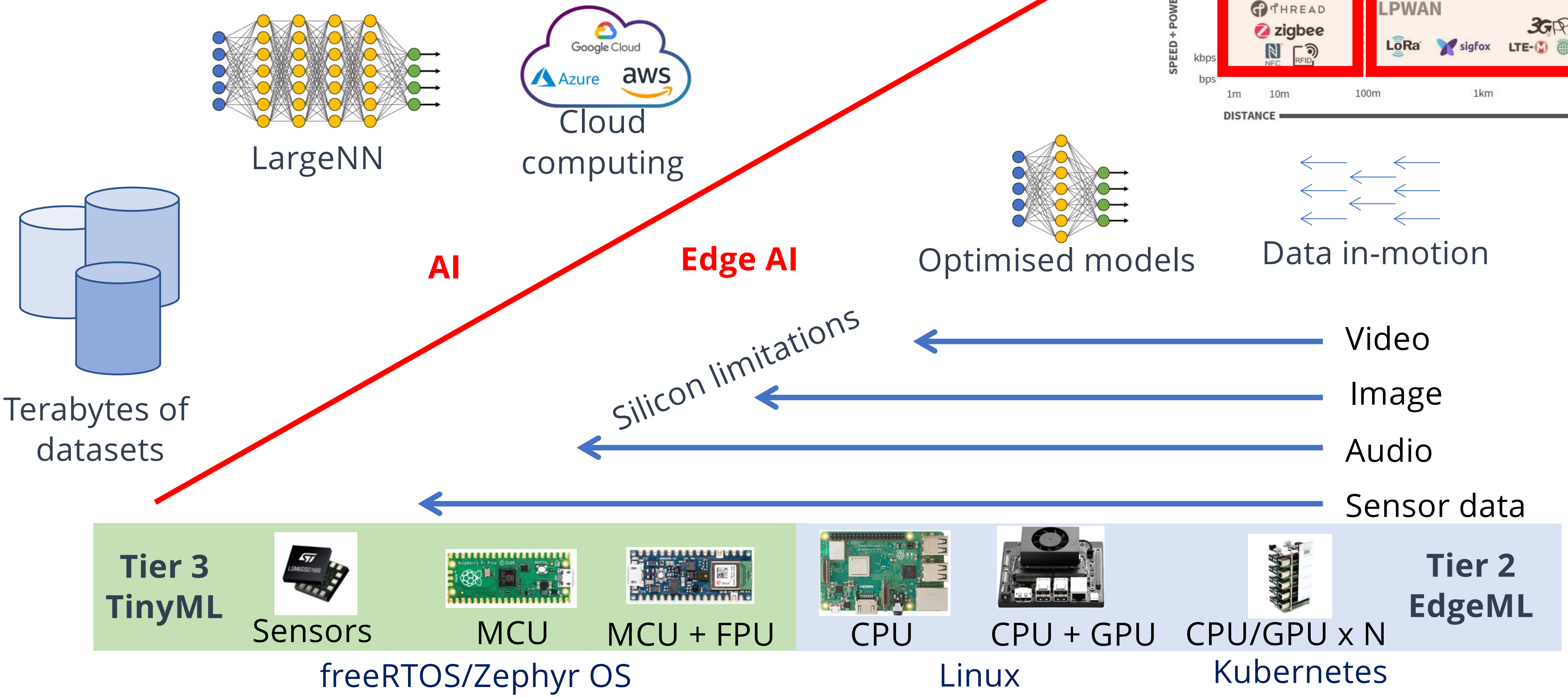
- Cyber-Disturbance Impact on Age of Data
  - Investigating how various cyber-disturbances affect the integrity and reliability of data in edgeAI systems.
- Towards Secure AI :
  - Improve fundamental understandings of how to secure AI: generative AI, and specific attacks (e.g. model inversion), inference-time security measures and classification-time security measures.
- Knowledge Graphs for Cyber-Disturbance Modelling:
  - Developing knowledge graphs that can effectively represent and analyse the complex relationships between cyber-disturbances and their attributes.



WS3: Edge Computing for AI

Computing Continuum





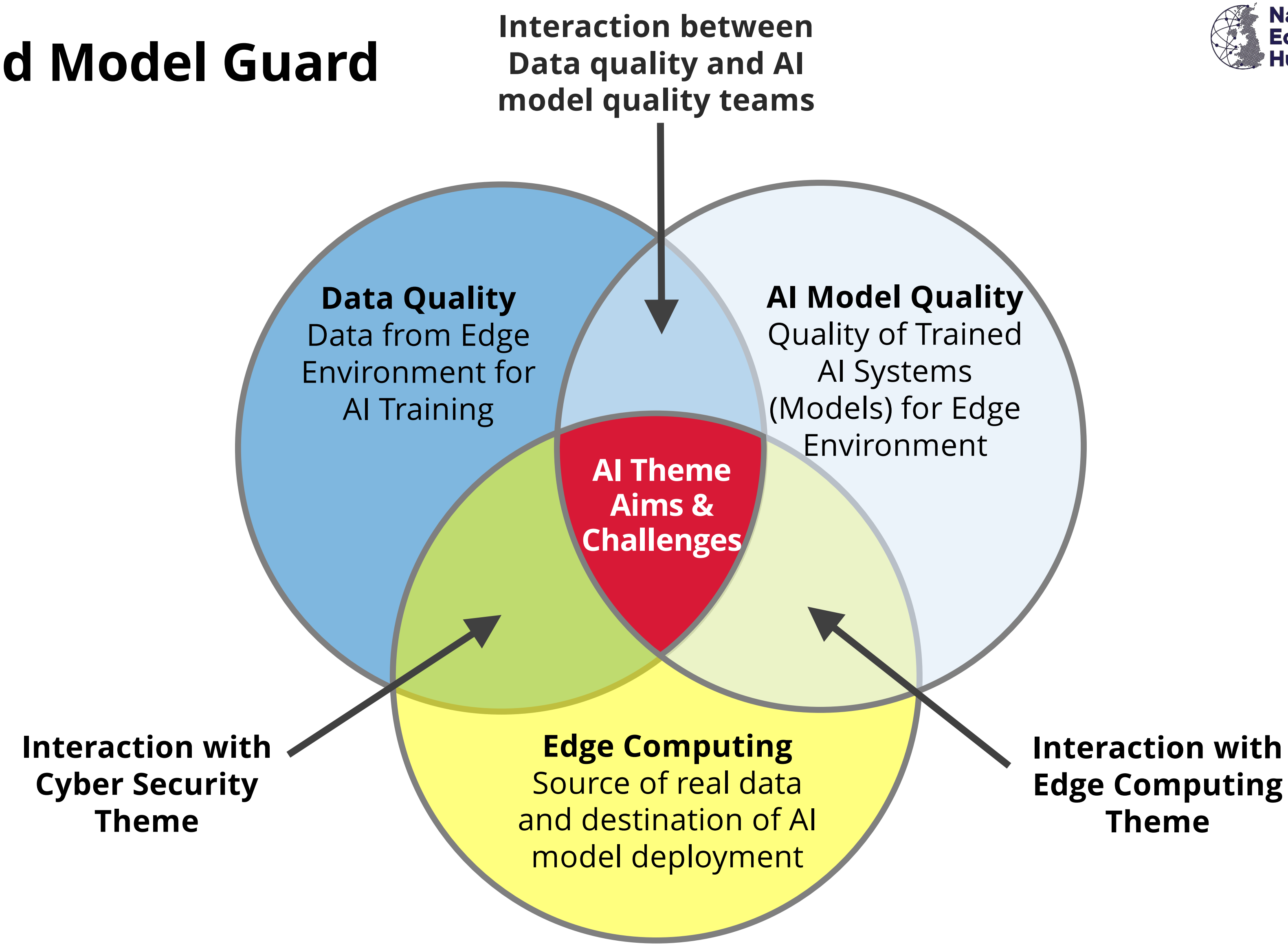
National Edge AI Hub

WS3: Edge Computing for AI

National Edge AI Hub

- **Edge Computing** systems **are complex** and can be **very different** from each other
  - Limited resources, hardware heterogeneity, several ML tools and techniques, intermittent communication, etc.
  - Makes it hard to deploy ML models to work well
- Aims to **simplify training and deployment** of complex Edge ML models
  - Execution of Edge ML models on different types and families of resources
  - Adaptation of Edge ML models to operational changes and failures in the end-to-end IoT system
- Create **new tools and techniques** that help practitioners achieve the above

WS4: AI and Model Guard



Federated Learning (all WSes): Alignment with Federated Compute Services NetworkPlus

National Edge AI Hub

Main Types of Federated Learning

Horizontal Federated Learning (HFL)

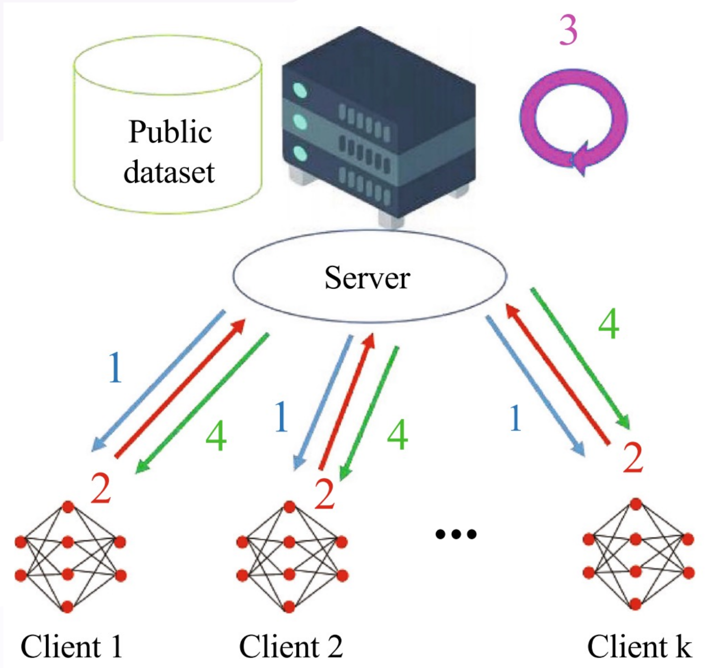
Example: Multiple banks can collaboratively train a credit scoring model without sharing customer information.

- all participants have data with the same features (e.g., different banks with similar customer transaction data)

Vertical Federated Learning (VFL)

Examples: A bank and an e-commerce platform can jointly train a model, with the bank providing users' financial data and the e-commerce platform providing shopping records to predict credit risk.

- all participants have data with different features (e.g., a bank with customer credit history and their shopping records)

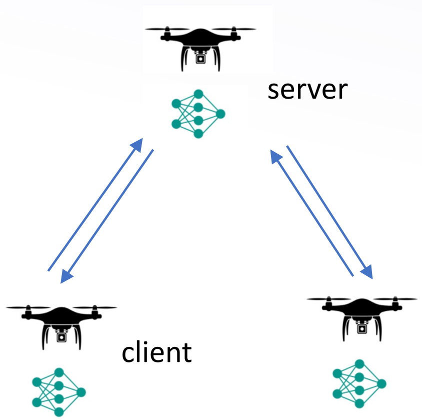


A schematic of federated learning. It includes four steps: 1: The central server sending the initialized global model to the client. 2: The clients then train locally and submit the local updates to the server. 3: The server performs the model aggregation. 4: the server sends the aggregated model to the clients [1]

[1]. Feng, Yunhao, et al. "A survey of security threats in federated learning." *Complex & Intelligent Systems* 11.2 (2025): 1-26.

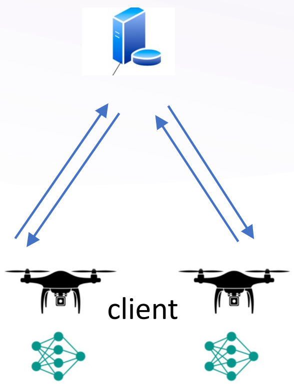
Federated Learning in UAVs

National Edge AI Hub



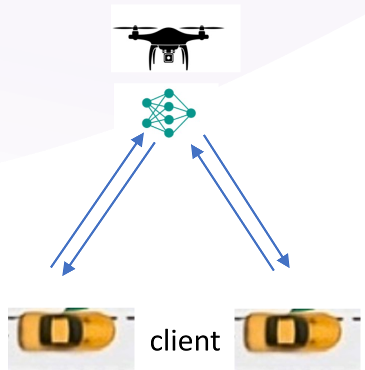
A Federated Learning Framework Composed **Entirely of Drones**[1], primarily designed for: **Object Detection and Tracking**, **Flight Path Optimization**, and **Obstacle Avoidance**.

UAVs as both server and clients



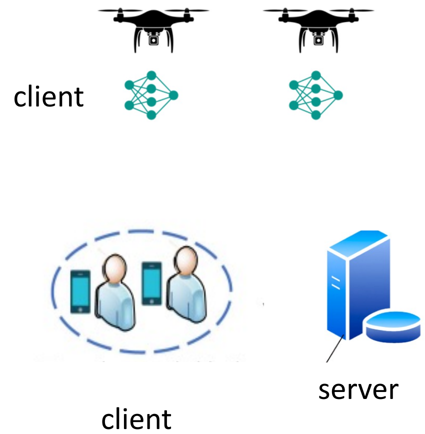
Model centric FL workflow in PySyft

UAVs as clients



A Federated Learning Framework **with Drone as the Server**[2], primarily designed for **Road Traffic Management and Congestion Control.**

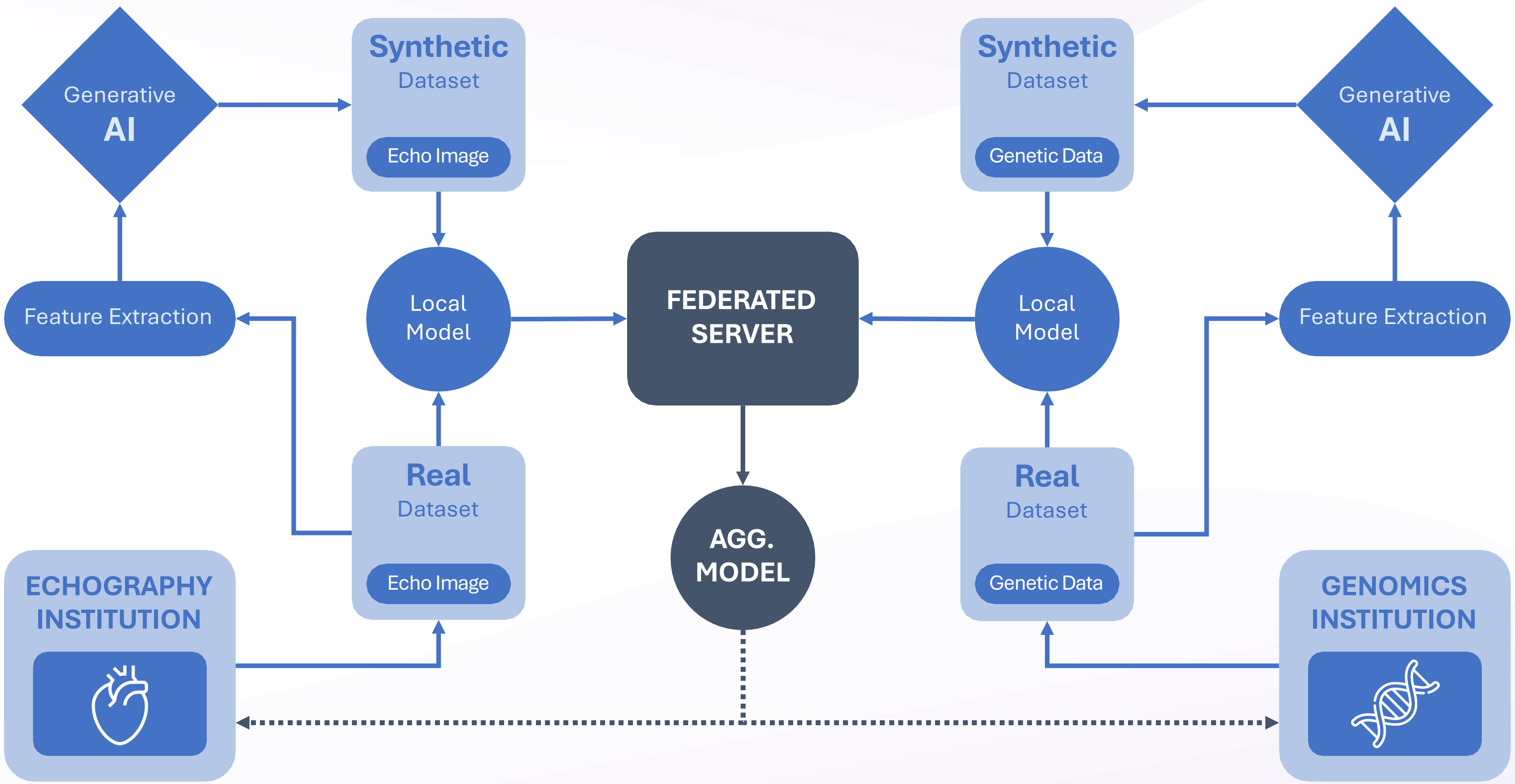
UAVs as server



Hybrid methods

[1]. Yazdinejad A, Parizi R M, Dehghantanha A, et al. Federated learning for drone authentication[J]. Ad Hoc Networks, 2021, 120: 102574.
[2]. Al Farsi A S, Khan A, Mughal M R, et al. Privacy and Security Challenges in Federated Learning for UAV Systems: A Comprehensive Review[J]. SECURITY AND PRIVACY, 2024.
[3]. Wang Y, Su Z, Zhang N, et al. Learning in the air: Secure federated learning for UAV-assisted crowdsensing[J]. IEEE Transactions on network science and engineering, 2020, 8(2): 1055-1069.

Implication of Generative AI on VFL



National Edge AI Hub

Federated Learning: Attacks (Varun/Shishir)

Attack on Federated Learning

Main Categories of FL Attacks

Integrity Attacks

These attacks aim to degrade the performance of the global model, causing it to make incorrect decisions.

Methods：

1. Disrupting Convergence (Model Poisoning)
2. Data Poisoning
3. Backdoor Injection

Privacy Attacks

These attacks attempt to infer or reconstruct private user data, leading to privacy breaches.

Methods：

1. Inference Attack
2. Gradient Leakage Attack
3. Model Stealing Attack

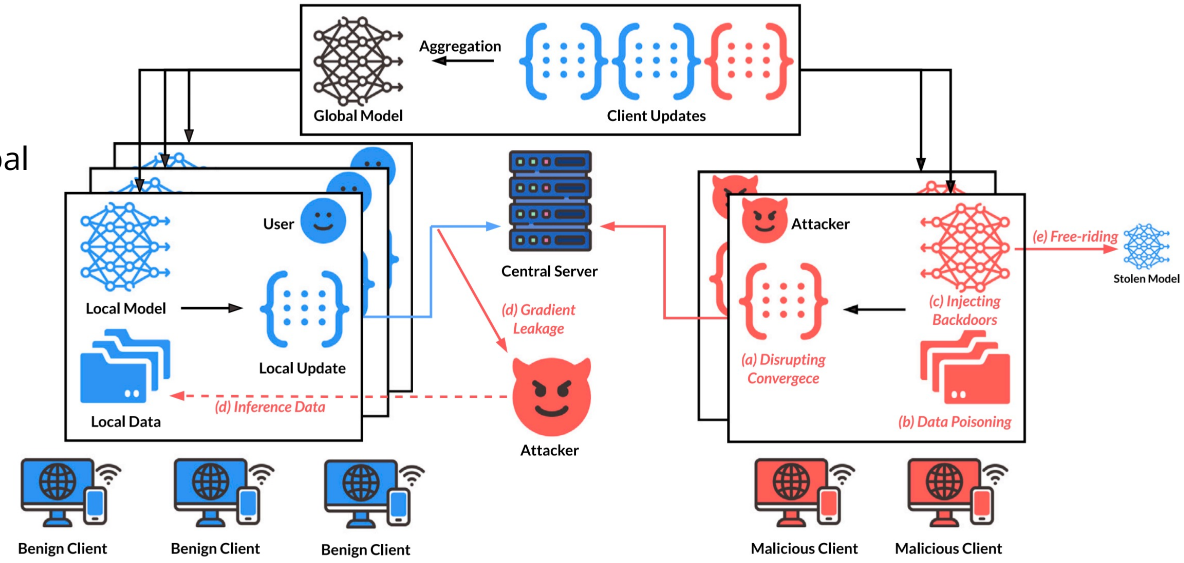


Fig. 1. An overview of common vulnerabilities in FL. Malicious attackers can: (a) manipulate model updates to prevent the global model from converging; (b) tamper data labels to induce erroneous predictions after training; (c) inject backdoors into the global model; (d) reconstruct data or inference data properties by eavesdropping model updates; (e) steal the global model while contribute nothing. [1]

[1]. Xie, Xianghua, et al. "A survey on vulnerability of federated learning: A learning algorithm perspective." *Neurocomputing* (2024): 127225.

Potential Risks of Federated Learning

National Edge AI Hub

Model Poisoning

Attack Method: Malicious drones/satellites inject tampered gradients to corrupt the global model.

Examples:

1. Misclassifying enemy drones as friendly.
2. Manipulating AI-driven traffic control to create congestion.

Data Poisoning

Attack Method: Injecting incorrect or mislabeled data during local training.

Examples:

Introducing false GPS coordinates to mislead other drones' navigation systems.

Backdoor Attack

Attack Method: Embedding hidden triggers that activate malicious behavior under specific conditions.

Examples:

Ignoring enemy vehicles with specific camouflage patterns.

Gradient Leakage Attack

Attack Method: Reconstructing original training data by analyzing gradients.

Examples:

Extracting sensitive images captured by drones.

Inference Attack

Attack Method: Determining if a specific sample was used in model training.

Examples:

Identifying if a drone has surveyed a restricted area.

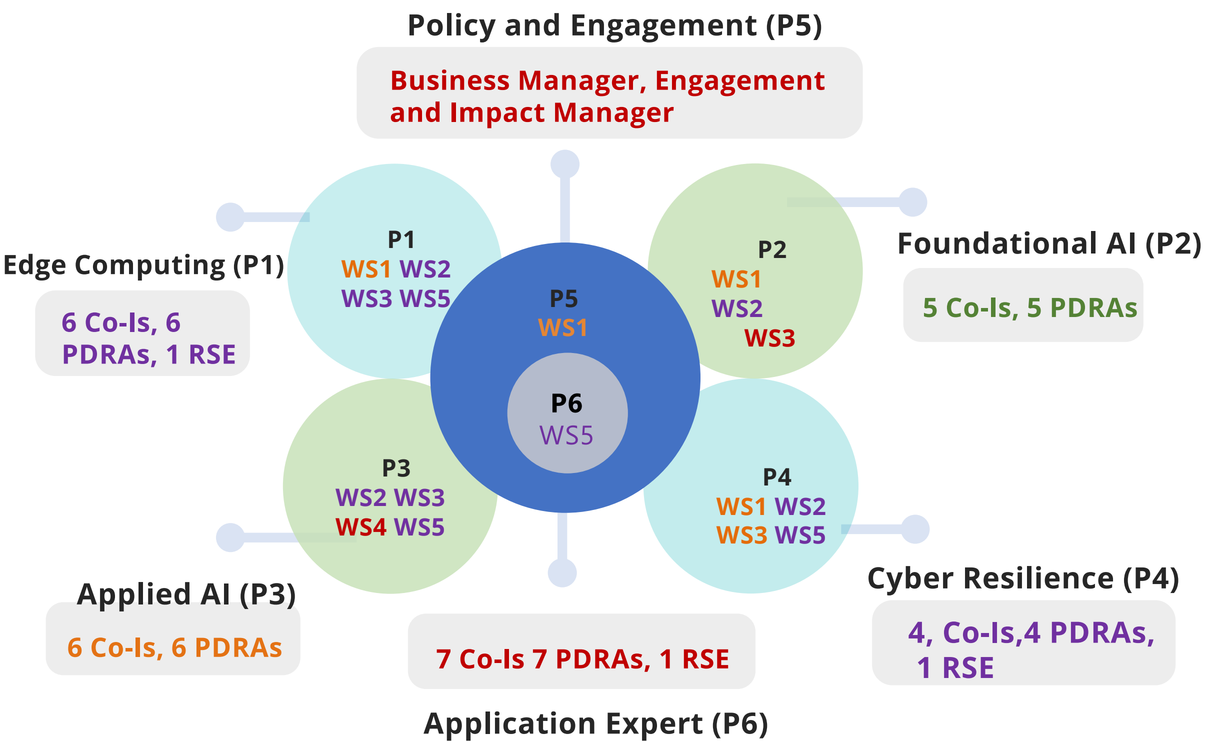
Model Extraction Attack

Attack Method: Reverse-engineering the model by repeatedly querying it.

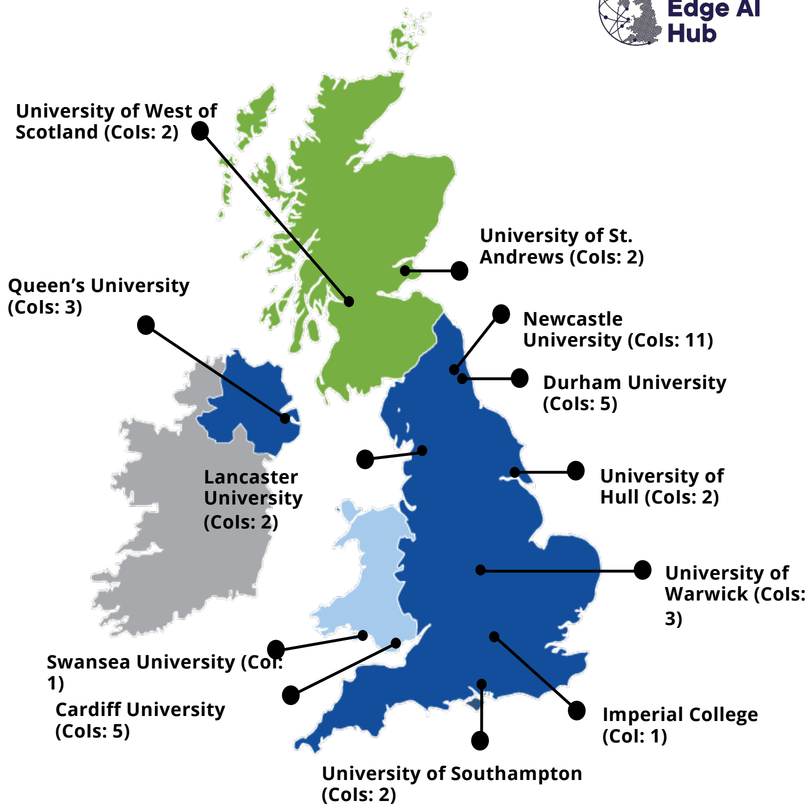
Examples:

Stealing an AI-based traffic control system to gain insights into urban infrastructure.

The consortium



Mapping of Research and Co-Is to Workstreams



National Edge AI Hub

Overlapping project management & administration expertise	Complementarity of partners' technical expertise	Covering all aspects of the development leading to the end-customer

Next edgeAI technologies for handing cyber-disturbances require multidisciplinary expertise!

National Edge AI Hub

Get in touch

Address

Urban Sciences Building, 1 Science Square,
Newcastle upon Tyne NE4 5TG, UK

Email

hub@edgeaihub.co.uk

Web

https://edgeaihub.co.uk