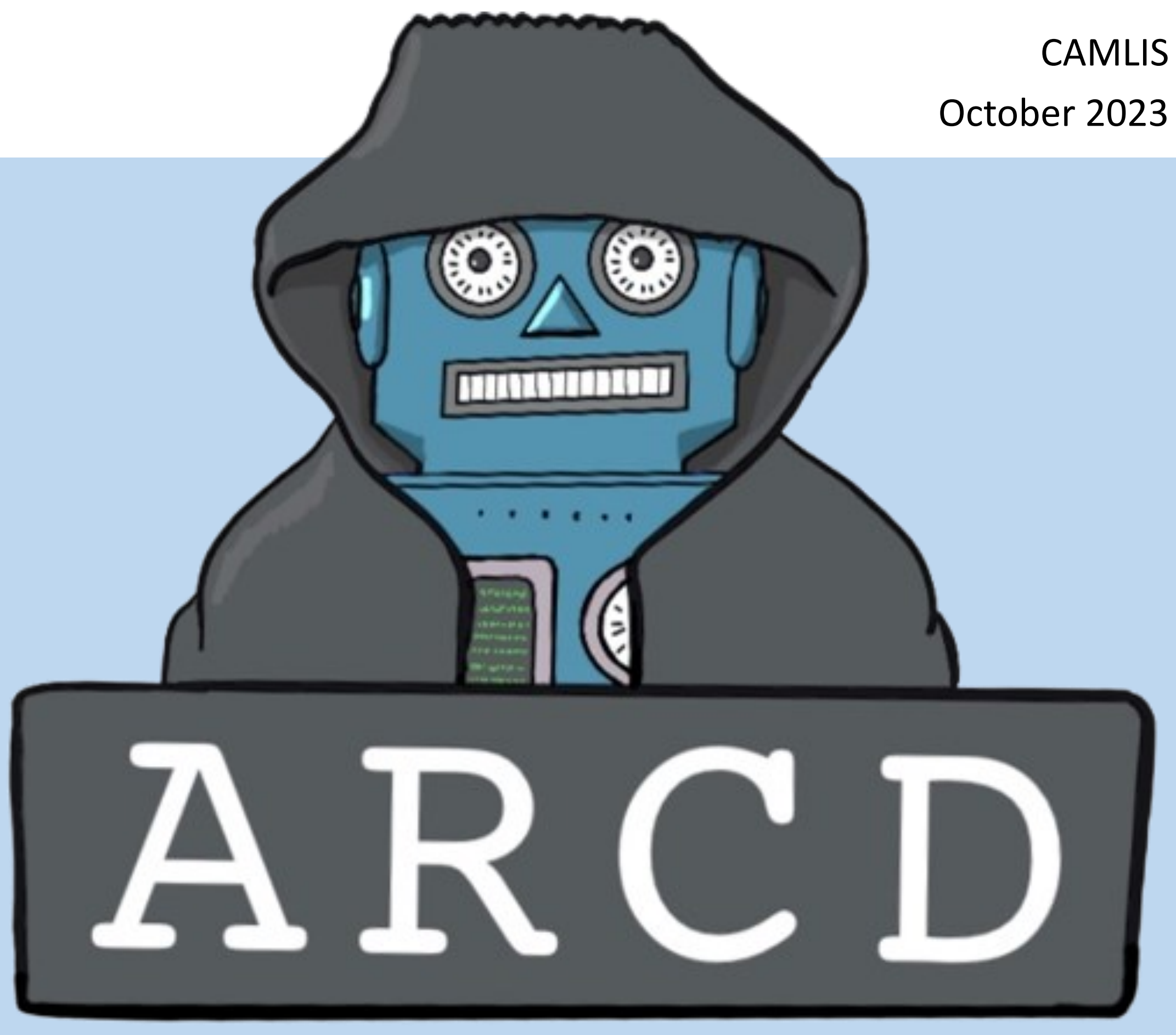
CAMLIS
October 2023

APPLYING MACHINE LEARNING TO ATTRIBUTE CYBER ATTACKS

Autonomous Resilient Cyber Defence (ARCD Programme

Informed Cyber Sensing (ICS) Project

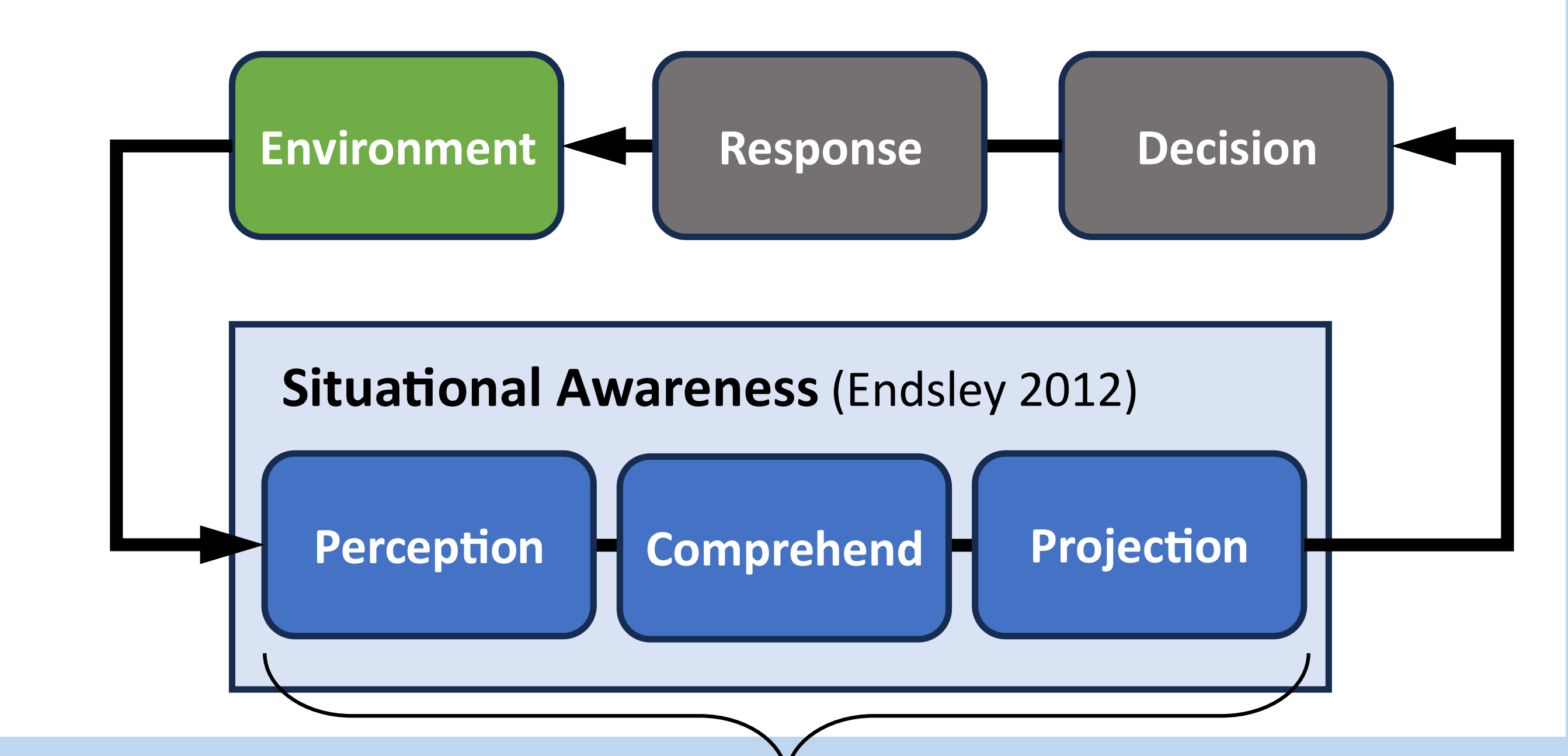


AI-Powered Cyber Attacks

AI is revolutionising cyber-attacks. With attackers using AI to create malware, ransomware and to automate phishing attacks. As AI continues to advance, so do the capabilities of cyber-attackers, not only in the speed, nature, complexity and volume of the attacks, but also in the evasion techniques used by attackers.

Situational Awareness of a Cyber Attack

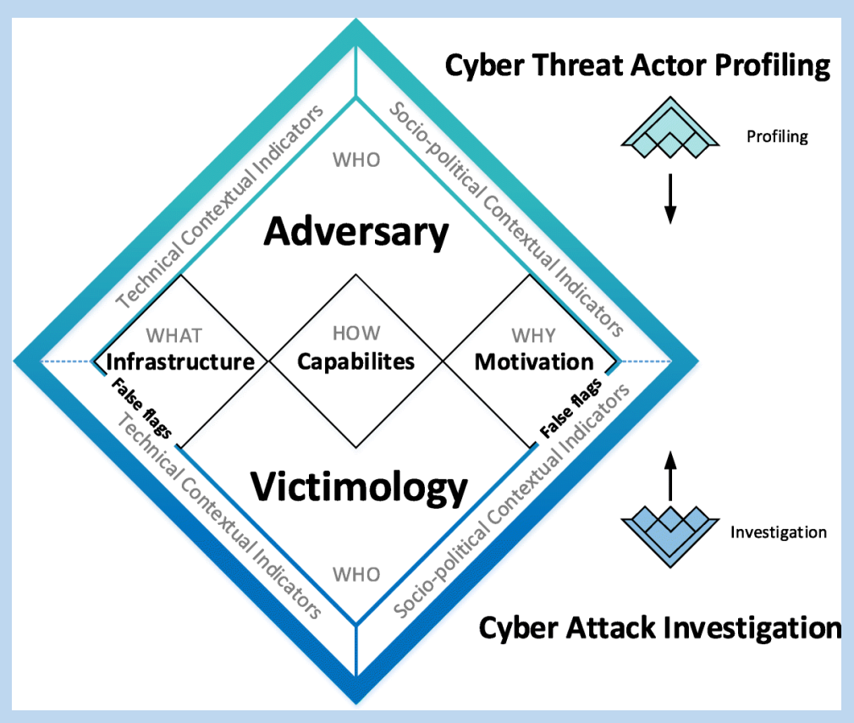
The Informed Cyber Sensing (ICS) project within ARCD, is undertaking research to develop ML techniques to provide insight into the threat and behaviour of the attacker.



Intelligence Preparation of the (Cyber) Environment (IPOE):

- Define the cyber environment
- Understand the cyber attacker's intent
- Evaluate the threat
- Determine the attacker's course of action

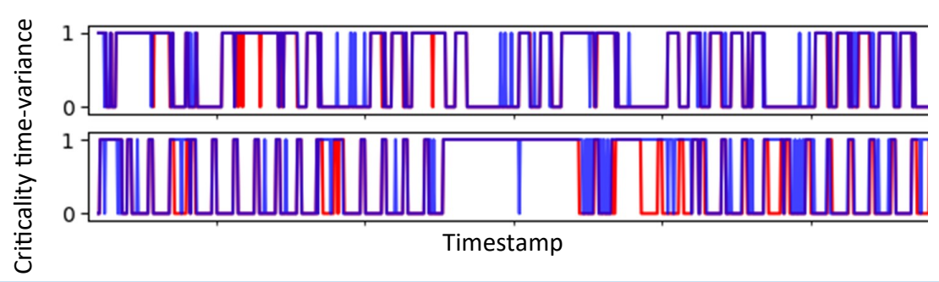
Attribution


Cyber Threat Actor Profiling
Profiling
WHO
Adversary
WHAT
Infrastructure
HOW
Capabilites
WHY
Motivation
Victimology
WHO
Technical Contextual Indicators
Socio-political Contextual Indicators
False flags
Technical Contextual Indicators
Socio-political Contextual Indicators
False flags
Investigation
Cyber Attack Investigation

Intelligence Preparation of the (Cyber) Environment

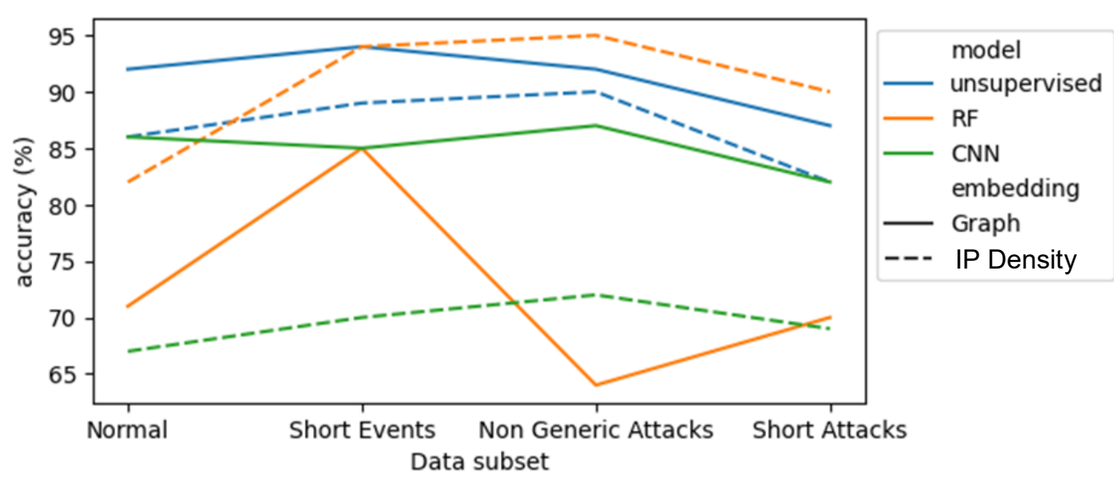
Critical Asset Cyber Terrain Identification (CACTI)

Creating a prototype tool to support identification of critical devices within a newly encountered network, utilising Graph Neural Networks (GNN) and related technologies (i.e. relational GNNs and Generative Adversarial Networks (GANs)). Phase 1 of the work demonstrated the model over classified a little, comparing ground truth (red) against model (blue) results. Phase 2 of the work is undertaking research to respond to cyber-attacks within a dynamic environment in which criticality of assets changes over time.



Topological Data Analysis for Network Data

To address what methods of transforming the data (e.g., Wasserstein Distance for persistence diagrams) into a TDA compatible format are the most effective for preserving anomaly/attack relevant features. This work demonstrated that a TDA-based intrusion detection system is a viable method of detecting anomalous activity. Where event types were altered, i.e., removing long events, the IPD method for RF showed a distinct improvement.



Applying ML to Cyber-Defence

The UK MOD Autonomous Resilient Cyber Defence (ARCD) research, funded through Dstl, is looking to:

> ***"develop self-defending, self-recovering concepts. Autonomous analysis, decision making and re-configuration of networks at machine speed"***

for Generation After Next (GAN) MOD capability, (completing March 2025) aiming to achieve the following outcomes:

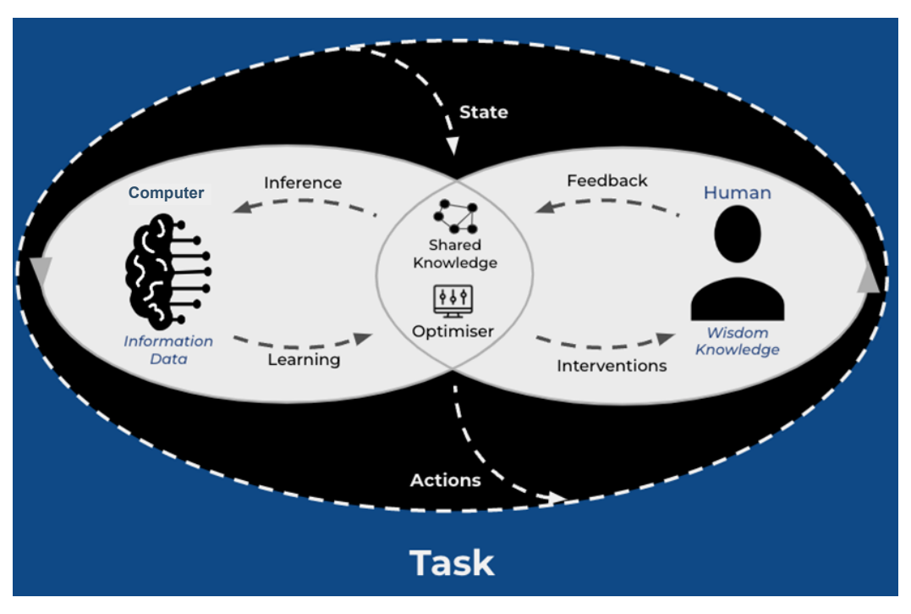
- A **concept demonstrator** (or demonstrators) capable of autonomously responding, using ML concepts, to cyber-attacks in the context of a military environment and mission.
- Improved understanding of the **strengths, and limitations of AI/ML technologies** and their applications to cyber defence.
- **Greater capacity in the UK supplier base** to support AI/cyber research areas.

Attribution

Modelling Adversarial Behaviour

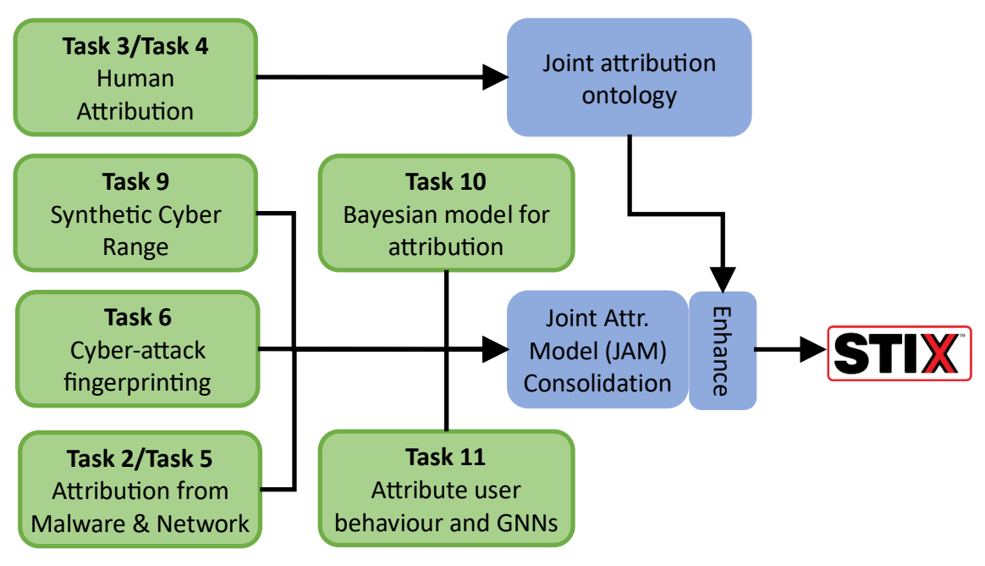
Develop a Blue Agent capable of counterfactual reasoning to predict adversarial actions.

TUPLA is modelling the decision making behaviour of human experts (pen testers, red teamers), developing and demonstrating an interpretable + effective Red Agent using that model, training the Blue Agent on that same model and implement causal inference (e.g., counterfactual reasoning).



Attribution of Cyber Attack with ML

We have applied ML to the problem of autonomously attributing a cyber-attack to a threat actor. K-Nearest Neighbour (KNN) algorithms showed improved accuracy (~95%) attributing a nation state from an Advanced Persistent Threat Group over previous work, *(Dstl Serapis U60),* using a Deep Neural Network (DNN), which returned an accuracy figure of close to 80%. Attribution tasks under ICS, including this work, could be integrated into JAM.


Task 3/Task 4
Human Attribution
Joint attribution ontology
Task 9
Synthetic Cyber Range
Task 10
Bayesian model for attribution
Task 6
Cyber-attack fingerprinting
Joint Attr. Model (JAM) Consolidation
Enhance
STIX
Task 2/Task 5
Attribution from Malware & Network
Task 11
Attribute user behaviour and GNNs

So What?

ML algorithms have shown promising results when applied to solving discrete challenges. This has enabled automating traditional human approaches to IPOE and the attribution of malware samples, networking data and telemetry logs.

ARCD Track 1 Frazer-Nash Consultancy
arcd@fnc.co.uk
Author: Steve Little (Frazer-Nash Consultancy)
Patrick Ewen (Dstl)