ARCD TRACK 1

Newsletter

Welcome to the ARCD Track 1 Newsletter!

As we near the end of the four-year ARCD (Autonomous Resilient Cyber Defence) research project, concluding in March 2025, we reflect on our journey leading the Ministry Of Defence's (MOD) exploration into Generation-After-Next (GAN) Cyber Defence concepts.

The ARCD programme has been a hub of innovation and collaboration, highlighted by dynamic events such as ARCD roadshows, September Demonstration Days, and key presentations at Black Hat in Las Vegas, CAM LIS in Washington, and the AMLUCS conference.

Our extraordinary research and development have been driven by the shared knowledge, innovation, and hard work of our diverse supply chain. Contributions from SMEs, academic institutions, non-traditional defence suppliers, and large organisations have been crucial in shaping the success of ARCD.

Contact

Email: arcd@fnc.co.uk

Web: fnc.co.uk/arcd

Autonomous Resilient Cyber Defence

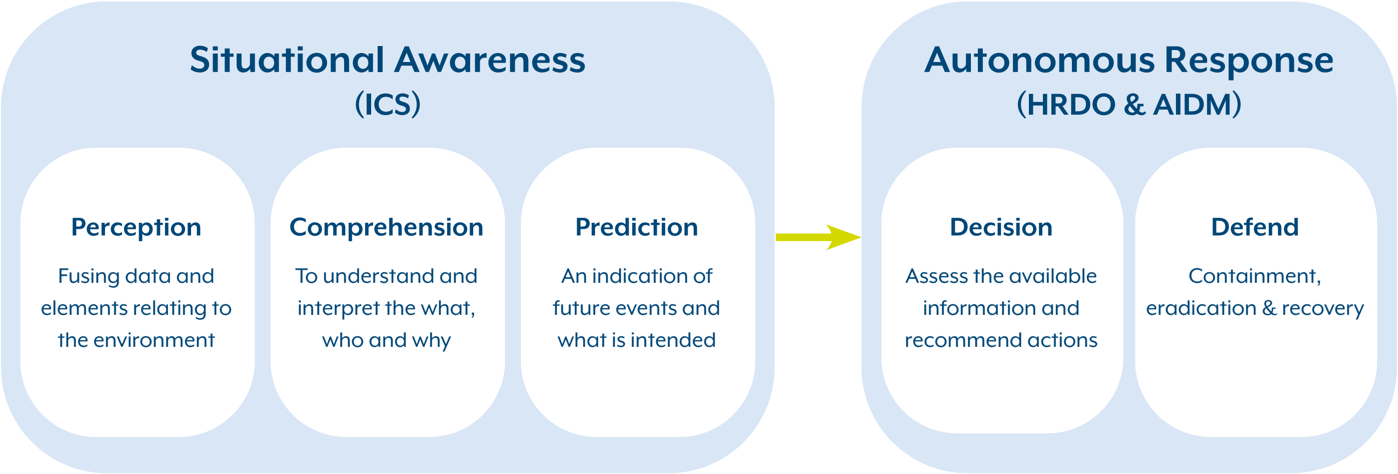
Cyber-attackers are increasingly using AI-driven techniques to identify vulnerabilities, predict patterns, and exploit weaknesses at scale, posing persistent and substantial threats to the MOD.

Due to the volume and complexity of these attacks, the MOD cannot scale its human defence capabilities adequately. ARCD, funded by the Defence Science and Technology Laboratory (Dstl), aims to develop self-defending, self-recovering concepts for military operational platforms, aspiring to achieve 'Full Auto' Cyber Defence.

ARCD's goals include:

- Creating a concept demonstrator capable of autonomously responding to cyber-attacks in a military context.
- Enhancing cyber and Machine Learning (ML) skills in the UK supply chain.
- Understanding the strengths and limitations of applying ML technologies in Cyber Defence.

Frazer-Nash leads the development of cyber defence concepts (Track 1), engaging with a diverse supply chain through Serapis Lot 6 to develop AI/cyber agents. As we enter the final year of work, several developed concepts are now exploitable to support human cyber defenders.



Situational Awareness (ICS)

Perception: Fusing data and elements relating to the environment

Comprehension: To understand and interpret the what, who and why

Prediction: An indication of future events and what is intended

Autonomous Response (HRDO & AIDM)

Decision: Assess the available information and recommend actions

Defend: Containment, eradication & recovery

Programme

ARCD is a unique contract for Dstl, being firm price and outcome based, necessitating a large and diverse supply chain.

Unlike typical transactional or mutual partnerships, Dstl sought a symbiotic relationship for ARCD, with a partner responsible for technical and commercial delivery, guided by Dstl.

Dstl awarded Frazer-Nash a 3-year firm price contract, with progress measured against key performance indicators. Frazer-Nash has awarded over 100 tasks to suppliers and hosted numerous stakeholder and industry events, making the ARCD Programme a real success.

Supply Chain

Over the past four years, we've had an incredible journey working with a vibrant mix of SMEs, academia, non-traditional defence suppliers, and large organisations, contracting over 100 tasks and funding more than £20m of research.

A highlight this year was the ARCD Demonstration Day in September 2024, showcasing our remarkable research and achievements. Positive feedback from key stakeholders reaffirmed the importance of this critical research and the progress made. We look forward to building on these successes and preparing for the next Demonstration Days in February 2025.

In addition to our daily work with Dstl Military Advisors, we've been busy with the ARCD Roadshow, delivering tailored briefings to Navy Command, The Defence Cyber Range, 591 Signals Unit, MAB, and the Defence Digital Rapid Reaction Troop, among others. Plans are in place for the Roadshow to visit Army HQ early in 2025. These roadshows have increased awareness of ARCD research and improved understanding of operational challenges and integration requirements.

Several near-term exploitation opportunities have been identified, including Green Pattern of Life and Red Agent with the Defence Cyber Range, and user trials of ICS tooling within MAB and Defence Digital.

To book your ARCD roadshow, please contact arcd@fnc.co.uk.

Dstl has agreed to a Triangle Project Agreement with DARPA, enabling collaboration across several similar projects. As ARCD concludes in March 2025, marking the end of our four-year contract, we are committed to continuing our collaboration and maintaining momentum.

If you have any queries, please reach out to the team at arcd@fnc.co.uk.

High Risk & Disruptive Options

Alex Revell – HRDO Technical Lead

The objective of HRDO is to explore how AI can be applied in cyber defence through research projects funded in response to periodic problem book releases. A key aspect of this is understanding the strengths and limitations of AI, which depend on the underlying challenges of the technology.

The challenges include:

- Explainability of 'black box models': How well can humans understand AI decision-making?
- Efficiency regarding the compute and data requirements for training and running AI models.
- The ability of AI solutions to generalise to various cyber threats and military scenarios.

HRDO projects are low TRL, addressing proof-of-concept for high-risk, cutting-edge solutions. To date, HRDO has contracted over 50 proposals aiming to provide AI solutions for cyber defence. These projects include novel AI solutions and applications of cutting-edge AI to specific cyber threats in various military scenarios.

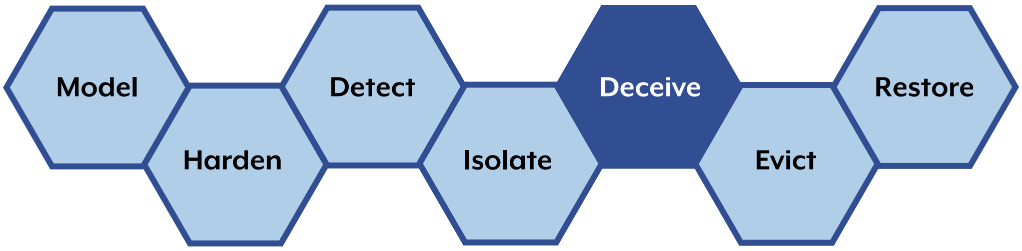
This update highlights some projects focusing on AI for cyber deception as a solution to current and future cyber threats. Deception tactics involve techniques to advertise, entice, and allow potential attackers access to a controlled environment, thereby protecting the underlying system. Dynamic deception enables network defenders to apply these techniques in response to ongoing cyber-attacks.

There exist several challenges in implementing effective responsive deception strategies:

Optimisation:
An optimal deception strategy should consider aspects such as attacker motivation or preference, defensive cost and likelihood of success.

Authenticity:
For an attacker to be deceived by deceptive artefacts they must imitate genuine assets. Efforts to evaluate the authenticity of AI generated deception is aimed at supporting this concept.

Adaptable:
Cyber deception involves configuration of tailored deceptive components. To replicate this, AI based automated deception should be adapted to respond to evolving threats. Our PoL task focused on green traffic generation to evaluate the feasibility of AI generated network traffic that can be employed to imitate active decoy networks.



Deception is one of seven tactical groups defined in the MITRE D3FEND framework, a knowledge base of cybersecurity countermeasures that can be used as part of a cyber defence strategy.

More information on the HRDO tasks can be found from page 9. Deception based tasks are 4, 28, 31, 36.

The research outcomes provide an understanding to what extent ML can be used to provide a cyber deception capability. The range of deception-based tasks are at varying stages, but so far, the following AI based deception has been investigated:

1. Can LLMs be used to provide database decoys?
2. Can AI support optimisation of deception strategies against different adversarial profiles?
3. Can AI generated artefacts resemble genuine network assets or data to deceive an adversary?

Deploying deception-based infrastructure is limited to existing techniques and reliant on human input to configure. Evaluation of different new techniques and strategies using a variety of AI approaches helps to better understand a future autonomous deception capability with wide ranging benefits. Building on the existing effective cyber deception techniques, research into aspects of optimisation, deception authenticity and adaptability further leverages an autonomous and rapid response capability to extend a defensive armoury dealing with next generation cyber threats.

Informed Cyber Sensing

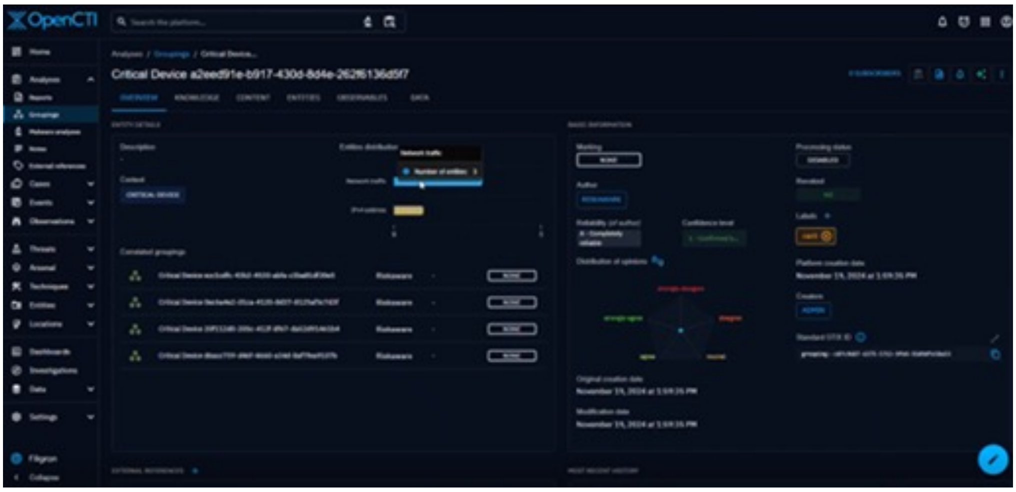
Steve Little – ICS Technical Lead

Since the start of ARCD ICS has focused on developing methods for autonomously characterising the operating environment, optimising data collection and improving understanding of adversary intent and behaviour. These activities are typically human intensive processes.

For generation after next cyber defence, autonomous response and recovery agents need to have knowledge of the operating environment and the intent and behaviour of the attacker. At the start of ARCD this was the primary purpose of ICS, however a number of capabilities have been identified as being able to support current cyber defenders, and trials have been undertaken or are being planned across MOD, including Defence Digital, 591 Signals Unit and the Defence Cyber Range. ICS is a component of MOD's work to provide cyber situational awareness, integrating capabilities from other programmes.

- To identify critical assets in a network, **Riskaware** has developed a supervised machine learning system which classifies Internet Protocol (IP) addresses into critical and not-critical categories by looking for patterns in network traffic for the Critical Asset Cyber Terrain Identification (CACTI) system.
- **Riskaware** has integrated three components, to form **CyberAware Subsystem**
  - **CyberAware Predict**, which predicts the next steps of evolving cyber-attacks based on network scans, real-time monitoring and Cyber Threat Intelligence.
  - **CyberAware Resilience**, which models cyber-attacks using advanced common vulnerability scoring to determine the potential impact of cyber-attacks on missions that depend on network assets. It can be used to support mission planning modelling the dependency missions have on cyber assets.
  - **CACTI**, (see above).
- **Montvieux** has developed the **Joint Attribution Modelling (JAM)** framework, which provided a standard structure for data, tools and models, covering malware text reports, network traffic captures and log telemetry. Additionally, JAM is being augmented with specialised Advanced Persistent Threat (APT) level attribution models and the tool has been refactored as a scalable cloud-based service to further development and access.
- **Elemendar's READ** tool uses Named Entity Recognition models to process human authored, unstructured Cyber Threat Intelligence (CTI) reports, into structured CTI data.
- In support of Human/Machine teaming, through a series of experiments, **Tulpa** has captured expert knowledge of attack tactics, techniques, and procedures into a causal knowledge graph. Building on this work Tulpa has implemented their red adversarial agent and blue predictor agent technology (**Red Faction**) into key ARCD environments, such as Cyborg and PrimAITE.
- **BMT** has created 'Attack Planners' using Finite State Machines to extend documented OT cyber-attacks (e.g. Stuxnet, Industroyer, Black Energy) to support analysis of attacks on a Maritime Platform Management System (PMS) and mimic attacks from adversaries with varying capability and resources.

Most of the above tasks utilise the Open-Source platform, OpenCTI, which is used by a number of cyber agencies for the creation, ingestion and dissemination of threat information. It can develop a threat landscape by understanding a threat actor's current behaviours, their ongoing campaigns and by monitoring malware and vulnerabilities.



AI Decision Making

Ian Miles – AIDM Technical Lead

AIDM has made great strides in testing autonomous cyber response & recovery concepts within more realistic & representative environments, taking our total commitment to just under £10M across 33 projects, adding to the UK's portfolio of cyber defence research, including the MOD and National Security.

The last year has seen many landmarks, including:

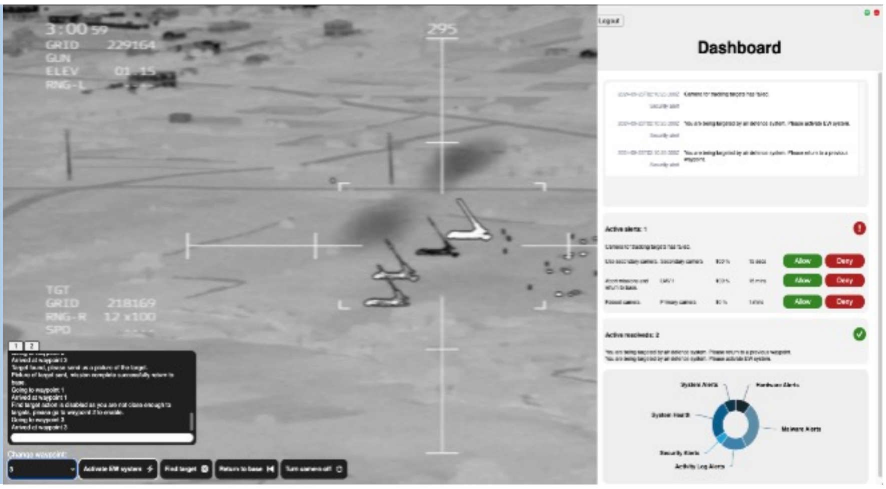
A collaborative technical paper detailing a snapshot of our most mature autonomous response and recovery research was supported by 30 authors across 11 AIDM organisations and underpinned a presentation at Black Hat USA. Recent presentations at CAMLIS and AMLUCS, including papers from Cambridge Consultants and BMT, continue to support external peer review, raising interest in our suppliers' research and generating international collaboration opportunities.

BMT (with ADSP) collaborating with **Dstl** to deploy their Multi-Agent Reinforcement Learning (MARL) autonomous response concept into Dstl's highly representative Royal Navy Platform Management System (PMS) proxy system in what we believe to be a world first for automated cyber defence of Military OT/Industrial Control Systems.

There has been strong progress towards demonstrating real autonomous cyber defence benefits. Highlights include:

- ADSP's 'ARCD Demonstration Agent' project, a holistic cross-track demonstration of ARCD capabilities with QinetiQ (Track 2). Agents trained in PrimAITE are transferring to the PalisAIDE demonstration environment, with independent Track 2 evaluations introducing a feedback loop to improve performance.
- Exalens and Frazer-Nash deploying their cyber first aid agent to a Dstl drone system. Initial field trials showed successful integration for cyber responses, with further testing in January.
- Trustworthy AI developing an emulated environment with red and blue agent actions using Cobalt Strike and Elastic Defend. Accenture will conduct ARCD's first human red teaming exercise on this agent.
- BAE Systems collaborating with DARPA CASTLE teams for more robust agent testing against unseen adversaries and network topologies.
- Cambridge Consultants extending their Generic Vehicle Architecture (GVA) simulation environment, demonstrating co-operative MARL agent defence against attacks on navigation and vehicle-to-vehicle communications. Early discussions are exploring deployment to QinetiQ's GVA Digital Twin.
- BT developing holistic network-focused reward structures with their military defensive cyber teams to improve understanding of real-world impacts and performance.
- HRDO concepts graduating into mature research in more complex environments (Montvieux's World Models and ADSP's zero-shot LLM agents).
- A workshop bringing together ATI and researchers from AIDM's mature MARL projects to explore challenges and opportunities, with a paper due in early 2025.

Tulpa's team of nationally recognised experts in Human Machine Teaming (HMT), autonomous cyber defence and human sciences exploring HMT and building trust in ARCD capability. 10+ military stakeholder interviews identified early user requirements, and a range of HMT prototypes will elicit user feedback to inform a roadmap for adoption of autonomous cyber defence capability.



Arrived at waypoint 3
Target found, please send us a picture of the target.
Picture of target sent, mission complete successfully return to base.
Going to waypoint 1
Arrived at waypoint 1
Find target action is disabled as you are not close enough to targets, please go to waypoint 2 to enable.
Going to waypoint 2
Arrived at waypoint 2

Camera for tracking targets has failed.
Security alert

You are being targeted by air defence system. Please activate EW system.
Security alert

You are being targeted by air defence system. Please return to a previous waypoint.
Security alert

Active alerts: 1

Camera for tracking targets has failed.

Use secondary camera	Secondary camera	100%	15 secs	Allow	Deny
Abort mission and return to base	UAV1	100%	15 mins	Allow	Deny
Reboot camera	Primary camera	10%	1 mins	Allow	Deny

Active resolveds: 2

You are being targeted by air defence system. Please return to a previous waypoint.
You are being targeted by air defence system. Please activate EW system.

Track 2 - QinetiQ

Will Bowers – Track 2 Programme Lead

Track 2 – Test & Evaluation (T&E) is responsible for performing Experimentation and Evaluation (E&E) of the agent performance.

It provides the environments to train, evaluate and demonstrate the agents; the evaluation schemes and tools to measure the performance and measures of effectiveness of the agents; and an actuation and integration (A&I) capability to provide agent, environment and evaluation orchestration. As QinetiQ comes to the final quarter of the final year of the current programme, the following update will show the incredible work undertaken by T&E and their supporting functions.

The Training and Demonstration Environments (T&DE) team have delivered multiple product iterations over the last year. Following a re-architecture and significant feature enhancement, PrimAITE v3.0 was delivered in spring 2024, and the current release (v3.3.1) is providing capability to multiple Track 1 projects, including AIDM U75c Demonstration Agent. PrimAITE has also received external interest from organisations such as the ATI, NCSC and other areas of Dstl. Imaginary Yak was consolidated into release 2.0 and has been made available for those wishing to conduct experimentation with emulation environments; it also delivered a ground-breaking project which saw it integrate with the Dstl RL-based Red Agent tool, Automatic Jack. The Demonstration Environment, PalisAIDE, has undergone a huge development over the last year, growing from a simple prototype into a full virtualised cyber range. Through the use of Terraform, Ansible, and a number of off-the-shelf products, PalisAIDE now models the enhanced "Use Case 6" military scenario and is deployed to an Azure cloud platform where it integrates with ARCHON (see below). This presents a test-bed for the QinetiQ E&E team to conduct defensive agent evaluation, assess the practice of agent transfer between environments of differing fidelity, and explore the impact of sim-to-real. Finally, DaRIUS was delivered at v1.0 and has been presented to the Dstl customer, as well as representatives from MoD STRATCOM.

The A&I team have been enhancing the functionality of ARCHON with new capabilities to support the evaluation workflow (ARCHON is a component-based solution that provides the orchestration of the interaction of an agent with its environment, using an open standard sensor harness that will present the required system inputs and the collection of sampled data required to support evaluations). This has started with the integration of the metrics tool ANANSI from the Experimentation and Evaluation (E&E) team. This will provide the E&E team with data that has been pre-processed by ARCHON and can be used directly for evaluation. Alongside adding new features, the A&I team have been expanding the ARCHON data model and environment wrappers to integrate the new versions of the environments released by the T&DE team. New agent wrappers have also been created allowing agents in use by the E&E team to be deployed into both PrimAITE and PalisAIDE environments with minimal additional effort. We have contracted an external party to revise a capability developed in Track 1 to demonstrate the plug and playability of ARCHON.

Team members from the three ARCD T&E Tasks have been engaging with ADSP (Track 1) who are delivering the ARCD Demonstration Agent project, training their agents for evaluation in PalisAIDE – a wholly collaborative, cross-track ARCD task. Track 2 have now completed evaluations of a number of agents.

These evaluations have allowed E&E to test and improve the evaluation process; make large improvements to the speed of evaluations; and provide validation of how metrics at different levels can be used to assess different agent behaviours. Work is progressing to document the overall evaluation process and details of tools and techniques that have been used as part of the research programme.

Development is progressing collaboratively across the ARCD T&E Tasks to allow agents to be evaluated in both PrimAITE and PalisAIDE via integration with ARCHON.

If you have any questions on the update above, please email arcd-track2@qinetiq.com.

Alan Turing Institute

The AI for Cyber Defence (AICD) Research Centre, established in collaboration with Dstl as part of the ARCD programme, is advancing research into cutting-edge AI methodologies to deliver impact in Autonomous Cyber Defence (ACD).

AICD plays an essential role across both Tracks 1 and 2 of the programme, providing technical leadership and foundational research to underpin the development elective ACD solutions in the military domain.

Led by principal investigators Vasilios Mavroudis and Chris Hicks, AICD's research areas include:

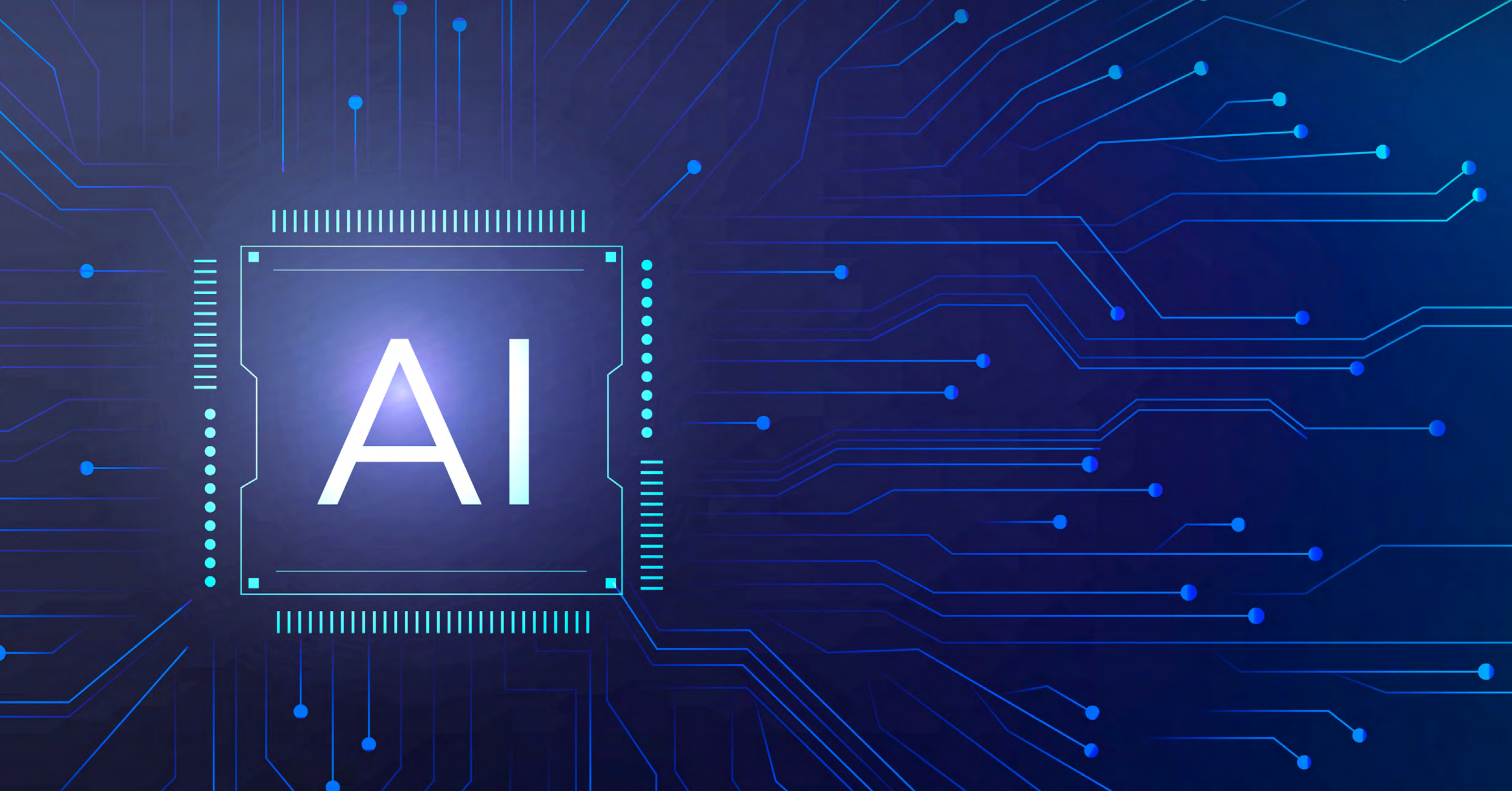
- Autonomous Cyber Operations and Network Defence: Developing ACD approaches that will function correctly in degraded, denied, and disrupted operational environments.
- AI for Systems Security: Showcasing immediate ACD impact and establishing a risk-baseline with real-world vulnerability discovery.
- Validation and Assurance. Creating robust environments for training and evaluation and detecting and mitigating vulnerabilities in AI models.
- State-of-the-art approaches. Determining how elective the state-of-the-art in AI is for ACD and developing better as required.

AICD's contributions span the validation and assurance of ACD environments and models, the development of novel ACD approaches for degraded and denied environments, and new algorithmic breakthroughs in efficient decision making.

Notable recent activities include the creation of enhanced training environments such as CybORG++, which enables the evaluation of AI-driven defensive agents in realistic scenarios, and research into mitigating backdoor vulnerabilities in DRL models. The centre's role across both tracks is imperative to aligning cutting-edge research with operational impact. For example, it supports Track 1 with innovations in cyber sensing and autonomous decision-making, while in Track 2 it provides evaluation frameworks and empirical evidence to test the feasibility and reliability of new solutions in military relevant environments. AICD operates within a collaborative framework, working closely with a range of stakeholders, including government agencies and academic institutions, to ensure its research addresses operational priorities. Its open science model promotes knowledge transfer and capability-building across the wider research and defence communities.

Through its work within ARCD, AICD seeks to provide robust and adaptable solutions that enhance military ACD capabilities and support the operational readiness of the UK's cyber defence framework.

For further information, please contact aicd@turing.ac.uk and our website.



ARCD Track 1 Tasking Overview

The following pages provide an overview of each task contracted on the ARCD programme.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 1: QTSL – Autonomous Decision Making for Cyber Defence [complete]

A 6-month follow-on project aiming to use the Mission Planner tool to remotely conduct vulnerability scans and remediation on Virtual Machines (VMs) that were deliberately defined with vulnerabilities. The project was able to demonstrate remote vulnerability scanning, using OpenVAS, but could not complete the full remediation process within the project timescales. OpenVAS was also found to miss several of the deliberate vulnerabilities and Nessus was recommended for future work in this area.

Task 2: Camulos – Interpretable Causal Models for Cyber AI Agents [complete]

An 18-month research project focused on the development of novel cyber defender RL models that display interpretable behaviour and causal explanations within a reference 'game' simulation environment, such that models' behaviour and explanations can be readily understood by cyber defender operators, operational planners and decision-makers. Overall, it was noted throughout this research that Causal RL is still an embryonic field of research and significant advances in scalability and non-toy applications are required before meaningful application to a cyber security challenge will be realised.

Task 3: University of Kent – Explainable Generic Design, Self-Evolving Intelligent Security Systems for Cyber Attack Detection [complete]

A 12-month study with an aim to develop next-generation intelligent security systems to monitor network activities and categorise threats in real time. The team intend to combine online clustering techniques and prototype-based approaches to construct an ML-based security system to achieve a transparent structure and explainable internal reasoning of decisions made, such that the system can continuously learn new knowledge (through discovering new prototypes representing unseen patterns) from new data but free from the problem of system obesity (namely, the system becomes oversized and less transparent due to more knowledge learned from data).

Task 4: Cambridge Consultants – Generic Persona Classification to Generate Pattern of Life [complete]

This 5-month project successfully developed a proof-of-concept bot (PoLly) with the ability to generate believable and variable network traffic which is tuneable to exhibit different behaviours from different network users. In the foundational phase of the research, the team have successfully shown variation in PoL data generation against each aspect of the bot architecture:

- For different personas: 3 representative military personas were developed and can be tuned.
- For different activities undertaken such as whether an email might contain an attachment or not, or internet use and search behaviour, etc: from the underlying application information.
- For the time spent on, and frequency of activities across, any given day: where an HMM is used to generate time dependent events (based on both the parameters defined by the personas combined with the set of applications).

The network traffic produced by such PoL bots as in this research, could subsequently be used for a training or deception-based use case (i.e., green agent traffic as background noise in AI training, or honeypot). The next phase of this work can be seen under Task 28.

Task 5: Montvieux – Data-Efficient Decision Making [complete]

A 6-month project which addressed the challenge of how to carry out Data-Efficient Decision Making by investigating the use of World Models and Model-Based Reinforcement Learning (MBRL), to:

- Generate synthetic training data to effectively increase data efficiency of decision-making, and
- To reduce the amount of training data required to produce policies in cyber environments.

The authors conclude that 'Accurate world models of abstracted cyber environments [can] be trained and [can] accelerate learning of policy optimisation methods'. Key takeaways include a method for predicting observation transitions in discrete observation spaces that have increased data efficiency in abstract cyber domains. The project has also observed that model-based approaches are limited in simple environments, and further research is needed to understand the full potential of MBRL in cyber domains. Additionally, improving data efficiency of RL algorithms helps the ARCD program by increasing the compatibility of RL to cyber domains.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 6: Frazer-Nash Consultancy with OxBrdgRbtx – Quantum Hybrid approach to Reinforcement Learning to train agents with less data [complete]

A 5-month project where the team challenged themselves with applying quantum technology to the cyber domain (previously unexplored in the literature) to address the problem of insufficient data available to train an autonomous RL agent classically. Their results went some way to proving their overarching hypothesis that 'QML, implemented on a D-WAVE, extracts additional insights from each training piece of data, reducing the data requirements for training such models.' While the research demonstrated some clear emerging trends, a short extension was required to confirm initial conclusions by extending training. These have been written up against Task 26 below.

Task 7: Decision Lab with Actica Consulting – Causal Inference through Neuro-symbolic AI [complete]

In this 6-month task, The Decision Lab team have applied Neuro-symbolic AI to the problem of Causal Inference in the cyber security domain (specifically identifying and responding to malicious insider threats). In conclusion Decision Lab have:

- Implemented an LNN model and use its functionality to explore Judea Pearl's Ladder of Causality, although further iterations for the refinement of the causal diagram would still be required.
- Found that the model performance varied based on the user and the scenario (from performing well to seeing a more variable performance level).
- Demonstrated that the LNN approach (in some cases) can determine the difference between malicious and non-malicious users, despite there being a large overlap in their behaviours.

This project successfully implemented a causal inference solution to a cyber-security task and further explored the Ladder of Causation to tackle the first question. The remaining two were only considered at a high level, to prioritise development of a functional model, therefore this work provides a foundation on which these questions could be investigated further.

Task 8: Decision Lab with Actica Consulting – Data-Efficient Cyber Investigation Determination and Evaluation (DECIDE) [complete]

A 5-month Project where the team successfully demonstrated that:

- AI training data can be utilised to identify high risk components in a model network, and
- The DECIDE tool can evaluate the quality of agent training, distinguishing between agents which successfully and unsuccessfully complete their tasks and the actions taken.

Decision Lab's approach provides human analysts with unbiased explainable rankings for component cyber risk and has the potential to complement the Multi-Criteria Decision Analysis (MCDA) tool used by the Cyber Investigations Management Unit (CMU). The human approach to analysis does not scale effectively, whereas DECIDE is particularly valuable for capabilities that are large or complex, and for attack scenarios with many actions available to the malicious actor. The DECIDE tool has made it possible to identify when agents appear successful in training (i.e. improving their mean average per episode reward) but are not learning to successfully complete their task. During the task, the team found and made significant amendments to the PrimAITE simulator, including introduction of a more random red agent. This resulted in demonstrating the ability of 'MARMOSET' to train a defensive agent.

Task 9: Decision Lab with Actica Consulting – Data Efficient Model Inference (DEMI) [complete]

A 6-month project which successfully showed that the algorithm greatly improved data efficiency.

- DEMI was able to train to attain similar levels of performance as the ATI implementation while using approximately 20% of the iteration.
- Of note is that the DreamerV3 agent adapted and used for DEMI required no hyperparameter optimisation to attain similar levels of performance, despite not being designed for or tested previously in the cyber domain. This has positive implications in terms of transferability and generalisation of agents to diverse settings, without having to devote as much effort to tuning as is often the case with Reinforcement Learning algorithms.

Overall, this work shows promise that this methodology, leveraging the representational capabilities of 'world model' agents, can indeed provide a more data-efficient solution for cyber defence.

Task 10: Decision Lab with Actica Consulting – White-box Models for Explainable AI [complete]

A 5-month Project where the team implemented a Neuro-symbolic AI model called a Logic Tensor Network (LTN), as it combines the high predictive power of neural networks with the inherent explainability of symbolic AI. Using this, the team designed and developed a high-performance, XAI tool around a synthetic malicious insider dataset capturing the activities of 2,000 employees over 17 months. Whilst the model and UI was successfully developed against the XAI objectives, it exhibited poor performance when classifying malicious behaviour. This was concluded to be due to the limitations of the dataset and that to improve performance, increasingly complex data processing and modelling would be required; likely in opposition to the concept of XAI.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 11: University of Kent – Reward shaping for network defence based on reinforcement learning [complete]

Enhancing decision-making in network defence settings using reinforcement learning is the central objective of this project. Twelve-months of research investigating how to improve sample efficiency through developing heuristics for intrusion detection that can be used to define and extract high-quality (dynamic or static) shaping rewards that can easily transfer between environments. This enables them to be used as part of a high-level self-organising meta-controller able to detect new activities and guide the RL process. Reward shaping is a justified method for improving the sample efficiency of RL as it makes the learning process more informed. Beyond reward shaping, the project explores techniques such as graph neural networks, properties of the decision-making scenarios and simulations, factorised representations, Monte Carlo tree search, and activity recognition.

Task 12: Riskaware – Causal Inference for Cyber Security (CAICS) [complete]

During this 8-month project Riskaware have successfully developed a prototype causally-aware model-based Reinforcement Learning (RL) agent that can conduct surgical and precise interventions when defending a network and achieves better performance than causally-blind benchmark scenarios. Over all scenarios, their algorithm achieved better performance than the state-of-the-art benchmarks in these tests. Performance exceeded that of model-free and Neural Network (NN) model-based agents by between 17% and 40%. In addition, network health was also improved by between 12% and 24%.

During this work, the team demonstrated:

- that Causal Inference (CI) can be applied to adversarial, multi-agent cyber-security relevant scenarios to examine the behaviour and dynamics of agents, and
- how causal modelling can aid in creation of trained RL agents by reducing the required interactions with the training environment needed to achieve optimal performance.

Task 13: University of Kent – Machine Learning for Causal Inference [complete]

In this 10-month Project the University of Kent team were challenged with enhancing decision-making in both stationary and dynamic temporal settings using the causal inference framework within machine learning. Their research task focused on online decision-making in cybersecurity, investigating challenges such as confounding variables and adversarial noise influencing system states. Over the course of this project the team:

- Explored the integration of causal inference in decision-making within stationary environments.
- Extended the application of causal inference to decision-making in dynamic temporal settings.
- Investigated methods for obtaining counterfactual information without relying on simulator queries.

Their work on this project extends common machine learning algorithms using the causal inference framework, offering improved decision-making methodologies for cybersecurity. Key achievements on this task include:

- Improvement in performance.
- Proposing a reward prediction model, which can be seamlessly integrated with NCB and NC-UCT, providing crucial causal information without the reliance on resource-intensive simulations.
- Developing the Nested Causal Bandit (NCB) approach, enhancing decision-making in stationary environments and showcasing its superiority over conventional Multi-Arm Bandit (MAB) methods: NCB offers a practical avenue for informed decision making in the changing cyber security landscape, with an observed improvement of up to 75% in cumulative rewards compared to traditional approaches.
- Extending the NCB framework to dynamic temporal settings, introducing the Nested Causal UCT (NC-UCT).
- Proposing a reward prediction model, which can be seamlessly integrated with NCB and NC-UCT, providing crucial causal information without the reliance on resource-intensive simulations.

Task 14: Cambridge Consultants – Mapping Cyber Causes and Effects [complete]

A 6-month project exploring the construction of a general-purpose framework: the Cyber Security Effects (“SFX”) framework, for analysing and modelling security (realistic cyber environments, attacks and defences) in cyber-physical systems with the overall objective to:

- provide greater depth and breadth of attributes and attacks,
- whilst remaining sufficiently abstract and flexible to enable application to a wide range of scenarios,
- thereby allowing autonomous agents to evaluate response and recovery methods to machine speed cyber-attacks.

SFX embeds knowledge about cyber-physical system components and the effects on those components arising from across the full range of attack and defence Tactics, Techniques, and Procedures (TTPs). It can be used to:

- quickly construct models of realistic cyber-physical systems and easily evaluate the applicability and effects of the full breadth of attack and defence TTPs on those models,
- support autonomous agent decisions to be supported by expert knowledge i.e. reasoning about potential causes of observations and/or adverse impacts of actions on the system,
- enable streamline updates as systems, threats, and defences evolve.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 15: Frazer-Nash Consultancy – Knowledge Graphs [complete]

A 5-month research task originally aiming to develop an ontology and conduct a feasibility study into completion of information or deconfliction of states in a knowledge Graph (KG), which is able to explain real time activity through probabilistic output and suggestions of countermeasure, and evolving ontologies in the event of unseen information. The findings initiated a change to the original scope to develop a full implementation of problem-specific Bayesian Networks (BNs) for relevant problems seen in the cyber domain. In summary, considering a network traffic dataset and attack scenarios, the following three separate BNs were constructed to:

- Resolve conflicting or missing information through predicting properties in nodes,
- Predict the next action by a suspicious user, and
- Predict the malicious score of next attack by matching to an established attack pattern.

Task 16: BMT with Mind Foundry – XAI [complete]

In their five months research task BMT and Mind Foundry considered a range of XAI techniques (including Feature Importance, Learning Process & MDP and Policy-Level explanations) which can provide relevant, global or local explanations of autonomous MARL defender agent decisions. They developed a Proof-of-Concept XRL Demonstrator through which cyber–Incident Responders/Managers can gain a sense of scenario history and agent actions from a variety of 'familiar' step-by-step and a post-mortem visualisation. Their evaluation determined how useful these explanations were to cyber stakeholders answering why MARL agents have taken given remedial actions in the event of a cyber-attack. This showed that while participants could use the contextual information to reason why actions were taken, they found it difficult to interpret negative SHAP (SHapley Additive exPlanation) scores and noted wanting a higher-level XAI explanation. Generally, participant response was seen to be more positive when features matched their own mental model of what might influence a remedial action, and as such it became more important to use the XRL information when MARL agents took actions with which they did not agree. The team concluded that explanations must be role based, operationally/mission relevant and timely, and that while this PoC successfully broke ground on how XAI techniques and explanations should be presented and evaluated, further work is required to mature the concept for deployment on real autonomous systems.

The original scope of this work increased significantly as it became apparent that to provide the required contextual information required (the 'why'), the 'What, When, Where and How' is needed to be broken down. An additional and valuable output seen from this work was that XAI explanations such as those provided here are invaluable to software developers for machine learning RL assurance and debugging.

Task 17: BMT with ADSP – Data Efficient RL [complete]

In this 5-month project BMT and ADSP have proven the ability for the Self-Predictive Representations (SPR) meta-RL technique to generalise better to unseen tasks more effectively than the traditional RL technique PPO. Long-Short Term Memory (LSTM) and Attention based networks were also investigated, however do not show this same robustness to instantaneous scenario changes. Additionally, despite this success with SPR, meta-RL techniques are not necessarily more data efficient at retraining on the new task than PPO, i.e. they often require more data than traditional RL techniques to continue retraining on the new scenario.

To meet the aim: to investigate the ability of these Meta-RL techniques to successfully transfer knowledge between tasks without extensive retraining, three hypotheses were tested against 8 experiments on the IPMS RL environment. Recommendations include:

- Conduct testing to see if ARCD agents can adapt to new scenarios, rather than re-training from scratch.
- As meta-RL techniques are not as well tested as more established approaches caution should be taken if before rolling out into a production or near-production environment.
- Several suggested questions have been proposed that should be asked to determine whether a new agent should be trained from scratch or an existing agent should be adapted for a particular task.

Task 18: ADSP – (A Generalist RL Agent for Cyber Defence) [complete]

A 5-month project which successfully implemented and adapted DeepMind's ground-breaking Gato, meeting the objective to demonstrate the creation of a single, versatile agent capable of effectively operating across multiple cyber environments, irrespective of differences in observation and action spaces. This project was not guaranteed to work as it required the implementation of cutting-edge techniques. It paves the way for a consolidation of different approaches into a single unified defensive agent, able to perform on environments with disparate observation and action spaces.

Overall ADSP's Gato agent can achieve expert-level performance across four distinct environments. Each environment was chosen to demonstrate the versatility of the final Gato model (two cyber environments: NASIM, PrimAITE, and two non-cyber: CartPole, Gym-Flipit – which has a continuous observation space unlike the other three environments). In addition to demonstrating that the model performed on par with the expert agents from which it learned in the PrimAITE, Gym-Flipit and CartPole environments, of particular significance is that for the NASIM environment, the Gato agent outperformed the expert agent: completing the task in 5.80 ± 0.96 time steps versus the expert's 9.92 ± 3.09 time steps.

ARCD Track 1 Tasking Overview

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 19: Intellium AI – Inversion Resistant XAI [complete]

A 6-month project addressing the problem of model inversion attacks (reverse engineering the model to infer and reconstruct sensitive training data) on autonomous AI security systems, building a solution that autonomously recovers while preserving usefulness/security of an XAI model. Intellium AI showed that the techniques they implemented and evaluated were ineffective in responding to, but effecting when recovering from, inversion attacks against AI-based ARCD systems trained on tabular cyber network data. The research went on to demonstrate that the use of SHAP values (a representative Explainable AI (XAI) technique) are not able to be used as an attack vector, however it should be noted that this conclusion may not generalise to all XAI techniques.

Intellium AI also provide a system design in section 7.0 (Figure 3) of their final report that shows how their best Recovery mechanism can be incorporated into an AI system. Their contribution to the relatively narrow field of inversion attacks with XAI exploitation includes addressing the impact of this when using the SHAP technique and their own recovery mechanism. This goes beyond the state of the art in the research literature: their combined recovery mechanism (FCS and ANF) was effective at increasing the existing attack model's data reconstruction error even when SHAP values are exploited in the attack.

Task 20: Intellium AI – Mitigating Adversarial Attacks using Explainable Denoising Autoencoders [complete]

A 6-month project demonstrating the potential of Denoising Autoencoders (DAEs) to autonomously detect and recover from hidden adversarial attacks and preventing malicious samples from entering the network. Intellium AI's DAEs also alert the user of the attempted attack (and when it occurred) to enable both a faster response and the knowledge that their system is a target (and potentially breached). DAEs remove perturbations caused by adversarial attacks from incoming traffic by reconstructing clean samples, and reconstructing adversarial samples, back to the original state. With those disturbances gone, detection systems (i.e. the DAE 'plugged in' in front of the NN classifier) can identify malicious traffic and block it, thereby autonomously responding to, and recovering from, malicious threats.

So what? Adversarial attacks (those which subtly change data samples in such a way as to ensure misclassification by AI/ML models) pose a serious threat to autonomous AI cyber systems as they are used as a method to evade AI-based Network Intrusion Detection Systems and conceal malicious activities aimed at undermining the integrity of a network or a device. Using DAEs for anomaly detection and system recovery could be employed in autonomous military assets to provide resilient cyber defence capabilities exceeding current methods. This technique holds significant potential in real-world operational scenarios set in the unpredictable cybersecurity context, where incoming data tends to differ significantly from training datasets.

Task 21: Intellium AI – Reliable Response Through Federated Learning [complete]

A 6-month project collaborating with the centre for cyber security research to demonstrate that the combination of semi-supervised and Federated Learning methods shows the potential to autonomously defend against poisoning attacks, although addressing model poisoning in non-IID datasets remains a crucial research challenge. Additional challenges that require further research and solutions include low-latency responses and limited bandwidth, when creating an autonomous military asset in the real-world. This research targets the challenges of data labelling, such as barriers to data transfer across trust boundaries, or the scarcity of labelled data generated by cyber military assets.

So what? This research takes a step towards enabling future military cyber assets to make autonomous decisions and to address potential threats by learning from their own and similar assets' data, such as network data streams. Creating this "secure by design" concept federated learning system that can autonomously deal with cyber threats has exposed the limitations and challenges that need to be overcome before equipping it to a real-world military asset.

Task 22: Cambridge Consultants – XAI using Policy Dissection [complete]

A 4-month research task tackling black box concerns by using the policy dissection technique (which originally aligns neuron activations with the kinematic attributes of the agent behaviour to produce a desired action through stimulation of a given neuron). In this task Cambridge Consultants provide initial proof to their hypothesis that "[elements of] policy dissection can be integrated within cyber security environments to analyse NN activity, thereby identifying patterns of behaviour [from observations to neuron activations in the first layer]" for use as an XAI tool. Overall, while the technique does not immediately transfer to the cyber security domain, they conclude that it can be used to partially answer two key questions: What does the RL "perceive" of its environment? And what action is taken given the perceived observations?

Over the course of this task the team have shown:

- Extraction of neuron activations throughout the network to facilitate explainability within cyber environments and how to use extracted information to determine the state of the cyber environment as 'seen' by the RL.
- That the importance of different neurons of the NN can be extracted for different actions taken by the RL.
- The potential for using the "Juvenal" network developed, to address challenges in connecting the environment state to resulting actions, using extracted activations in the final hidden layer of the network.
- Further work is required to better understand the decision-making process and automate the recognition of this process to facilitate the development of an XAI monitoring system.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 23: Frazer-Nash Consultancy with Howard Science – Genetic Grammar for Explainability Wrappers [complete]

Typically, SOTA XAI methods, such as LIME and SHAP, focus on providing explainability layers from regression analysis over local data neighbourhoods, and do not show how the combination of input parameters influence the base model predictions or explain why there may be a sudden change in behaviour in certain scenarios resulting in a lack of assurance when defending critical networks. This research aimed to address this by trialling a novel extension to genetic algorithms (GA). The key question to answer was: "Can a powerful extension of the genetic algorithm methodology provide a useful explainability layer for a complex black box model, across a wide spread of the input space?"

While it was found that the GA approach can replicate various simple use cases and that for these simple cases, the GA can be turned into a helpful explainability layer; the methodology used in this research is unsuitable, without further development, for problems with several valid outputs for a single input set. This impacts its suitability for probabilistic models, such as reinforcement learning algorithms such as Proximal Policy Optimisation (PPO), or problems with multiple valid null actions.

Task 24: Advai – A Purposefully Uncertain Recommender Agent for Cyber Defence [complete]

Six months of research tackling the challenge that, as the size of networks grow and attacks become more sophisticated, it becomes more challenging for human defenders to effectively defend against cyber-attacks. RL is a typical choice for cyber defence agents; however, this task explores two challenges with the RL approach:

- Limited interactions with a real-world environment leading to limited quantity and diversity of training data.
- Lack of robustness and resilience to small changes in the cyber environment.

The Possibilistic Q-Learning (PQL) proof-of-concept recommends the most appropriate defensive actions to the human user. PQL is an adaption of Possibility Theory and uses uncertainty quantification techniques which allow the RL agent to be aware of its own knowledge of the cyber network. PQL outperformed existing SOTA strategies by efficiently recommending defensive actions (particularly effectively under high-risk conditions with numerous critical nodes). This approach has potential to increase availability by judiciously responding to threats, but it also minimizes the need for cyber specialists to analyse recommended actions. Additionally, PQL ensures fast, consistent response times and reliable recovery from attacks, which is crucial during rapid cyberattacks. Moreover, PQL agents are scalable, capable of defending networks with thousands of nodes.

Task 25: BAE – XAI Model Architecture Using Theory of Mind [complete]

A 6-month project where BAE evaluated the suitability of models inspired by a branch of psychology named Theory of Mind, their goal being to find ToMnets capable of characterizing the behaviours that emerge during interactions between different types of Blue cyber-defence and Red cyber-attacking agents across a range of network topologies. The team demonstrate attack prediction plots showing the predictions over what an observed agent(s) within the Yawning Titan cyber defence environment will do next and how this evolves over time. Their approach involved making predictions over the goals, intentions, and character of the observed agent(s) within a given environment (without requiring any insight into their internal mechanisms), based on access to past observations (e.g., previous episodes and the trajectory that led to the current time-step). The aims of the research are to:

- Allow humans to better understand the decisions made by cyber security agents.
- Improve the decision making of cyber security agents through granting them the ability to reason over the beliefs, goals, and desires of Red RL agents.

Task 26: Frazer-Nash Consultancy with OxBrdgRbtx – Quantum Extension [complete]

A 3-month extension to Task 6, focusing on three primary goals to enable confirmation of the emerging conclusions:

- Fuller training of the 32 node DBM to allow for better comparison (Only 65K~ steps done).
- Optimisation of D-WAVE through parallelisation to reduce QPU cost.
- Test a range of larger networks (was shown that they could be added to D-wave with little cost).

The Task 6 and 26 research is able to now conclude that:

- A quantum-hybrid approach can provide improved learning per data point, when compared to an equivalent classical approach with an identical sized network. Seen for a range of DBM sizes (32 to 256 hidden units) and for a larger network (10 to 24 nodes).
- The apparent improvement provided by a quantum-hybrid approach is not as significant as the improvement provided by choosing a more complex, cutting-edge algorithm solved with a much larger neural network.
- For the D-WAVE Advantage system, there is a limited increase in solving time for larger problems. On current hardware, a model with 256 units can be effectively used for reinforcement learning, which is sufficient size for a range of problems.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 27: Frazer-Nash Consultancy with OxBrdgRbtx – Novel Application of Quantum Machine Learning to Maximise the Value of Training Data (PPO application) [complete]

In Tasks 6 and 26, the team applied Quantum technology to the cyber domain to address the problem of insufficient data available to train autonomous RL agents classically. Early research in this field indicated that QML offers a route to training ML models with less data due to the increased expressivity of qubits. The team applied this theory and showed their hypothesis that 'QML, implemented on a D-WAVE quantum annealer device, extracts additional insights from each piece of data, reducing the data requirements for training such models compared to equivalent classical methods.' Moreover, it was apparent that RL algorithm choice is key, initiating a second phase (conducted in this task: 27) to understand whether a quantum-hybrid approach implemented using the PPO algorithm shows further benefit.

The DBM approach was integrated into a PPO agent by replacing its neural networks, to investigate whether the benefits of SOTA algorithms and leading python libraries show additional data efficiency gains beyond those already demonstrated, also allowing for a direct comparison of DBMs to neural networks. PPO uses a neural network to train its value function and policy function. Therefore, to isolate the potential benefit seen using a DBM, each network was replaced both individually and together.

Early results showed successful integration of the quantum hybrid approach into PPO enabling a 50% reduction in the data required to train this agent in comparison to classical methods. Additionally, further experiments suggest that the efficiency of a DBM increases as the size of the action space increases, and that using current hardware, near real world problems can be addressed, with this gap expected to close with the next hardware release from D-WAVE. Additionally further experiments suggest that the efficiency of a DBM increases as the size of the action space increases, and that using current hardware this near real world problems can be addressed, with this gap expected to close with the next hardware release from D WAVE.

Task 28: Cambridge Consultants – Co-ordination of multiple PoL emulation bots [complete]

This task further developed the PoL research from a single bot, in Task 4, to multiple interacting bots. Overall, this work has enhanced the TRL of the HMM based PoL bots and increased the available features. Additional work is required to mature this and evaluate the levels of realism in comparison to real-world patterns of life the framework used should provide the flexibility required to achieve this. This phase was centred around two key research areas:

- Reaction to stimulus to enable multi-bot communications – Developing the existing bot to be able to realistically coordinate network traffic between several deployed bots.
- Generative modelling research – Investigating the potential to use generative modelling within the bot, either for content creation or determining the actions of a bot.

These developments required an updated HMM architecture (extended to include status, activity and task) to increase event detail and to facilitate stimulus and communication triggers between bots. This allowed bots to mimic real-life behaviours through reacting to external inputs and autonomously coordinate complex activities, such as attending meetings, responding to emails and going to lunch. Additionally, real-time events and content creation enabled the bots to perform coordinated network traffic, further increasing model realism and varied network traffic. This included a Python-based VoIP system, emulating the Mumble application (featuring human interaction patterns such as speech, silence and overlapping conversations), and a LLM to provide 'realistic' content in email communications between bots. Results on the content creation aspects, showed that LLMs could be prompted to generate content which would pass lightweight examination as an initial email (first of a chain). However, the quality of subsequent emails rapidly degrades.

Generative modelling was also investigated as an alternative to PoL agent creation. It was shown that although LLMs can be used to create PoL agents, significant limitations are present. Most notably the latency from the LLMs was almost prohibitively slow due to the amount of information required to provide the context and history of the agent. Consequently, with current technology an LLM-agents-based simulation is not currently capable of running anywhere near real-time.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 29: University of Exeter – Quantum machine learning for complex networks [ongoing]

Studies of complex networks encompass a wide class of graph-based problems, which typically emerge in telecommunication and information networks. Understanding inherent properties of these networks and their inherent correlations, is vital for identifying security risks. At increasing network (graph) size the complexity of calculating topological properties makes their estimation computationally infeasible on classical processors. Here, distinct operation principles of quantum computers can provide an advantage for graph-based problems, and potentially address their dynamics.

In this 12-month task the University of Exeter will develop quantum solvers for estimating properties of graphs, embedding them as quantum Hamiltonians and studying their spectra and ground state degeneracy. Leveraging the powerful tool of geometric quantum machine learning, they will address the challenge of identifying anomalies based on topological features. Choosing suitable use-cases in collaboration with defence and crypto groups, they will apply these methods in silico and test the demonstrator as quantum network analyser with largely increased capabilities. The resulting quantum software will lay the background for future-proof complex network analysis.

Task 30: Roke – Multi-Objective Reinforcement Learning for Cyber Defence [complete]

Reinforcement Learning (RL) is a popular method by which autonomous decision making can be achieved. Such algorithms only focus on a single objective: Single-Objective RL (e.g. minimizing the intrusion of red agents on the network) limiting a defensive agent's resilience in the face of many competing and evolving objectives. Conflicting objectives, such as restoring a machine from a back-up image, must be carefully balanced with the cost of associated down-time, or the disruption to network traffic or services that might result.

As a result, Roke's research considered a multi-objective network defence game, requiring consideration of both defending the network against red-agents and maintaining critical functionality of green-agents. Two Multi-Objective Reinforcement Learning (MORL) algorithms, (Multi-Objective Proximal Policy Optimization (MOPPO), and Pareto-Conditioned Networks (PCN)), are used to create two trained Autonomous Cyber Defence (ACD) agents whose performance is compared on the CybORG gym, Cage Challenge 2 scenario (adapted to enable a Multi-Objective Cyber Defence game).

Task 30: Roke – Multi-Objective Reinforcement Learning for Cyber Defence [complete]

Reinforcement Learning (RL) is a popular method by which autonomous decision making can be achieved. Such algorithms only focus on a single objective: Single-Objective RL (e.g. minimizing the intrusion of red agents on the network) limiting a defensive agent's resilience in the face of many competing and evolving objectives. Conflicting objectives, such as restoring a machine from a back-up image, must be carefully balanced with the cost of associated down-time, or the disruption to network traffic or services that might result.

As a result, Roke's research considered a multi-objective network defence game, requiring consideration of both defending the network against red-agents and maintaining critical functionality of green-agents. Two Multi-Objective Reinforcement Learning (MORL) algorithms, (Multi-Objective Proximal Policy Optimization (MOPPO), and Pareto-Conditioned Networks (PCN)), are used to create two trained Autonomous Cyber Defence (ACD) agents whose performance is compared on the CybORG gym, Cage Challenge 2 scenario (adapted to enable a Multi-Objective Cyber Defence game).

Task 31: Smith Institute – Information Asymmetry Exploitation [complete]

This research task (5 months) involves the development of a cyber-defence agent based on discrete optimisation that can deploy deceptive tactics to delay and entrap adversaries. Smith Institute will use deceptive elements such as decoys or honeypots. This will be achieved by drawing on techniques from network-based optimisation. This approach explicitly uses the network topology to the defender's advantage. It will also be more adaptable to new situations as the algorithm will not require re-training for a new network or attack scenario, unlike reinforcement learning (RL) techniques for example.

In benchmark scenarios of increasing network and attack complexity, their work will demonstrate the power of optimisation-based methodology to deliver generalizable and scalable cyber defence. This will open the door to new hybrid strategies able to cope with the evolving cyber landscape.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 32: Advai – Continued Development and Validation of the Purposefully Uncertain Recommender Agent for Cyber Defence [complete]

Task 32 extends Task 24's PoC agent, through experiments designed to determine whether the PQL agent, inclusive of a new observation wrapper and prior knowledge, can withstand more challenging scenarios (e.g. red agent numbers and strategies, defence strategy, network laydown and numbers of critical nodes). Results, gained from a 35-node version of the PrimAITE environment, show with limited training the PQL algorithm can handle up to 118 attacks per episode (representing machine speed attacks), learn an optimal policy, and outperform the PPO agent by 22 times the average reward per episode. PPO uses an actor-critic framework which leads to a large variance in the learning process. Therefore unsurprisingly, PPO can perform correct actions however it struggles to learn an optimal policy. In comparison the PQL agent can deal with the substantial challenge presented by scenarios with large action and observation spaces.

Overall, results show PQL can converge to a good policy with very limited training, while current SOTA methods like PPO struggle without significantly more training episodes. This result demonstrates the increased training efficiency and scenario and network scale to which such new approaches might be applied. It could also serve to increase the time a network can operate normally by appropriately responding to threats, targeting the growing challenge for human defenders to effectively and efficiently defend networks against increasingly sophisticated cyber-attacks.

Task 33: Tulpa – LLM Translation Layer for C2 Multi-Agent Cyber Defence [complete]

Tulpa demonstrated that a Large Language Model (LLM)-based multi-agent system can successfully defend a network in a cyber simulation. This conclusion is subject to the following key findings:

- An LLM agent can defend the network as well as, if not better than, a state-of-the-art Reinforcement Learning (RL) algorithm: benchmark experiments yield an average reward score of –125.4 (LLM) in comparison to –273 (RL), where the less negative a score the more successful the defence agent).
- However, LLM reasoning often results in unpredictable or inappropriate actions (likely as a result of bias to repeated words or those holding certain positions within a prompt, hallucinating actions, etc) thereby limiting their application as defence agents in high-risk autonomous cyber defence environments.

Tulpa assigned a group of LLMs to different roles within a team: Logging, Detection, Prediction, Critic, Action and Green Advocate agents. The best performing compositions tested were those with 5-6 constituent LLM agents, and from adaptive teams (i.e. those which periodically 'lost' access to one of their agents) which demonstrated a greater variety of actions leading to better results potentially due to the reduction of bias, and the increase of specific task focus.

Task 34: BT – Project Orchid [complete]

In this research BT investigated how LLMs can be used to design a Hox gene based Evolutionary Algorithms (EA) that in turn control a novel compressed Transformer Neural Network. This work addressed the challenge that; whilst Generative AI has proven to be a powerful technique, the compute and memory costs are a barrier to cyber security and defence deployment. This novel approach to designing a more efficient Transformed model that is inspired by complex biological lifeforms utilise the regulatory function of Hox gene sequences to dictate cell placement and neural function during morphogenesis. The proposed model was then evaluated for use in the cyber domain for clustering of threat intelligence data to demonstrate a PoC and measure model performance compared to existing techniques.

The ORCHID tool was developed in two parallel packages to allow for the test and evaluation of both the EA and LLM components. To evaluate the tool an EU dataset of global cyber-attacks was used which contains a comprehensive, interdisciplinary, and continuously updated database of cyber incidents worldwide. First, the Hox gene EA was compared against a base model to which it outperformed by 10-20% in testing. The Hox gene model includes a novel cellular automaton based genetic masking function that enhanced the EA to achieve the measured performance. Evaluation was performed for different mutation rates of the Hox gene EA to measure the model fitness – a measure of optimality of a solution for the target problem. Second the Hox gene EA was compared against the K-means clustering algorithm, using traditional approaches to defining the number of data cluster, for analysing the threat data and grouping based on the reported attack type. It was reported that ORCHID achieved a peak improvement of ~15-20% in the performance of the tuned LLM to create threat clusters from raw cyber threat data.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 35: Smith Institute – Decision Transformers for Context-Aware Cyber Defence Agents [complete]

Smith Institute highlight the requirement within cyber defence for multi-step causal reasoning, network awareness, and multi-modal data ingestion. Their 5-month research task will create a new kind of autonomous cyber defence agent architecture (the Decision Graphormer) which provides an alternative approach to traditionally chosen RL based solutions. This integrates the decision transformer (for causal decision making), capable of leveraging multi-modal data, reasoning over distant time points, and generalizing effectively; and the recently published Graphormer architecture (for static graph-based interpretation), where the positional encoding of the Graphormer provides agents explicit, immediate awareness of underlying network topology.

Smith Institute note that: Success will disrupt the cyber-AI status quo. Naïve RL fails to solve the problem of needing to parse globally the available data, agent behaviour and network topology, and must inevitably hit a performance ceiling. If successful, Decision Graphormers will open a new paradigm, with power through long-range thinking in space and time, transferability without retraining, and the ability to consume multi-modal data as different applications demand.

Task 36: ADSP – Decoy Agents – A Generative Approach to Deception [complete]

Hypothesis: LLMs can be effectively utilised to create advanced and convincing decoy entities that are indistinguishable from real system components to cyber adversaries.

In this 4-month task, ADSP plan to explore the efficacy of LLM's (proven to be extremely capable of writing code to fulfil a given set of instructions, across a range of tasks) to create realistic decoys on-the-fly to deceive attackers and deflect from the intended target (servers, PLCs, etc.). Specifically, this means leveraging an LLM to produce data (either directly, or via writing a script) that is indistinguishable from the data produced by the 'real' component, so that attackers are convinced enough to laterally move to (or attack) a 'fake' node, rather than the intended target. A feedback loop will refine the decoys based on attackers' interactions, continuously enhancing their believability.

The expected outcome is a robust, self-evolving deception framework. It is anticipated to challenge even the most sophisticated attackers, offering a fresh and cost-effective strategy for cyber defence, aligning with generative deception model objectives and staying ahead of adversarial tactics. By programming LLMs to autonomously design, deploy, and adapt decoys, reliance on manual effort and expertise is reduced.

Task 37: ADSP – LLM Agents for Cyber Defence: A Zero-Shot Approach [complete]

ADSP's research investigated to what extent LLMs might be used as a defence agent to reduce reliance on traditional, labour-intensive training methods (such as Reinforcement Learning (RL)). They used a zero-shot approach, posing the hypothesis that LLMs possess sufficient contextual knowledge to generate informed actions based on natural language descriptions of cyber scenarios. They were able to demonstrate that their Azure OpenAI GPT 4 agent defends the network with a 90%-win rate on the IPMS environment and a 100%-win rate on the PrimAITE environment.

Zero Shot is an ML approach where a model handles unseen scenarios without needing specific training. A Zero Shot capability works by leveraging known contextual information that is related to the unknown scenario.

Key findings from this research include:

- New Environments Integration: LLM agents demonstrated the ability to be rapidly deployed in new environments.
- Memory Module Integration: The integration of a memory module allowed LLM agents to reference past strategies and apply them to future scenarios.
- Explanation Module Integration: The development of an explanation mechanism provided insights into the agents' decision-making process.
- Adaptability Testing: Adaptability tests revealed that while LLM agents could maintain performance in slightly altered environments, their success was contingent on the relevance of their pre-existing knowledge.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 38: ADSP – Talk-To-Your-Components: Human Programming Interfaces [complete]

In this 3-month task, ADSP will deploy LLMs equipped with retrieval augmented generation (RAG) to interpret the complex and voluminous cyber security data into human-readable output. The RAG approach enables the LLM to dynamically retrieve information from relevant documentation, guidelines, and examples to provide context-aware responses. This process transforms the otherwise opaque data into clear, concise summaries and insights that human analysts can easily understand. The particular novelty of this project is the intention to show how the LLM can read documentation from relevant cyber-APIs and learn how to take appropriate action in a given circumstance, in the same way that a human could.

The core hypothesis of the project proposal is that Large Language Models (LLMs) can take machine-oriented information such as logs or packet data, convert it into a format that is easily digestible for humans, and interact with cyber-APIs simply by reading the relevant documentation for the API, to take appropriate action.

The outcome is the production of human-readable information that is easily comprehensible and enables cyber security professionals to swiftly interpret logs and other packet information, as well as allowing the LLM to take actions by reading relevant cyber-API documentation, significantly reducing response times and improving the overall efficacy of cyber defence measures.

Task 38: ADSP – Talk-To-Your-Components: Human Programming Interfaces [complete]

In this 3-month task, ADSP will deploy LLMs equipped with retrieval augmented generation (RAG) to interpret the complex and voluminous cyber security data into human-readable output. The RAG approach enables the LLM to dynamically retrieve information from relevant documentation, guidelines, and examples to provide context-aware responses. This process transforms the otherwise opaque data into clear, concise summaries and insights that human analysts can easily understand. The particular novelty of this project is the intention to show how the LLM can read documentation from relevant cyber-APIs and learn how to take appropriate action in a given circumstance, in the same way that a human could.

The core hypothesis of the project proposal is that Large Language Models (LLMs) can take machine-oriented information such as logs or packet data, convert it into a format that is easily digestible for humans, and interact with cyber-APIs simply by reading the relevant documentation for the API, to take appropriate action.

The outcome is the production of human-readable information that is easily comprehensible and enables cyber security professionals to swiftly interpret logs and other packet information, as well as allowing the LLM to take actions by reading relevant cyber-API documentation, significantly reducing response times and improving the overall efficacy of cyber defence measures.

Task 39: BAE – Theory of Mind as Belief Models for Reinforcement Learning Agents [complete]

BAE continued their work on determining the motives and behaviours of autonomous agents given increasingly such agents are being trusted with performing critical tasks. In their previous phase (Task 25), BAE show that ToM can be used to predict characteristics, attack paths, and target users of a hostile Red agent. In this project, the team used these outputs to inform RL agents with predictions on their adversary. Their goal was to assess if using a belief model in this manner could improve the performance of cyber defence agents thereby realising the creation of a novel methodology that allows network agnostic defence strategies that can anticipate different adversaries on networks of various different sizes and topologies.

By repurposing their successor representation (SR) evaluation metric from phase 1 as a loss function to optimize SR predictions directly, and by reviewing the architecture of GIGO-ToM's, significantly more accurate predictions of the behaviours and objectives of previously unseen Red agents were seen. BAE were also able to demonstrate that this performance is robust to network topologies consisting of up to 250 nodes.

Task 40: ADSP – Generalised RL Agent Wrapper Extension [complete]

ADSP successfully show that it is possible to create adaptors which remove the need for duplicated work in RL projects. They also demonstrate examples of leveraging the power of adaptors to provide new solutions to existing problems (for example, investigating how to create a central analysis platform that is agnostic to the underlying RL libraries used to train and run models). This work could be used to lay the foundations of a new way of approaching RL work within ARCD, serving to reduce the time spent on agent and environment integration work and free up development teams to focus more effort on solving new problems.

ADSP achieved their three core objectives:

- Creation of an Adaptor Framework: allowing the training, running and evaluation of RL agents through a unified UI system, to which developers could add their own environments. This package allows interfacing with environments and agents other than just those that feature the ARCD Common Action and Observation Space (CAOS) interface.
- Adaptor Creation Using an LLM: the adaptor framework was extended to enable the use of an LLM to assist in writing adaptors, streamline the new environment onboarding process for users of the adaptor framework.
- Creation of an Agent Choice Engine: a further extension of the adaptor framework to add the Agent Choice Engine, which can select an agent from a suite of agents at each runtime timestep, making use of known metadata about each agent in the suite during decision making.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 41: Cambridge Consultants – Juvenal Network Phase 2 [complete]

In a second phase of work, following on from Task 22, Cambridge Consultants use the policy dissection technique (extraction of neuron activations in an agent) combined with the introduction of an auxiliary neural network (the Juvenal network) to provide a more detailed classification of behaviour, thus allowing explanations along the lines of: "Update ACL for Node 3 in the cyber network due to a suspected denial of service attack. Observation: very high traffic on Link D."

This proof of concept is designed to show how an agent can have explainability and performance built in without one significantly compromising the other, and how this explainability can be introduced in the training phase of the agent in order for this to be achieved. The work concludes that incorporating explainability as a target during training can be achieved without loss to agent performance. Cambridge Consultants achieve the creation of a hierarchical approach that introduces an AI responsible for explanation, separate from the AI responsible for action, and provides a useful general method for explainability.

Task 42: University of Exeter – Biological AI – Neuron Soup [ongoing]

The University of Exeter team will take neurons from pluripotent stem cells and place these on high density multi-electrode arrays (HDMEA) in a clean room environment. These are then fed data via MXWServer software in the form of neuron stimulation. The neurons are 'rewarded' (e.g. with heat and light) or 'punished' (e.g. with cold and darkness) according to their response. This method will be applied to cybersecurity, through proposing appropriate solutions within the Yawning Titan and/or PrimAITE environments. The system would autonomously learn to 'bounce' the ball of the attacker just as with a game of Pong, to demonstrate the proof of concept.

Task 43: BT – Kraken [ongoing]

In a 4-month project, BT's research will explore:

- Distributed neural architectures: What topology of intern-neural network is optimal, how much overhead will it place on the underlying mesh communication network?
- Resilient neural topologies: Which topologies are most resilient to targeted attack versus A.I performance? Previous work in our team has demonstrated significant differences in resilience arise depending on the topology, e.g. scale-free, random, or small-world networks.
- What algorithms can be reused from the swarm robotics domain in the cyber defence context? One positive aspect of borrowing from this domain, is the computing hardware was always assumed to be constrained, hence the algorithms are energy efficient.

As biological neurons are inherently dynamic with high levels of plasticity, BT intend to mimic this process to create self-healing functionality in a cyber security context. In biological terms they will compare this to the neural networks (NN) within some marine species, e.g. comb jelly fish or the distributed nervous system of Octopi, which demonstrate extreme resilience to damage and loss of function.

The aim of BTs project is to create an adaptive self-organising A.I system, which can provide intelligent cyber defence services, while under sustained cyber-attacks on itself. Plus, create highly resilient distributed neural networks, supporting tactical mesh cyber defence platforms and data services.

Task 44: Riskaware – Hierarchal Deep RL For Self-Healing and Traffic Management [ongoing]

Riskaware will use Hierarchical Deep Reinforcement Learning (H-DRL) to respond to cyber-attacks, with a focus on both taking recovery actions whilst maintaining the legitimate/healthy network. They aim to answer whether a H-RDL agent can be trained to provide a multi-action response and restoration strategy with consideration for a healthy network. Temporal abstraction will be used to support the multi-action response, which has yet to be demonstrated with the multi-objective considerations. The project will further introduce challenges associated with dealing with MANET type networks, adding complexity to the agents understanding of a network.

Task 45: Frazer-Nash Consultancy with Neurosynapse – Vector Symbolic Architectures for Efficient Network Defence and Recover [ongoing]

In this 5-month project, FNC will investigate the potential benefits VSA can have on automated cyber defence. To date cutting-edge research has shown that VSA is at the point where the various cognitive steps of sensing, understanding, planning, acting, and learning are starting to be represented in a novel unified form that is both compact and efficient. However, they hypothesise that these benefits will enable autonomous VSA agents to be developed that can adapt to cyber threats in a dynamic network topology, alongside offering reduced latency for faster responses and low enough power requirements to be easily distributed across a network of edge devices. Key research questions include:

- Is VSA suitable for representing network topology as a cognitive map?
  - APPENDIX B - Can VSA encode cyber-attack observations and actions?
  - APPENDIX C - Can autonomous VSA agents be created such that they are capable of generating respond and recover actions in network scenarios with malicious actors.
  - APPENDIX D - Can these agents continue to operate effectively under dynamic network topologies?

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 46: University of Surrey – Digital Twin and 5G [ongoing]

The University of Surrey propose to leverage Digital Twin technology within their existing full stack O-RAN 5G testbed and RL environment. They intend to create their own virtual replicas of critical systems, within which they will integrate AI-driven analytics and Reinforcement Learning, to predict attack trajectories and dynamically adapt responses at real-time. They will fully capture all operational states of the system, including uncertainties, asset lifecycles, and cyberinteractions, providing a more comprehensive and dynamic analysis that enhances the robustness and security of such systems against cyber threats.

Key research questions for this work include:

- How can the DT framework be integrated with AI-driven analytics and RL algorithms for real-time threat prediction?
- How can RL be used to dynamically adapt security measures in real-time?
- What are the performance metrics and benchmarks for evaluating the effectiveness of the integrated AI and DT system?
- How effective is the integrated DT and AI system in real-world 5G O-RAN environments?
- How does the system perform under different types of multi-stage cyberattacks?

Task 48: UKAeris – Enhanced Spiking Neural Networks (eSNNs): Overcoming Training and Scalability Challenges with Hybrid Techniques [ongoing]

Deployed IT often faces limited computational resources and insufficient connectivity to fixed bases, hindering the use of networked computational resources. Quick decision-making is crucial, especially in resource-constrained platforms like small UAVs. This project investigates the advantages of using enhanced Spiking Neural Network (eSNN) powered agents in these environments over traditional Artificial Neural Network (ANN) powered agents. UKAeris will develop defensive cyber agents using the CybORG environment to solve the CAGE challenge 2, which simulates complex cyber-attack scenarios. This involves creating agents that can autonomously detect, respond to, and mitigate cyber threats in real-time. Agents trained with ANNs and SNNs will be compared, to evaluate their energy expenditure, efficiency, and performance. This approach will enable a thorough assessment of the potential benefits of eSNNs, particularly their ability to enhance decision-making speed and resilience in dynamic and constrained cybersecurity contexts.

Task 49: ADSP – Probabilistic Graphical Models for Agent Planning [ongoing]

This project aims to implement a probabilistic graphical model (PGM) as a blue agent in a cyber-security context. Bayesian Networks have previously been applied to cyber security applications to predict the next attacking agent action. Rather than predicting the next action that a red agent is likely to take, ADSP hypothesise that they can use the forward inference ability of PGMs to predict the outcomes of potential action of the blue agent. This will allow the agent to select actions with a higher likelihood of success and anticipate the outcomes leading to planning capability. ADSP intend to use an RL environment to train the PGM.

Task 50: ADSP – Generalised RL Agent Adaptor Extension [ongoing]

This project aims to extend and develop the existing proof-of-concept agent and environment adaptors into a beta release package. This will enable other groups working on ARCD RL projects to utilise the adaptor functionality. ADSP will collaborate with Advai to evaluate the usability of the adaptors for their upcoming project and identify additional features and priorities. Working on these additional features and incorporating regular feedback, ADSP will maintain flexibility in ensuring features may be prioritised as needed. Through this phase, ADSP aim to finalise the software and prepare it for a production-level release.

U75A HIGH RISK AND DISRUPTIVE OPTIONS

Task 51: ROKE – Autonomous Pareto-Optimal Policy Selection for Cyber Defence [ongoing]

Real-world decision-making problems often have multiple objectives, some of which conflict, or whose priorities change over time. Examples include networking within aircraft to maintain flight systems, or the collaboration between drones in an ISR activity. Autonomous Cyber Defence (ACD) agents aim to maintain the integrity of these networks, whilst preserving mission critical functionality.

Roke demonstrated in their previous work (Task 30) that ACD can be treated as a multi-objective problem and evaluated the performance of defensive agents based on state-of-the-art multi-objective reinforcement learning (MORL) algorithms (see Summary of Previous Work below). The result of this training is not a single RL agent, but a collection of agents, that approximate a set of optimal policies (on the Pareto Front, PF). However, the choice of policy depends on dynamic variables such as mission priorities, user preferences, and the state of environment. Roke proposes to address two key points in this project reflecting ARCD requirements:

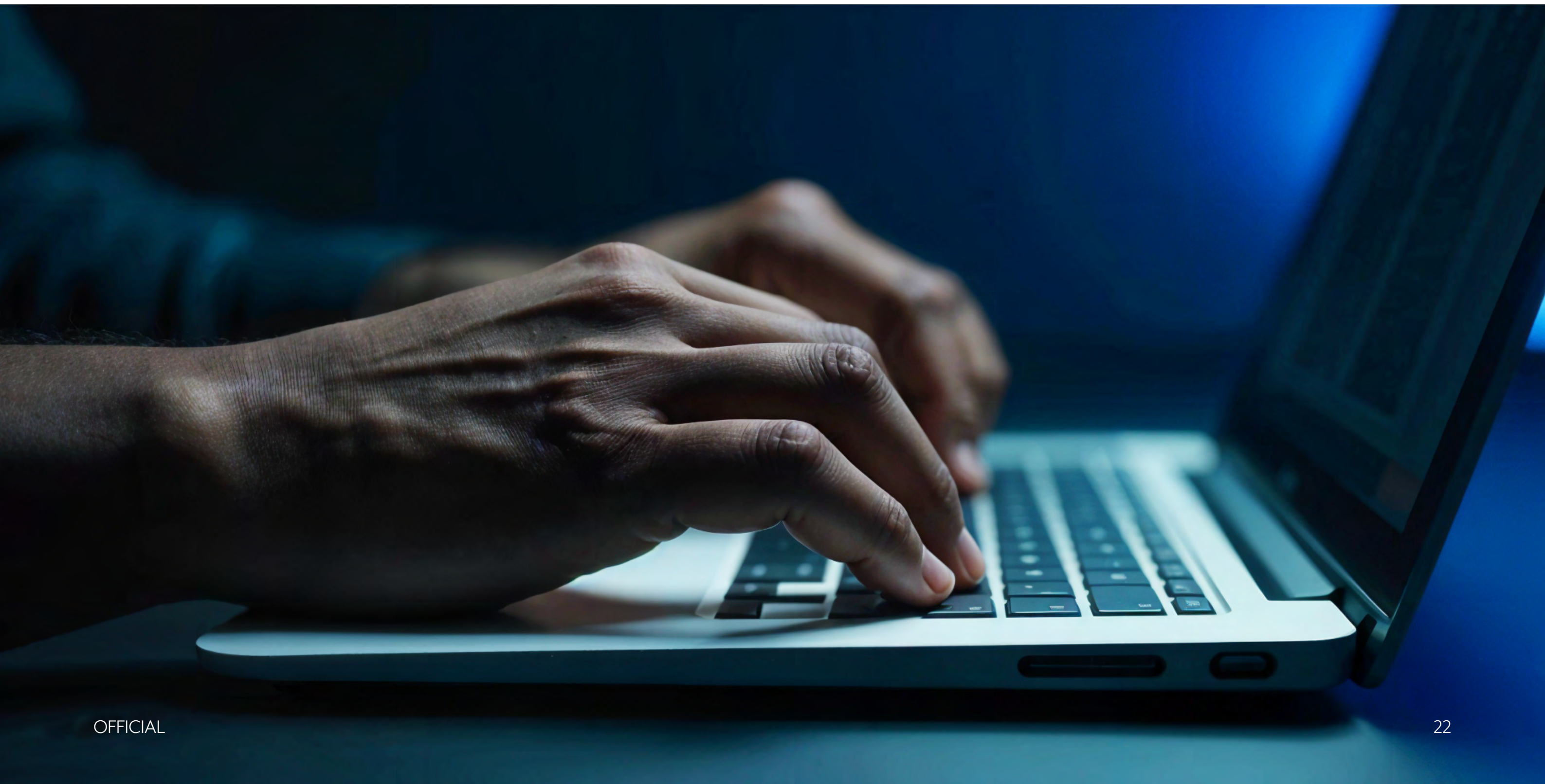
- Algorithmic dynamic selection of optimal policy based on inputs such as operational objectives, sensor inputs, intelligence, historical performance, or anticipated future performance.
- A demonstration of MORL in a richer simulation environment, featuring a diverse set of objectives such as more complex green functionality requirements available in recent PrimAITE scenarios, and more sophisticated red agents.

Task 52: Advai – Continued Development and Validation of the Purposefully Uncertain Recommender Agent for Cyber Defence [ongoing]

This task aims to continue development of Advai's Possibilistic Q-Learning (PQL) Agent (see tasks 24 and 32) and demonstrate their approach on more sophisticated cyber scenarios and realistic red agents within PrimAITE v3.3.

The objective of this project is to continue the maturation of the PQL agent to handle more sophisticated cyber simulations in higher-fidelity environments with more realistic threat models. This will be achieved by improving its ability to respond dynamically to evolving situations by utilising multiple agents, integrating real-time expert knowledge, and expanding its transferability capabilities. Increasing complexity will likely increase the dimensionality of the decision problem, for which various function approximation techniques that can model complex dynamics will be considered.

Finally, to increase the speed of iteration, training and testing situations will be aligned more closely with real cyber-defence missions by increasing cyber deployment realism in the network, red agents, and rewards.



ARCD Track 1 Tasking Overview

U75B INFORMED CYBER SENSING

Task 1: Riskaware – CACTI [complete]

The Critical Asset Cyber Terrain Identification (CACTI) project undertook a data-science led investigation of the applicability of selected techniques to identify critical assets within a network. This covered the use of GNNs and related technologies such as Relational-GNNs and Graph Attention Networks (GANs), to develop a methodology to create a prototype tool to address the requirement for identification of critical devices within a newly encountered network. This covered the use of GNNs and related technologies such as Relational-GNNs and Graph Attention Networks (GANs), to develop a methodology to create a prototype tool to address the requirement for identification of critical devices within a newly encountered network. This work was extended through Tasks 16 and 19 to provide a dynamic view on support of the Intelligence Preparation of the (cyber) environment.

Task 2: Frazer-Nash Consultancy – Autonomous Cyber Attribution Follow-On Work – U60 Extension [complete]

Serapis Lot 6 U60 applied machine learning to the problem of autonomously attributing a cyber-attack to a threat actor using machine learning. Frazer-Nash's work focused on using malware analysis tools to generate features that can be used by machine learning algorithms. This task provided an improved understanding of the machine learning models that are applicable to automated cyber attribution and was combined with U75b Task 5 to understand how information from multiple sources, such as malware binaries, network traffic and telemetry, can be used to provide broader coverage of a cyber-attack.

Tasks 2 and 5 were integrated into a Proof-of-Concept JAM (Joint Attribution Model) framework.

Task 3: Frazer-Nash Consultancy – Exploring Behavioural Contexts of Cyber-Attacks [complete]

This project approached the problem of cyber attribution from understanding the behaviour of the attacker, and how attacker behaviours change in relation to the victims of the cyber-attack. This was split into two phases. Phase 1 was a literature review of historic cyber-attack case studies, each involving attackers with different motivations, identifying key features of targeted organisations using the COM-B model of behaviour to create attacker 'profiles'. Phase 2 identified and compared different methods used to analyse crime data, including directional acyclic graphs (DAGs).

Task 4: Frazer-Nash Consultancy – Research into Behavioural Cyber Attribution Frameworks [complete]

Serapis Lot 6 Task U60 recommended the need for more research into how cyber attribution models contribute to cyber attribution. This work was split into two phases: a literature review into current cyber attribution models and behavioural factors; and in-depth comparative analysis of identified models and their relevant attributional factors, recommending how different stages of threat intelligence could be automated by adding a layer of behavioural factors over existing cyber threat intelligence systems.

Task 5: Montvieux – An attribution framework integrating malware, networking, and telemetry data [complete]

Initial work on Serapis Lot 6 U60 showed that multiple attribution facets (ATT&CK class, nation state, and APT) can be inferred from labelled malware samples using ML techniques. Network traffic and telemetry data was also identified as a potentially rich source of relevant detections. Task 5 trained new baseline models to handle networking and telemetry data and integrated this work with that from Task 2 into a 'Joint Attribution Model (JAM)' framework.

Tasks 2 and 5 were integrated into a Proof-of-Concept JAM (Joint Attribution Model) framework.

Task 6: Montvieux – Cyber attribution fingerprint classification via deep learning [complete]

This task applied techniques from the field of digital imagery forensics to machine speed cyber threat detection and attribution. Creating a proof-of-concept attack fingerprinting classification tool by extracting uni-dimensional time series from network logging data and network packet captures, encoding as multi-channel two-dimensional imagery field datasets with state-of-the-art (SOTA) computer-vision based classifiers to achieve classification on attack typologies.

ARCD Track 1 Tasking Overview

U75B INFORMED CYBER SENSING

Task 7: University of Liverpool – Hidden Network Model: Learning and Attacker Identification [ongoing]

Starting with the premise that behaviour displayed by malicious agents differs measurably from each other. UoL will develop novel methods that are able to estimate the probability that an agent is responsible for a given identified attack. This task comprises of two distinct steps, initially learning a formal model of malicious network behaviour for each potential attacker as a reference point for their future attack behaviour, this work generalises their learning algorithm for the construction of Hidden 1-Counter Markov Models (H1MM).

Step 2 provides an evaluation as to what extent the observed behaviour during exploitation of the network complies with any of these learned models. To deliver this task UoL have recruited an experienced postdoc with expertise in probabilistic models, learning algorithms and implementation of proof-of-concept tools.

Task 8: Raytheon UK – Topological Data Analysis for Network Data [complete]

Due to the complexity of network data and malicious actors obfuscating their activity, network traffic monitoring and intrusion detection are challenging problems. Task 8 developed a proof of concept using Topological Data Analysis (TDA) to summarise data while preserving the high level "shape". TDA's have been shown to offer advantages in anomaly detection, activity clustering and as the basis for activity prediction using network logging data.

Task 8 investigated using Machine Learning (ML) models with persistence images to gain further insights such as classifying anomalies and showed some really promising initial results.

Task 9: Montvieux – Live Validation of Prototype Cyber Attribution [complete]

There is a portfolio of Open Source and Machine Learning (ML) tools which are able to observe and understand different aspects of threat actions, specifically understanding Tactics, Techniques and Procedures (TTPs) being employed by adversaries, and potentially who they are. Building on Task 5 this task implemented an instrumented an online sandbox environment to support more representative tool testing (and refinement). The sandbox was opened to the internet as a honeypot collecting 'real' data. Montvieux collected three different datasets, which have been obfuscated to remove any personal information, and which are available for subsequent tasks.

Task 10: Montvieux – Bayesian Networks for Cyber Threat Attribution [complete]

Machine Learning (ML) and Deep Learning (DL) models are inherently static and deterministic once trained, this is at odds with cyber-attacks and the adversaries who perpetrate them. Bayesian Network models are capable of systematically dealing with uncertainty and incomplete observations. This task investigated the application of Bayesian Networks for incorporating noisy and partial observations, as well as human expert prior knowledge, into causal reasoning models of Advanced Persistent Threats (APTs).

This task built a proof-of-concept Bayesian Network model for inferring adversary Tools, Techniques and Procedures (TTPs). This task built a Bayesian Integration Network prototype, demonstrating the potential for Bayesian Networks to form the basis for a centralised attribution framework, integrating multiple heterogeneous data sources.

Task 11: Montvieux – Autonomous Cyber Attribution Model Advancement [complete]

Typically, cyber attribution is used to infer techniques and strategies being employed, the goals motivating an attack and who is responsible. This work is exploring how to uplift technique inference to APT inference via the associations contained within the ATT&CK knowledge base, by generating a dataset linking TTPs to APT Groups, train a Machine Learning (ML) model on this data to infer a probability distribution over APT Groups. This task explored how to uplift technique inference to advanced persistent threat (APT) inference via the associations contained within the ATT&CK knowledge base, by generating a dataset linking TTPs to APT Groups, train a Machine Learning (ML) model on this data to infer a probability distribution over APT Groups.

U75B INFORMED CYBER SENSING

Task 12: Elemendar – Active Learning for continuously improving extraction of CTI entities [complete]

Elemendar's READ tool uses Named Entity Recognition models to process human authored, unstructured (Cyber Threat Intelligence) CTI reports into structured CTI data. Elemendar's READ tool uses Named Entity Recognition models to process human-authored, unstructured Cyber Threat Intelligence (CTI) reports into structured CTI data. This task extended the READ system to detect model staleness or concept drift in the models, enabling analysts to configure custom rules to refine the model training process. The system will be able to select samples from the user-generated data for use in model training based on model behaviour, analyst actions and interests. The system was able to select samples from the user-generated data for use in model training based on model behaviour, analyst actions and interests.

The research made enhancements to the READ, including custom domain-specific overrides to entity extraction, however it was not possible to automatically select data from unstructured CTI reports. The research made enhancements to the READ tool, including custom domain-specific overrides to entity extraction, however Elemendar were not able to automatically select data from unstructured CTI reports.

Task 13: Accenture – Log Source Qualification [complete]

Log sources are essential to provide environmental visibility to detect notable security events, amongst benign indicators and false positives, however these sources need to be assessed and categorised to determine the data quality and relevance to security monitoring. This task comprised of three phases: Analyse and categorise detection logic and machine data; understand the relationships that exist between data logs and cyber-attacks (mapped to the MITRE ATT&CK framework); and incorporate findings to dramatically improve the ability to detect, understand and remediate cyber-attacks. The research demonstrated the use of detection logic against machine data to derive cyber activity on a network.

Task 14: Frazer-Nash Consultancy – File System Analysis for Automated Digital Forensics [complete]

This task investigated how a large set of file system metadata can be used as a knowledgebase, to classify ingress data from a file system, as suspicious, attribute it to a set of similar activity within the system and use this to understand which techniques an adversary is currently employing to attack a system. Snapshots were taken of the file system and then machine learning (ML) techniques were used to process the machine-readable data at machine speeds, enabling trial obfuscation and reconnaissance activities to be identified. This task was unique in exploiting file system metadata, which could potentially be pivoted to different operating systems.

Task 15: Tulpa – Modelling Adversarial Behaviour to Enable AI Predictions Analogous to Counterfactual Reasoning [complete]

This task developed a prototype knowledge graph, based on causal models, that encodes the adversarial behaviour of human experts. It investigated how causal knowledge graphs might be best combined with multiple ML techniques (inverse reinforcement learning, deep reinforcement learning, multi-agent co-training and transfer learning) to train explainable/scalable defensive agents capable of mimicking counterfactual reasoning to enable autonomous agents to be trained to undertake actions like automated vulnerability detection and adversarial prediction.

The research established techniques to explore, visualise and explain agent behaviour and to compare it with human derived Structural Causal Models. This task pioneered a causal neuro-symbolic implementation of cyber red agents which intelligently attacked simulated networks, and blue predictors which anticipated what their red adversary might do next.

Task 16: Riskaware – Critical Asset Cyber Terrain Identification (CACTI) Phase 2 [complete]

Previous work on the CACTI project (Task 1) explored different Graph Neural Network (GNN) architectures and data enrichment strategies resulting in a demonstrable capability (Technology Readiness Level (TRL 2)) this showed significant promise for identifying critical assets. Introspection analysis was also conducted to show how the model performed when treating networks as time-series, providing insight into criticality probabilities over time. This task harnessed the Phase 1 CACTI findings to further develop a capability that can, in future phases be integrated into autonomous agents. It responds to potential cyber-attacks within a dynamic environment in which criticality can change over time with the inclusion of models for classifying criticality in time-varying situations and use the CACTI model architecture to classify device type.

U75B INFORMED CYBER SENSING

Task 17: Montvieux – Consolidation of Joint Attribution Modelling (JAM) Framework [complete]

Attribution is an important part of the forensic investigation process following a Cyber-attack, but it remains a highly complex and predominantly manual task. It requires the labour-intensive analysis of many disparate, noisy, heterogeneous data sources to form a coherent overall picture of probable threat actors. The JAM framework, previously developed jointly by Montvieux and Frazer-Nash Consultancy (U75b Tasks 2 and 5) successfully applied Machine Learning (ML) to this data orchestration problem, providing a standard structure for data, tools and models, and implemented a simple two-source experiment covering malware text reports and network traffic captures. This task expanded JAM to handle new data sources (such as log telemetry), scaled up to production quality and prepared for integration into the ARCD environments. Additionally, it drew upon work done in U75b Task 11 to augment the Tasks 2 and 5 models with specialised Advanced Persistent Threat (APT) -level attribution models and refactored JAM into a scalable cloud-based service.

Task 18: BMT – Autonomous Resilient Cyber Defence – Informed Cyber Sensing (Prediction, Deception) [complete]

Where a cyber-attack cannot be prevented, understanding the timely prediction of the past and future path of the attack is crucial to understand, potentially enabling attackers to deploy deception tools. This task addressed the challenges of advancing prediction and deception techniques through the development of three proof-of-concepts (POC). POC1 – Prediction Module to analyse attack data to understand and predict the attack path and potential targets; POC2 – Deception Engine generating a Dynamic Deception Environment (DDE) to monitor and understanding an attacker's behaviour, continually adapting to an attacker's interests and activities to maintain engagement and allow collection of relevant data; and POC3, integrating POC1 and POC 2 to further improve modelling of attacker's behaviour, prediction and effectiveness of deception, this enabling a better defensive response.

Task 19: Riskaware – ICS CyberAware Subsystem [complete]

This project developed a Cyber Enrichment Engine integrating three Riskaware capabilities suitable for enriching agent data that can harness different measures of cyber risk on monitored environment: CyberAware Predict (developed as part of Dstl Predictive Cyber Analytics (PCA) Programme), CyberAware Resilience and Critical Asset Cyber Terrain Identification (CACTI). The Cyber Enrichment Engine uses the Structured Threat Information Expression (STIX) object notation for all Application Programming Interface (API) interactions to post ICS data to the service and by logging internal communication between components. The task successfully developed the Cyber Enrichment Engine to offer data enrichment services in support of cyber defence. Riskaware delivered the Cyber Enrichment Engine and supporting CyberAware Resilience, CyberAware Predict and CACTI source code and docker containers, facilitating future exploitation in frameworks such as ARCHON. A short technical guide was provided detailing the specification of the API, STIX data model and guidance for integrating the Cyber Enrichment Engine into ARCD clients.

Task 20: Riskaware – CACTI Phase 2 MAB Trial [ongoing]

MoD seeks to understand device criticality on their networks to better understand how to defend them. Phase 2 of CACTI is complete and provided an opportunity to classify critical devices in a network from Zeek conn logs. This trial also provides the opportunity to label a deployed network for future use across ARCD. Outcomes from this trial will be a trained CACTI model able to classify critical devices on the MoD network, as well as a criticality labelled dataset for future use across ARCD.

Task 21: Elemendar – Optimisation of READ [ongoing]

In the ever-changing domain of cyber defence, the capability to process Cyber Threat Intelligence (CTI) efficiently and accurately is critical for maintaining security posture. The READ tool, developed to enhance the autonomy in processing CTI, has shown significant technological advancements in addressing this need. However, challenges persist in the preparation of CTI data and the systematic evaluation of analyst performance, which are essential for effectiveness in operational environments. This project is focusing on user onboarding and training, providing a training session to ensure users can effectively utilise READ, while actively evaluating them for continuous feedback and process improvement. By streamlining and providing guidance for CTI data preparation, this project ensures READ's users are able to continue to process the most relevant and current intelligence, thus improving threat identification and mitigation. Lastly, by introducing a continuous evaluation framework for READ users' effectiveness, regular reports and recommendations foster an environment of ongoing refinement. This project will contribute to a more adaptable and responsive cyber defence posture, ensuring that both the READ tool's users are well-equipped to address the challenges of modern cyber warfare effectively by focusing on CTI data preparation, user proficiency, and continuous performance evaluation.

ARCD Track 1 Tasking Overview

U75B INFORMED CYBER SENSING

Task 22: Riskaware – CyberAware Subsystem Phase 2 [complete]

This task matured the CyberAware Subsystem, increasing situational awareness for cyber defenders and decision-making ability for automated agents. Under CyberAware Subsystem Phase 1, Riskaware developed an enrichment engine that intelligently orchestrates three previously developed Riskaware capabilities, CyberAware Predict, CyberAware Resilience and Critical Asset Cyber Terrain Identification (CACTI) with the goal of providing additional cyber context for cyber defence agents. The resulting enrichment extends the ARCHON Structured Threat Information Expression (STIX) data model for compatibility with ARCD environments and agents. The project integrated OpenCTI and Nessus using two new connectors to enhance the CyberAware Subsystem. The OpenCTI aggregator provides an additional service for analysts to gain situational awareness through use of OpenCTI dashboards, presenting enrichment data supplied by ICS services in a human readable format. The output provides a case for Nessus to provide CyberAware Subsystem with software vulnerability information required for its analytics. Riskaware delivered source code for CyberAware Subsystem along with each connector service for Nessus and OpenCTI, including instructions for building and deploying each service within Docker.

Task 23: Tulpa – Red Faction [ongoing]

In this task Tulpa will integrate both its red adversarial agent and blue predictor agent technology for use in the ARCD environments PrimAITE and provide a CHAOS- and STiX- compatible output for ARCHON and OpenCTI. Tulpa will provide modularised code with standardised inputs and output APIs for communication with PrimAITE. This design will decouple agent decision making from any environment-specific dependencies using a higher-level, machine-readable information representation which maps to environment-specific actions and observations via configurable environment adapter modules. This architecture has already been developed and demonstrated under Task 15, in which the agent is now decoupled from (but still able to interface with) the CybORG environment. This integration into PrimAITE will provide an initial stepping stone supporting delivery of agents into more realistic environments supplemented by an investigation and production of a roadmap for how the agents could be integrated into PalisAIDE at a later date. For the red agents, Tulpa will focus on integration of its SCM agents into PrimAITE developing additional functionality to enable training, testing and demonstration of blue agents across the ARCD supply chain.

Task 24: Montvieux – JAM Follow On [ongoing]

As part of the latest JAM project (U75b Task 17), MV included a proof-of-concept (PoC) SSL method for discovering embeddings optimal for Cyber attribution from unlabelled data generated in the ARCD PrimAITE environment. For this project Montvieux have proposed to generalise and adapt this approach in order to reduce the requirement for labelled Cyber data and test its applicability and effectiveness for training RL agents.

Task 25: Riskaware – CACTI Phase 3: Semi-Supervised Learning [ongoing]

The objective of CACTI Phase 3 is to use semi-supervised self-training in order to reduce the impact of having incomplete criticality labelled datasets which currently limits the real-world usability of CACTI. Previous work in the first two phases of CACTI focused on developing a Graph Neural Network (GNN) using model data from network logs (Zeek and NetFlow) to identify critical devices. Phase 3 will extend the current CACTI Command Line Interface (CLI) to apply semi-supervised self-training to the model architecture developed under phases 1 and 2. This will result in being able to use semi-supervised self-training to increase the amount of criticality labelled data and as such be able to train CACTI in settings where a fully labelled dataset is not possible.

Task 26: Riskaware – CyberAware Resilience Trial with 591SU [ongoing]

591 Signals Unit (591SU) at RAF Digby has a requirement to produce visualisations of red team cyber-attacks for blue team training. CyberAware Resilience should be able to meet this 591SU requirement, so a short trial will be conducted through ARCD with Riskaware under Task 26. Studying how real-world blue teams are trained to perform cyber defence might inform how autonomous blue agents could be trained using CyberAware Resilience output in the future.

U75B INFORMED CYBER SENSING

Task 27: BAE Systems – U75b Problem Book 3 – Foundational Secure by Design v1.0 [ongoing]

BAE Systems will be reporting on their research, findings and insights in hardware assurance and secure design of SoCs to mitigate against cyber vulnerabilities. BAE Systems has extensive experience of tracking cyber operations of nation states and have also carried out numerous cyber assurance tasks on their own platforms, aiming to identify weaknesses ahead of deployments. The task will split the problem area into four main components: hardware integrity, hardware exploitation, firmware integrity, firmware exploitation. The first aspect of research will be the verification of hardware and firmware and how systems integrators could detect malicious changes. The second aspect will incorporate a threat focused view to ascertain how exploitation has been carried out in operational environments to date, and how knowledge of such threat vectors can contribute to a system design that reduces the footprint for malicious cyber activity. A report will be produced to present the findings of this work and is due to be completed in March 2025.

Task 28: Cambridge Consultants – Reducing cyber risk in the SBOM (Software Bill of Materials) supply chain for complex SoC [ongoing]

Cambridge Consultants will be assessing cybersecurity risks associated with the Silicon IP (SBOM) used in System on Chip (SoC) devices and how these risks can be reduced. State of the art in SBOMs for SoCs and their associated cyber risks will be investigated and a methodology with guidelines and recommendations on minimising these risks will be developed under this task. This will be achieved by considering security risks through the whole lifecycle of the Silicon IP, from initial selection through to final programming. Cambridge Consultants will be producing a report on the best practice approach for the use of Silicon IP in the design and manufacturing of a complex SoC.

Task 29: Thales – Foundational Secure By Design DCS [ongoing]

Data Centric Security (DCS) underpinned by Zero Trust (ZT) has gained considerable traction in recent years and been the subject of on-going research, standardisation and gradual implementation globally. Much of this has been directed at the enterprise ecosystem and the drive towards a single information environment across military domains. For Task 29, Thales will capture a set of options for DCS implementation and analyse these in the context of ARCD relevant scenarios for future airborne platforms and off board support infrastructures to assess the utility and appropriateness of applying DCS in this context. Thales will produce a report and presentation documenting the results of the research analysis along with the options and scenarios investigated and recommendations for further research.

U75C AI DECISION MAKING

Task 1: Cambridge Consultants – CO-DECYBER Phase 2 [ongoing]

Three-year follow-on project applying a MARL approach using Deep Q-Networks (DQNs) for cyber defence of a 'platooning' scenario with leader-follower logistics vehicles. A bespoke training environment has been built based on NATO GVA utilising a simplified model of the Distributed Data Service (DDS) protocol. The project also includes a 'Live AI' end demonstration, where trained agents running on representative compute resources respond to simulated cyber-attacks to explore the challenges of real-world edge deployment. This project was presented at the SECAI 2023 Conference and CAMLIS 2024. The team are engaged with the GVA Digital Twin project under another area of Dstl to explore exploitation routes and opportunities to conduct research in more representative environments.

Task 2: Aleph Insights with the University of Liverpool – MIDGARD Phase 2 [complete]

Nine-month follow-on project using Gaussian Processes and Bayesian Optimisation for cyber defence in an air defence radar operator scenario. The research explored cyber defence decision making in a simulated "world" game, where the MIDGARD agent is defending the network (modelled in Yawning Titan) whilst a radar operator 'player' deploys air defence assets. The research explores the challenges of training an agent using sparse data in a noisy context. A small addendum to this project is described under U75c Task 17. The funding for this project included a PhD with the University of Liverpool ("RL for Physically-Aware Cyber Defence"), which started in October 2022.

Task 3: BMT with ADSP – MARL for Operational Technology (OT) [complete]

Eleven-month follow-on project expanding to a MARL approach using Multi Agent Proximal Policy Optimisation (MAPPO) for cyber defence of an abstract maritime Integrated Platform Management System (IPMS). Key findings included multi-agent approaches outperforming single agent, agents independently taking on cyber defence roles (e.g. container, eradicator) and agents successfully defending a network when presented with partially complete detection alert data. The follow-on to this project is detailed at U75c Task 15. This project was briefed to the Minister for the Armed Forces at the Defence Concept Paper Launch in Westminster Summer 2023, and a technical paper was presented at CAMLIS 2023.

Task 4: Illumr – Genetic Optimisation for RL [complete]

Four-month project following-on from previous work under SECTORED PBS159. Train and test 4 SOTA RL algorithms (DDQN, DQN, PPO and A2C) using both gradient descent and evolutionary / genetic optimization in the CAGE2 Challenge. Compare performance of Convolutional NN (CNN), and Recurrent NNs (Gate Recurrent Unit (GRU) and Long Short-Term Memory (LSTM)). Genetic optimisation routines trained faster than traditional gradient descent, with a slight general improvement in performance. Expectation is benefits will increase with more complex NN. architectures (see Task 9).

Task 5: BMT – Managing High Dimensionality in Cyber Defence Decision Making [complete]

Four-month literature review exploring state of the art approaches to dimensionality reduction, in support of the ARCD goal of demonstrating concepts on a high-fidelity representative environment. The review covers supervised and unsupervised learning (primarily split into feature selection and feature extraction methods), and RL. Cyber-specific analysis was conducted to identify and characterise high value data sources for cyber defence decision-making, to support AI specialists lacking cyber expertise. Findings are available in a full report or Aide Memoire format and were presented at the poster session at CAMLIS 2023.

Task 6: BAE DI – Deep RL for Autonomous Cyber Operations (ACO): A Survey [complete]

Five-month study, surveying relevant DRL literature and conceptualizing an idealised ACO-DRL agent. The report provides i.) A summary of the ACO domain properties; ii.) An overview of benchmarking environments with comparable properties to ACO; iii.) An overview of approaches for scaling DRL to domains that confront learners with the curse of dimensionality; and iv.) A survey and critique of current methods for limiting the exploitability of agents within adversarial environments. A survey and critique of current methods for limiting the exploitability of agents within adversarial environments. A paper is under final approval for the Journal for Autonomous Agents and Multi-Agent Systems (JAAMAS).

U75C AI DECISION MAKING

Task 7: University of Kent with University of Newcastle – Playing Cyber Games [complete]

Nine-month project exploring the use of cyber-attack symptoms for cyber decision making, rather than causes. An initial literature developed a symptoms taxonomy aiming to provide coverage across relevant Mitre ATT&CK TTPs. The project adapted the PrimAITE environment to implement a basic RL Red Agent but experienced difficulties implementing the symptom-based approach.

Task 8: ADSP – Minimum Viable Product (MVP) Agent Integration [complete]

Three-month fast paced project to implement the first integration of a Track 1 agent (a basic out of the box PPO agent) into a Track 2 environment (PrimAITE V1.0), followed by stress-testing of the environment. Integration was successful and environment modification recommendations were raised, which were implemented in PrimAITE V2.0. This project included the first known ARCD demonstration of an ML cyber defender outperforming a rules-based agent developed with a human analyst. A second MVP project has commenced, summarised at Task 19.

Task 9: illumr – Genetic Optimisation for RL [complete]

Six-month follow-on to U75c Task 4, expanding on genetic optimization research to more complex CNN and RNN architectures within the CAGE2 environment. The project also explored transfer learning, finding DDQN approaches to be more robust than PPO in defending against new, and previously unseen enemies.

Task 10: Trustworthy AI – Generalised Cyber Defence of Military Tactical Networks [complete]

Six-month project researching generalizable blue agents utilising deep RL and Heterogeneous Graph Neural Networks. A new environment framework, Cyber Adaptive Environment for Simulating Autonomous Response (CAESAR), was developed to model tactical networks, with a rich defensive action space including decoys. Agents were trained using adversarial learning in 60 network topologies and demonstrated an ability to defend 20 un-seen topologies, all generated from GPT4 prompts developed in collaboration with a Dstl technologist. A long-term follow-on project is described at Task 20.

Task 11: BT – Automation of Response with RL [complete]

Four and a half-month project aiming to i) improve the level of 'primary' level offensive tooling to train blue agents, and ii) the identification of new, unseen threat pathways in a known environment. Building on their epidemiological modelling tool Inflame, BT utilised their in-house defensive cyber teams to develop RL-based red agents as more sophisticated adversaries to train against than is currently available in ARCD training environments. Software outputs have been provided with a BSD-3 license, enabling wider adaptation and re-use within Defence. A long-term follow-on project is described at Task 21.

Task 13: Smith Institute – Bayesian Games for Decentralised Multi-Agent Decision Making [complete]

Six-month project in response to the 'Diversification of AI approaches' Problem Book. The project will develop and investigate the use of Adaptive Social Learning (ASL) to train decentralised social-learnt agents, acting locally, to defend a system globally (Bordignon et al., 2021). Agents can be taught to be sceptical of information from other agents when whole sections of the network are compromised and isolated, enabling distinct behaviour from "secure" network states. ASL's distributed approach showed great promise, particularly in terms of generalisability, scaling and resilience. Indicative results in January 2024 showed ASL/PPO agents trained in a 9-node network achieving an 88.4% win rate when transferred to a 50-node network. Significant training time and performance benefits over the current SOTA approach (PPO) were also observed, which increased as network size increased. A long-term follow-on project is described at Task 26.

Task 14: Xewli – Topology and Weight Evolving Artificial Neural Networks (TWEANN) [complete]

Three and a half-month project exploring TWEANN for automated cyber response as an alternative the more widely used RL methods. Three and a half-month project in response to the 'Diversification of AI approaches' Problem Book, exploring TWEANN for automated cyber response as an alternative the more widely used RL methods. TWEANN uses genetic algorithms to evolve a population of neural networks until an optimal solution is reached (i.e. "AI designing AI"). Two neuroevolutionary techniques were explored: NeuroEvolution of Augmenting Topologies (NEAT) and CoDeepNEAT to evolve the topology and the weights of the connections between nodes in a neural network. The NEAT approach scored closed to the CAGE2 benchmark, with a topology similar to that the published human NN designs for CAGE2. CoDeepNEAT performed far worse. Ultimately both approaches were constrained significant compute requirements, leading to low confidence the approach would scale.

U75C AI DECISION MAKING

Task 15: BMT – MARL for OT Phase 3 [ongoing]

Twenty one-month follow-on to U75c Task 3, aiming to continue to progress research of SOTA MARL techniques, initially including generalisability and action masking in a maritime IPMS environment. The environment has undergone significant modification to reach the 'Secondary' level (i.e. TRL-5). Experimentation has shown considerable benefits for action masking and curriculum learning and a paper on this topic was presented at CAMLIS 2024. Real-world Multi-Agent architectures are also being explored. Exploitation on a 'real' IPMS system is our key future focus for this concept. A scoping study assessed potential physical target systems for the next phase of demonstration (TRL-6), including the Dstl IPMS Proxy System and Cyber-SHIP Lab at the University of Plymouth. The Dstl IPMS Proxy System was selected due to its military representativeness and focus on industrial control systems (see Task 23 below).

Task 16: BAE DI – Evaluating the Exploitability of Autonomous Cyber Operations Agent [complete]

Three-month exploratory project exploring exploitability of blue agents in the CAGE2 environment. Here, exploitability is measured using an approximate best response (ABR) that a new opponent can learn against the fixed policy of the agent being evaluated (Blue). Using the computed ABR, exploitability quantifies how much a player (in this case Red) gains through unilaterally deviating to the ABR. Results showed that whilst the CAGE2 winner appeared to be the most performant in that competition, it was highly exploitable by varying red agent strategies. ML-based red agents using curriculum learning were used as the ABR and were found to pose a higher threat to blue agent than the CAGE2 rules-based agents. A follow-on project is described at Task 22.

Task 17: Aleph Insights – MIDGARD Trial Addendum [complete]

One-month addendum task to U75c Task 2, with the aim of providing additional benchmarks, and evaluate the performance of an agent that optimises a cyber-score vs. a 'real-world' score. Findings related to benchmarking were shared with the ARCD Evaluation and Experimentation project. Findings related to benchmarking were shared with the ARCD Evaluation and Experimentation project and are reflected in current Evaluations.

Task 18: Exalens: Virtual Incident Response Agent (VIRA) Phase 2 [ongoing]

Seventeen-month project building on a DASA project that achieved ARCD's first end-to-end demonstration of autonomous cyber defence against a real cyber-attack on a real OT system (ROSbot). VIRA is a RL-based system aiming to offer first-aid to cyber-physical targets in the field, analysing threat alerts, and triggering or recommending understandable containment actions to a non-expert operator, to provide crucial time for detailed remediation and recovery. This phase aims to demonstrate on a representative military system to achieve TRL-6 (a Dstl Unmanned Aircraft System (UAS) concept – see Task 27).

Phase 2 has developed an "automated environment learning engine" enabling self-configuration to new environments. A "mixture of incident response experts" model architecture combines triage, correlation and action selection roles within an impact-aware "Cyber First Aid" incident response workflow. This will be the first ARCD concept to apply "Rules of Engagement" by modulating agent action selection, as recommended in the recent "Authorised bounds for autonomous agents" NCSC Briefing Paper.

Task 19: ADSP – MVP Agent Phase 2 [complete]

Three-month project which explored new opportunities for Track 1 posed by the upgraded ARCD training environments: PrimAITE V3 and Imaginary Yak. These environments provide a larger action space and an avenue to test multi-agent systems, but they also present integration complexities. The research was able to demonstrate that RL could learn to handle the more complex PrimAITE V3 simulation environment, and agents trained in PrimAITE could be transferred to the Imaginary Yak emulated environment, a milestone in addressing the Sim-to-Real challenge.

Task 20: Trustworthy AI – Generalised Cyber Defence of Military Tactical Networks – Phase 2 [ongoing]

Seventeen-month project building on from the promising generalisability potential shown in Task 10. The project aims to transfer its DRL / Heterogeneous Graph Neural Network approach to an emulator approach (TRL-5/6). The project uses an interesting 'shared network' between the simulator and an emulator running virtual machines and enact 'real' changes to the representative network. An ambitious, rich red and blue action space has been developed using real tooling including Cobalt Strike and Elastic Defend.

Defence Digital have supported the project by providing information on MOD's cyber infrastructure and Cyber Mission Data (CMD) suite of sensors and there is active knowledge sharing with the US CASTLE programme relating to graph-based RL. The project will undergo ARCD's first human red teaming operation (described at Task 30).

ARCD Track 1 Tasking Overview

U75C AI DECISION MAKING

Task 21: BT – Odin Phase 2 [ongoing]

Task 11 developed a RL agent to conduct red team attacks on a simulated network and demonstrated an ability for agents to learn an optimum path of action for particular strategies. This continuation builds on previous learning to assess a co-evolutionary approach to the development of defensive and offensive agents.

Progress thus far includes the development of more sophisticated red agent strategies including multi-stage attacks, enhancing observation space for the problem including capture of device types to aid generalisability and development of a network health score to build understanding of real-world impacts. Next steps will develop strategies for co-evolutionary learning and development of diverse training environments. BT are investing in a long-term vision for this concept including trials on enterprise infrastructure and collaboration with customers within MoD.

Task 22: BAE DI – Reducing the Exploitability of Cyber-Defence Agents [complete]

In many real world scenarios, the assumption is that Blue cannot learn once deployed. Therefore, a framework is required for obtaining Blue policies that are robust and resilient towards worst-case opponents prior to deployment. Task 16 measured exploitability through training Red cyber-attacking agents against Blue cyber-defence agents in CAGE Challenge 2.

In Task 22 BAE developed a novel principled adversarial learning framework for reducing exploitability, using Multiple Response Oracles. Results showed that training blue agents using this approach reduced their exploitability, increasing their robustness against varied red agents (a follow-on task is described at Task 31).

Task 23: BMT – MARL for OT Phase 4 [ongoing]

Task 23 is a 13-month project run in parallel with Task 15 to include the delivery of a TRL 6 demonstrator on the Dstl Maritime IPMS Proxy system. A 13-month project run in parallel with Task 15 to include the delivery of a TRL 6 demonstrator on the Dstl Maritime IPMS Proxy system. The project is exploring the significant engineering challenge in deploying ARCD agents onto a physical system.

Dstl have developed a “middleware” layer which the ARCD agents interact with to alleviate sim-to-real gap problems and deploy to the proxy environment. We have now successfully deployed our first trained cyber response & recovery agents in what we believe to be a world first for Military Industrial Control Systems. Due to security classifications this project involves significant on-site working at Porton Down with close collaboration with Dstl cyber security scientists.

Task 24: Montvieux – World Models Phase 2 [ongoing]

Our first HRDO graduate extends Phase 1 of this research (HRDO Task 5) which showed promise for World Models (WMs) to be used to reduce the amount of training required by cyber defence agents through ‘dreaming’ based on limited time spent training in an environment. Our first HRDO graduate extends Phase 1 of this research (HRDO Task 5) which showed promise for World Models (WMs) to be used to reduce the amount of training required by cyber defence agents through ‘dreaming’ based on limited time spent in a training environment.

This phase is addressing some of the Phase 1 challenges by enhancing generalisability through use of graph embedding of the network state and training on many different network topologies. As well as training in the PrimAITE simulation environment, Montvieux are training on their emulated SHAPESHIFTER environment (developed under a DASA project) to test the benefits of WMs applied to a higher fidelity environment and explore whether this could reduce the need for simulators for agent training.

Task 26: Smith Institute – Adaptive Social Learning Phase 2 [ongoing]

This 9-month project follows on from Task 13, which showed great scaling potential, to further explore ASL in more complex scenarios. Moving from the previous prototype in Yawning Titan, into PrimAITE V3.3, the research is exploring the potential limits for ASL scalability, and the resilience of ASL architectures to realistic interruptions such as dropouts or impacts of attacks. The research will also investigate the practical network engineering considerations when deploying this multi agent approach requiring communication between agents.

Task 27: Frazer-Nash – UAS Support to VIRA [ongoing]

9-month project aiming to provide a training and demonstration platform to support the VIRA concept (Task 18) in achieving TRL-6 through demonstration on a representative cyber physical military system. The UAS has been developed under the Dstl Counter Terrorism area and is a good example of a future autonomous cyber physical platform that may be deployed without a dedicated cyber operative with the additional benefit of being a live Dstl research project. The project has developed hardware in the loop simulators for both single and multi-system applications to generate PCAP datasets that can be used to train the VIRA agents. VIRA Integration is complete, and field trials have commenced, with a final demonstration in January 2025.

U75C AI DECISION MAKING

Task 28: ADSP – Deployment to the ARCD Demonstration Environment [ongoing]

Development of Track 2 Test and Evaluation capability, including significant enhancement to PrimAITE and imminent availability of PalisAIDE (the ARCD demonstration environment) provide new opportunities to train and evaluate ARCD agents in more complex cyber environment.

This project is exploring the agent's performance under varied network scenarios, red/green agent behaviours and more, enhancing its adaptability and effectiveness. The project is the most collaborative ARCD project to-date, requiring long-term tactical working relationships between Track 1 and Track 2. Examples include ongoing refinement of environments, independent evaluations and feedback to understand and improve agent performance, whilst providing a holistic demonstration of a range of capabilities and functionality across ARCD.

Task 29: Tulpa – Human Machine Teaming Prototype for Autonomous Response and Recovery [ongoing]

More knowledge is needed to (a) understand how ARCD agents should be integrated alongside human operators of varying levels of skill, and (b) set out a roadmap for how we can go from where we are now to a 'full auto' cyber defence capability. This project integrates human sciences research with software engineering to build a HMT console through which users can manage cyber defence agents operating at different autonomy levels in simulated scenarios that depict a range of autonomous actions and use cases.

Military users are being engaged in capturing and testing hands-on workflow experiences which allow them to explore and intervene in various aspects of ARCD agent deployment, from initial configuration and policy definition, through performance monitoring and online behaviour correction, to wind-down and after-action review. A fully working simulated prototype is being developed for user evaluation and a roadmap to 'full-auto' will be derived to inform future ARCD direction.

Task 30: 6point6 – Provision of Specialist Red Team Resource [ongoing]

Continuation of the Task 25 trial providing red teamer oversight of U75c Task 20 (TAI). This project includes attendance at monthly technical reviews, and independent review of artefacts to provide a view of quality, accuracy and risk from a red team perspective. Additionally, Accenture will conduct a 'lite' simulated red team exercise, lasting for 5 days, against the scoped target ARCD system, outlined by TAI. The intent for this limited exercise is to experiment with red teaming ARCD agents, providing initial feedback on the TAI and informing future ARCD direction regarding red teaming of autonomous cyber defence agents.

Task 31: BAE – Exploitability Phase 3 [ongoing]

In U75c Task 22 BAE introduced a principled adversarial learning framework, designed to reduce the exploitability of cyber-defence agents. The approach is based on a novel Multiple Response Oracle (MRO) algorithm using Game Theory to determine policy sampling probabilities sample over sets of Blue and Red policies. Solution concepts from Game Theory are applied to the bimatrix game, determining policy sampling probabilities.

Follow-on Task 31 aims to address two identified bottlenecks: i.) Lengthy wall-times for iteratively computing approximate best responses (ABRs), and ii.) The time-complexity for augmenting the empirical payoff matrix. BAE propose principled extensions to address these challenges, including: i.) A novel Potential-Based Reward Shaping (PBRS) approach, and; ii.) A policy pruning process to reduce the time-complexity for payoff matrix augmentation. Their extensions will be evaluated on versions of CybORG CAGE Challenges 2 and 4 (CC2&4) featuring graph-based observation wrappers, enabling the training of Deep Reinforcement Learning methods implemented with Graph Neural Networks (DRL-GNN). The project is also aiming to deploy to the DARPA CASTLE MATrEx environment, to test against CASTLE red agents in a variation of the CAGE2 Challenge. The project is also aiming to be the first ARCD agent to deploy to the DARPA CASTLE MATrEx environment, providing an opportunity to test against CASTLE red agents in a variation of the CAGE2 Challenge.

U75C AI DECISION MAKING

Task 32: ADSP – LLM Agents For Cyber Defence A Zero-Shot Approach – Phase 2 [ongoing]

This project builds on HRDO Task 37 where it was demonstrated promise for off the shelf 'zero-shot' LLMs as suitable alternatives to RL cyber defence agents, eliminating the need for costly bespoke training processes. In this second phase ADSP aim to review the rapidly evolving LLM landscape (including self-hosted services), reducing latency and expanding memory options to address some of the limitations found in Phase 1. This project also aims to gain understanding of how LLM agents perform in increasingly complex environments.

Task 33: Frazer-Nash – ARCD Exploitation Readiness Level Assessments [ongoing]

The ARCD project team has a large range of research projects within their portfolio and has a need to build a systematic understanding of readiness for exploitation. In collaboration with Dstl, Frazer-Nash are defining a maturity assessment framework, including consideration of traditional approaches such as Technology Readiness Levels, as well as emerging ML considerations such as those outlined in JSP 936: Dependable AI in Defence. The project will conduct assessments on the most mature ARCD concepts to capture understanding of their areas of maturity and current limitations to build understanding of exploitation opportunities and challenges.