## Day One - Tuesday 9th September

## **Next steps for Autonomous Cyber Defence**

09:30 - 10:00	Registration & Refreshments
10:00 - 10:15	Welcome & Introduction to AMLUCS
	Alex Revell, Frazer-Nash Consultancy
10:15 - 10:45	Keynote
	Sadie Creese, Professor of Cyber Security, University of Oxford; Founding Director, Global Cyber Security Capacity Centre; Strategic Advisory Board Member, World Economic Forum.
10:45 - 11:15	Adaptive by Design: Contextual Reinforcement
	Learning for Mission-Ready Cyber Defence
44.45.40.00	Dr Jake Thomas, Advai
11:15 - 12:00	Break & Refreshments
12:00 - 12:30	Autonomous Cyber Resilience: An Infrastructure-as- Code Approach for Coalition Military Networks Dr Konrad Wrona, NATO Communications and Information Agency
12:30 - 13:00	Deploying Autonomous Cyber Defence onto a Military Relevant Physical System  Alec Wilson, BMT
13:00 - 14:15	Lunch & Poster Viewings (The Annexe)
14:15 - 15:00	Panel: Operationalising AI - Integrating Machine Learning into Defence Cybersecurity Systems Helen Wilson, Dstl, with Ministry of Defence stakeholders.
15:00 - 15:15	Autonomous Defensive Cyber
	Wayne Gould & Lara Tolley, Dstl
15:15 - 15:45	Break & Refreshments
15:45 - 16:15	Simulation-to-Reality Gap via RL-trained ROSbots
	Dr Jack Smith, Awerian
16:15 - 16:45	Topological Extensions for Reinforcement Learning Agents (TERLA).
14.45 47.00	Tim Dudman, Riskaware
16:45 - 17:00	Day One: Closing Remarks
47.00 10.00	Alex Revell, Frazer-Nash Consultancy
17:00 - 19:00	Drinks Reception (Exhibition Space) and MU

Brought to you by





In partnership with



## Day Two - Wednesday 10th September

## **Red Agents, AI Security & Evaluations**

09:00 - 09:30	Morning Refreshments
09:30 - 09:45	Day Two: Opening Remarks
	Alex Revell, Frazer-Nash Consultancy
09:45 - 10:30	Keynotes: including the Impact of AI on cyber threat
	from now to 2027
40.20 44.00	Peter H & Annabel W, NCSC
10:30 - 11:00	RAGING MINOTAUR: Improving Defence against Al-driven Cyber-Attacks by Designing More Capable Autonomous
	Cyber Training Adversaries
	Althea Waites & Sharaz Anwer, Dstl
11:15 - 11:30	Lessons Learned in the Application of Reinforcement
	Learning Agents for APT Attack Path Generation
	Chad Caison, Six24 Cyber Labs, SIEGE team, DARPA CASTLE
11:30 - 12:00	Break & Refreshments
12:00 - 12:30	Mitigating data poisoning attacks against RL from
12:00 - 12:30	Human Feedback (RLHF) finetuning
	Human Feedback (RLHF) finetuning Dr James Titchener, Raytheon
12:00 - 12:30 12:30 - 13:00	Human Feedback (RLHF) finetuning Dr James Titchener, Raytheon Text2VLM: Adapting Text-Only Datasets to Evaluate
	Human Feedback (RLHF) finetuning Dr James Titchener, Raytheon Text2VLM: Adapting Text-Only Datasets to Evaluate Alignment Training in Visual Language Models
12:30 - 13:00	Human Feedback (RLHF) finetuning Dr James Titchener, Raytheon Text2VLM: Adapting Text-Only Datasets to Evaluate Alignment Training in Visual Language Models Dr Jake Thomas, Advai
12:30 - 13:00 13:00 - 14:00	Human Feedback (RLHF) finetuning Dr James Titchener, Raytheon Text2VLM: Adapting Text-Only Datasets to Evaluate Alignment Training in Visual Language Models Dr Jake Thomas, Advai Lunch & Poster Viewings (The Annexe)
12:30 - 13:00	Human Feedback (RLHF) finetuning  Dr James Titchener, Raytheon  Text2VLM: Adapting Text-Only Datasets to Evaluate Alignment Training in Visual Language Models  Dr Jake Thomas, Advai  Lunch & Poster Viewings (The Annexe)  A Statistical Pipeline for Uncertainty Quantification in
12:30 - 13:00 13:00 - 14:00	Human Feedback (RLHF) finetuning  Dr James Titchener, Raytheon  Text2VLM: Adapting Text-Only Datasets to Evaluate Alignment Training in Visual Language Models  Dr Jake Thomas, Advai  Lunch & Poster Viewings (The Annexe)  A Statistical Pipeline for Uncertainty Quantification in ACD Test and Evaluation
12:30 - 13:00 13:00 - 14:00	Human Feedback (RLHF) finetuning Dr James Titchener, Raytheon  Text2VLM: Adapting Text-Only Datasets to Evaluate Alignment Training in Visual Language Models Dr Jake Thomas, Advai  Lunch & Poster Viewings (The Annexe)  A Statistical Pipeline for Uncertainty Quantification in ACD Test and Evaluation Smith Institute
12:30 - 13:00 13:00 - 14:00 14:00 - 14:30	Human Feedback (RLHF) finetuning  Dr James Titchener, Raytheon  Text2VLM: Adapting Text-Only Datasets to Evaluate Alignment Training in Visual Language Models  Dr Jake Thomas, Advai  Lunch & Poster Viewings (The Annexe)  A Statistical Pipeline for Uncertainty Quantification in ACD Test and Evaluation
12:30 - 13:00 13:00 - 14:00 14:00 - 14:30	Human Feedback (RLHF) finetuning Dr James Titchener, Raytheon Text2VLM: Adapting Text-Only Datasets to Evaluate Alignment Training in Visual Language Models Dr Jake Thomas, Advai  Lunch & Poster Viewings (The Annexe)  A Statistical Pipeline for Uncertainty Quantification in ACD Test and Evaluation Smith Institute Modelling adversarial behaviour to enable Al
12:30 - 13:00 13:00 - 14:00 14:00 - 14:30	Human Feedback (RLHF) finetuning Dr James Titchener, Raytheon  Text2VLM: Adapting Text-Only Datasets to Evaluate Alignment Training in Visual Language Models Dr Jake Thomas, Advai  Lunch & Poster Viewings (The Annexe)  A Statistical Pipeline for Uncertainty Quantification in ACD Test and Evaluation Smith Institute  Modelling adversarial behaviour to enable Al predictions analogous to counterfactual reasoning

**Brought to you by** 





In partnership with

