# All-Hands Meeting 2026 Presentation Abstracts

## Contents

# Keynotes

## Speakers

**Opening Keynote**
**Fabio Pierazzi**, Associate Professor, University College London (UCL)

**Industry Keynote**
**Edith MacArthur**, IT Principal Security Architect, Evernorth Health Services

**Closing Keynote**
**Peter Haigh**, Deputy CTO, National Cyber Security Centre

# Sponsor Presentation

## CyBOK

**Title:** CyBOK in Action: Making Cyber Security More Accessible

Yulia Cherdantseva,
Steve Schneider,
Cliffe Schreuders,
Clare Johnson,
Steve Furnell

Wed 09 Sept
13:00 – 14:30

This showcase event will explore how CyBOK is helping to make cyber security more accessible, supporting a stronger and more inclusive cyber security community.

The session will begin with an update on CyBOK's evolution as a Community Interest Company (CIC), highlighting its commitment to remaining a community-driven resource.

Attendees will hear a presentation on a CyBOK-funded project focused on providing open-source practical challenges and learning resources, followed by a talk on how CyBOK is being used by the National Cyber Security Centre (NCSC) to support cyber security practice and capability development. The event will also showcase international collaborations that are advancing CyBOK's vision of a global cyber security knowledge framework.

The event will conclude with a panel discussion examining how cyber security can be made more accessible, and the role that shared knowledge, community collaboration, and practical resources can play in broadening participation across the field.

# Special Interest & Working Group Sessions

## Economics of Cyber SIG

*Topic 1:* Decision Support with CRQ Indicators and Metrics by Eireann Leverett (30 mins)

*Topic 2:* Economy-Security Interdependence and the Geopolitics of Technology Regulation: Digital Sovereignty and the Implications of Strategic Autonomy for EU-UK cooperation by Ben Farrand (30 mins)

Topic 1 Abstract:
Cyber Risk Quantification is needed to budget for cyber defences, yet uncertainty abounds. We are building momentum, but we must discuss many foundational elements such as “what is a population” in cyber risk statistics, and “what is a location or the unit of cyber risk aggregation. We can discuss why good cyber risk metrics must be fractal or scale free, and how micro and macro security economics are dealing with different uncertainties and parameters. Do we need a bureau of cyber risk statistics, and what makes a good metric or index for cyber risk?

Topic 2 Abstract:
This talk examines how contemporary legal and regulatory frameworks increasingly align economic and security rationales in the governance of digital technologies. Drawing on work at the intersection of political economy, cybersecurity law, and digital sovereignty, it argues that European regulatory instruments now operate not merely as market-shaping tools but as mechanisms for recalibrating power, reducing dependency, and mitigating risk within global technology markets. Measures such as data governance rules, cybersecurity certification regimes, and emerging digital industrial policies are presented as responses to perceived vulnerabilities in supply chains, infrastructures, and platforms, while simultaneously advancing strategic economic objectives. The talk explores how this economy/security alignment reshapes market behaviour by embedding security logics into ostensibly economic regulation, through a process conceptualised as ‘regulatory mercantilism'. It considers both the increased focus in the EU on European solutions and strategic autonomy has potential implications for private sector operators in the UK wishing to provide cybersecurity services in the EU, concluding with the need for increased regulatory alignment between the EU and UK, both to ensure the UK’s potential strategic aims under the industrial strategy, while allowing for a ‘pooled’ sovereignty of benefit to both public and private sectors.

Chair: Professor Siraj Shaikh (Swansea University),

Eireann Leverett,

Ben Farrand

Tues 08 Sept

14:45 – 15:45

## Learned Society Formation WG

Professor Shujun Li (University of Kent)

Tues 08 Sept

14:45 – 15:45

**Title**: LSFWG Session

If you are interested in knowing more about the planned learned society and contributing to shape future activities of the LSFWG and the learned society, you are welcome to attend the parallel session. The session will be largely interactive to allow participants to provide feedback to help inform future planning of the LSFWG's activities after the All-Hands meeting.

## Inclusive Careers WG

Professor Lynne Coventry (Abertay University)

Tues 08 Sept

16:15 – 17:15

**Title:** Inclusive Careers Panel Discussion

This session will be an informal and open discussion on the topic of inclusive careers, joining Lynne will be one of our Keynote Speakers, Edith MacArthur and ECR Neeranhan Chitare

## SANE SIG

Dr Vadim Safronov (University of Oxford)

Wed 09 Sept

10:45 – 11:45

**Title**: Towards Inference-Aware Cyber-Physical Infrastructures: Challenges in Security, Privacy and Accountability *(Interactive session)*

**Brief Summary:** Inference-aware cyber-physical infrastructures are transforming how intelligent systems sense, learn, act and share information across distributed environments. This interactive session will explore emerging challenges in security, privacy, accountability and architectural design for future intelligent infrastructures.

## Safe & Secure Robotics SIG

Professor John Clark (University of Sheffield)

Dr Jims Marchang (Sheffield Hallam University)

Weds 09 Sept

14:45 – 15:45

**Title:** Open questions and Interactions on safe and secure robotics

This session will be an informal and open workshop which will encourage open questions and dialogue on the special interest area of safe and secure robotics and cyber security.

## Research Practices SIG

Professor Steven Murdoch (University College London)

Wed 09 Sept

14:45 – 15:45

**Title:** Research Practices SIG Workshop

This meeting will be a workshop looking at the scope and plan of activities for this SIG including proposed deliverables including an approach to ethics for CRANE, a code of practice for the upcoming learned society, guidance for cybersecurity researchers and their institutions, and engagement with policymakers.

# Technical Talks – Session 1: Cyber Resilience: Organisations, Incidents and Capacity

## Talk 1

**Title:** Research insights from investigating and supporting cyber security in Small and Medium-sized Enterprises

Steven Furnell,
Maria Bada,
Jason Nurse,
Neeshé Khan,
Matthew Rand
Ram Herkanaidu

Tues 08 Sept
14:45 – 15:45

The Small and Medium-sized Enterprises (SMEs) account for over 99% of UK businesses and yet are continually revealed to be under provisioned in terms of cyber security, while experiencing incidents in practice. Consequently, SME cyber security represents a rich vein for related research, both in terms of gaining a deeper understanding of the problem and seeking means to improve the situation.

For the last 2.5 years the UKRI/RISCS funded CyCOS project (www.cycos.org) has been investigating the cyber security needs of SMEs and piloting a new approach for assisting the situation via communities of support. Through the various phases of the work, the project team has been accumulating significant experience of engaging with both the SME community, and the providers that already support them in understanding and addressing cyber security issues. As a result, there is a strong understanding of SMEs' cyber security posture and concerns, as well as the constraints that affect their engagement.

The presentation will summarise the findings gathered across various interventions within the CyCOS project – including a survey of over 370 SMEs, 24 in-depth interviews with SMEs and providers, 75 support journeys experiences, and the operation of support communities themselves. The findings include details of SMEs' access to support, and the suitability of what is available. The session will offer resulting insights in three key areas:

- What does SME cyber security look like?
- What support do SMEs need and access?
- What are the wider lessons for SME-facing cyber security research?

These in turn can help to set a basis for further engagement to support this important community.

## Talk 2

**Title:** Zero Trust After the Breach: Lessons from Real-World Incident Response

Andrea Ibiassi

Tues 08 Sept
14:45 – 15:45

When a breach hits, the instinct is clear: shut everything down. But is that always necessary and what does it cost the business?

In the face of increasingly sophisticated attacks, perimeter-based defences often collapse once compromised. This session explores how Zero Trust changes that equation, not during the breach itself, but in how organisations recover and rebuild securely.

Drawing on real-world incident response in large enterprises, the talk shows how to turn reactive recovery into long-term resilience. It highlights what actually works in practice: reducing lateral movement, limiting blast radius, and maintaining essential services while addressing practical challenges.

This includes rapid onboarding of tens of thousands of users under crisis conditions, enabling secure third-party access for recovery teams, and implementing segmentation across workforce, datacentre, branch, and OT environments.

Breaches create a rare window to enforce least-privilege access and accelerate Zero Trust adoption.

Talk 3

**Title:** Enhancing Impact Evaluation in Cyber Security Capacity Building

Phil Sheriff

Tues 08 Sept
14:45 – 15:45

The Cybersecurity Capacity Building is a tool through which not just governments, but also NGOs and the private sector, can pursue a range of objectives, from national security to development and trade related outcomes and impacts.

However, unlike other policy areas, the evidence base for CCB interventions is inadequate, and the process of evaluating the impact of CCB programs immature. This poses significant challenges to enabling best practice in evidence based policy in CCB programming and implementation.

This research project examines the reasons behind the lack of effective impact evaluation, through examining specific challenges, for example the diversity of drivers and the lack of data, as well examining the often poorly conceptualised logical foundations for CCB programming. The research also examines methods and practices from other well established policy areas and adapts them to the specific area of CCB. For example, the research examines the experimental vs theory based approaches to evaluation, and the balance between quantified and qualified data.

The research proposes a typology of CCB interventions, and then uses the typology to propose a unique framing of, and solution to, impact evaluation in CCB, through the example of interventions aimed at establishing and enhancing the effectiveness of national Computer Security Incident Response Teams (nCSIRTS). The solution combines Incident Response industry models with evaluation logic models to enhance impact evaluation. The CyberSecurity Capacity Building ScoreCard is suggested as a way of not only demonstrating impact, but also of identifying relative contributions of different types of interventions.

The presentation will aim to summarise the research project, demonstrate the scorecard approach, and suggest how it can scale to other CCB programs.

# Technical Talks - Session 2: AI-Automation and Trustworthy Cyber Defence

| | |
|---|---|
| Talk 1 | **Title:** Fine-Grained Authorisation for AI Agents: Enabling Continuous Trust Verification |
| Shivangi Gupta,<br>Budi Arief,<br>Rogerio de Lemos<br><br>Tues 08 Sept<br>16:15 – 17:35 | A significant recent change in the development and application of artificial intelligence (AI) is the transition from non agentic to agentic AI. In contrast to non-agentic AI systems, agentic AI systems are designed to act more autonomously through goal setting and multi-step<br>planning. This autonomy enables agentic AI systems to re-plan and accomplish user tasks through goal-oriented behaviour, thereby reducing the user's workload. However, this autonomy introduces several challenges, including potential security risks, such as privilege escalation due to overly permissive access rights assigned to an agent.<br>This research highlights the need for fine grained, dynamic authorisation in agentic AI systems. The re search focuses on identifying the limitations of existing authorisation frameworks, such as OAuth with delegation extensions, in the context of agentic AI systems. To address these limitations, we introduce an approach that combines JWT(JSON web token) feature of OAuth framework with decentralised delegation and authentication methods for reliable and secure agentic behaviour. Authentication will be handled using agent cards, while authorisation will be managed through a combination of tokens and a gateway integrated with a policy engine. The framework will include continuous checks at each step before the agent reaches the server, where the authorisation will be handled by the gateway. This approach will ensure that agents are authorised based on the assigned task, rather than relying on predefined scopes or permissions that may be overly broad or overly restrictive. |

## Talk 2

**Title:** Immune-Inspired Proactive Cyber Security: A Novel AI-Driven Defence Paradigm

Abdullahi Arabo

Tues 08 Sept

16:15 – 17:35

Cyber-security threats are increasingly adaptive, AI-enabled, and automated, placing growing pressure on security systems that remain predominantly reactive. While emerging technologies such as artificial intelligence, large language models (LLMs), automation, and future quantum capabilities promise enhanced defensive capacity, their integration into coherent, proactive security models remains under-explored. This short talk presents a novel, immune-inspired approach to cyber-security design, explicitly drawing on biological defence principles to address this gap.

The core originality of the work lies in systematically mapping human immune-system functions—such as innate and adaptive responses, anomaly recognition, memory, and learning—onto cyber-security mechanisms, moving beyond metaphorical analogy to operational design. Building on a structured review of current AI-driven cyber-security technologies, the project proposes a lightweight immune-inspired architecture and develops a minimal proof-of-concept demonstrating proactive “cyber immune” behaviour, such as automated anomaly detection or adaptive rule generation.

The prototype would be implemented using synthetic data and evaluated against simulated attack scenarios, enabling transparent assessment of technical feasibility, strengths, and limitations. The work demonstrates technical rigour through explicit design assumptions, controlled simulations, and critical evaluation, while maintaining clarity through a deliberately lightweight and interpretable implementation.

Beyond technical contribution, the project emphasises relevance to the wider cyber-security community by producing accessible outputs for small and medium-sized enterprises (SMEs) and community cyber-awareness groups.

These knowledge-exchange materials illustrate how emerging technologies may enable more resilient, adaptive, and proactive cyber defence in practice. This talk contributes original conceptual and practical insights into immune-inspired cyber-security, while fostering interdisciplinary dialogue at the intersection of AI, system design, and societal cyber resilience.

## Talk 3

**Title:** SABRE: An uncertainty-aware modular semantic–behavioural triage framework for adaptive ransomware detection

Henry Kabuye,

Tues 08 Sept
16:15 – 17:35

Ransomware defence in operational environments is not only a classification problem but also a decision problem: incorrect automated action can be costly, while uncertain cases increase analyst workload.

We present SABRE, an uncertainty-aware semantic–behavioural triage framework for adaptive ransomware detection. SABRE combines a semantic detector operating on Portable Executable-derived representations with a behavioural detector operating on time-window forensic telemetry. Each detector produces both a predictive score and an uncertainty estimate via Monte Carlo Dropout. These outputs are fused in a score-level decision layer and mapped to an interpretable triage policy supporting automatic containment, analyst escalation, or allowance.

The framework is evaluated on complementary ransomware datasets representing semantic and behavioural settings. Results show strong separation in the semantic setting and weaker separability in the behavioural setting, highlighting the practical importance of uncertainty-aware decision support when evidence quality is uneven. Rather than relying on point predictions alone, SABRE operationalises predictive uncertainty as a control variable for safer deployment. We further analyse model behaviour using feature importance and counterfactual explanations to support auditability. Overall, this work positions uncertainty-aware triage as a practical mechanism for reducing overconfident automated response in ransomware detection.

## Talk 4

**Title:** Agentic coding assistants: risks and mitigations

Mark Ryan

Tues 08 Sept
16:15 – 17:35

Large language models are increasingly deployed as autonomous agents - reading files, browsing the web, calling APIs and executing code on a developer's own machine. However, agentic coding assistants such as Claude Code expose a poorly understood new attack surface. We look at examples and mitigations.

The talk focusses on prompt injection: text from a webpage, email or repository that is processed by the agent may be interpreted as an instruction rather than data. This can compromise coding assistants, potentially leading to credential theft or poisoned code. We survey existing defences, and point out their weaknesses. The root of the problem is that transformers distinguish instructions from data only statistically, not architecturally. We hypothesise that a real architectural separation based on typed tokens, separate semantic and authority channels, or constrained attention is needed, and we speculate on how that might work.

# Technical Talks - Session 3: Cryptography, Privacy and Digital Authenticity

Talk 1

**Title:** One Proof, Many Implementations of Certified $\Sigma$-Protocols

Mukesh Tiwari

Tues 08 Sept
16:15 – 17:15

Sigma-protocols are a core building block in a wide range of privacy-preserving applications, but one particular application where they have found a widespread adoption is electronic voting. They remain a popular choice in electronic voting to maintain privacy and verifiability of elections and are still used in prominent electronic voting systems such as Helios, Belenois, and SwissPost. However, existing implementations of Sigma-protocols have two critical shortcomings: (i) the same code must be written multiple times, once for the front-end (JavaScript/TypeScript) and once for the back-end (Python/OCaml/Java), which is error-prone because the front-end and back-end codes are often written by different developers, leading to inconsistencies and bugs, and (ii) none of these implementations provide formal security guarantees.

In this talk, I will present a certified formalisation of a comprehensive library of Sigma-protocols in the Rocq theorem prover. My generic formalisation over abstract vector spaces enables me to uniformly express and compose a wide class of discrete-logarithm-based Sigma-protocols, while remaining independent of any concrete group or cryptographic instantiation. Moreover, the constructive nature of my formalisation allows me to extract and compile executable implementations in OCaml, Rust, and WebAssembly, thereby eliminating code duplication. I demonstrate the practical applicability of my framework by implementing a voting server and client for Approval voting and an IACR tallysheet verifier.

## Talk 2

**Title:** Privacy of Synthetic Data Generation via Voice Cloning

Marios Aristodemo,

Siamak Shahandashti

Tues 08 Sept

16:15 – 17:15

Voice is a data modality that contains a very high amount of personal information. An emerging and widely-used technology is voice cloning which poses the risk of violating the privacy of speakers due its capability of generating synthetic audio samples from human speech recordings. Despite recent advances in the design of cloning and anonymisation techniques for reduction of private information encoded in the synthetic voice, privacy concerns regarding speaker re-identification and attribute inference remain relatively under-explored.

Furthermore, it is unclear what the best configuration is to combine cloning and anonymisation in an end-to-end pipeline. We systematise and extend existing attacks on synthetic voice data and provide precise definitions of privacy metrics based on a set of “privacy games” including de-anonymisation, linkability, membership and attribute inference. Based on these definitions, we design a framework for evaluating the utility–privacy trade-off of voice cloning and anonymisation pipelines, and report on a set of initial results on evaluating several combinations of cloning and anonymisation techniques. More specifically, we consider two main pipeline configurations based on the order in which cloning and anonymisation are applied, two popular voice cloning systems, and five state-of-the-art anonymisation methods from the Voice Privacy Challenge.

Our results show that anonymisation before cloning provides better utility in general, and while some techniques provide promising results in withstanding de-anonymisation and inference attacks, linkability attacks achieve high success rates against almost all techniques.

Talk 3

**Title:** Combating fake media files with edit-tolerant signatures

Fentham Dan,
David Oswald,
Mark Ryan,
Xiao Yang

Tues 08 Sept
16:15 – 17:15

With the increasing usage of fake videos in misinformation campaigns, proving the provenance of a video is important. We are developing methods that allows a social media platform to incorporate a "Verify" button that would allow a user's device to verify the authenticity of a video with respect to a signature made by the recording device, even if the publisher has edited the video before it was published.

The publisher may want to redact confidential parts of the video, e.g. by blurring the faces of people not relevant to the content, or by cropping the video. We assume the recording device signs the video content, together with metadata including the recording time and GPS location. We assume that the recording device has a signing key, certified by an appropriate certificate authority. The requirement to support editing means that a simple digital signature on the video file and the metadata won't work.

In our method, the recording device divides the video into blocks. For each block, it computes a discrete cosine transform (DCT) from the pixel values in the block. It then truncates or quantises the DCT (as done in JPEG encoding), and signs hashes derived from it. A publisher can edit the video and the signatures; the signed material for blocks that are edited will be eliminated from the data that the user receives, ensuring confidentiality of the redacted material.

# Technical Talks - Session 4: Securing Connected Critical Systems:

## Talk 1

**Title:** Security Assurance of Semiconductor Manufacturing (SASM)

Sridhar Adepu

Weds 09 Sept
10:30 – 11:50

This talk is to present our EPSRC funded project on supply chain security in semiconductor design and manufacturing processes.

To advance the state of practice for semiconductor security, we jointly develop a novel assurance framework to help equipment manufacturers (supplying semiconductor fabs) manage security risks across their supply chains. The framework encompasses an auditing regime based on measures

in compliance with the NIST Cybersecurity Framework (CSF) 2.0. The framework allows for varying levels of security assurance, ranging from self-disclosure to external auditing, specifying appropriate assessment methods for each of the CSF functions of govern, identity, protect, detect, respond and recover.

To facilitate knowledge exchange, we conducted a series of workshops in the UK and Germany by engaging with the electronics manufacturing industry (including both providers of equipment and services) to assess the state of supply chain security for the sector. Conducted jointly with researchers from SEMI, Bavarian Chip Alliance, and supported by Techworks UK.

**Approach and Methodology**

We developed the assurance framework by combining emerging industry practices with the latest research literature. This is achieved by catalogue of evidence that would be used to inform the state of the art in assurance methods across semiconductor manufacturing, security assurance methods employed in other sectors such as Critical National Infrastructure (CNI) including data centers, and would include technical documentation, configuration, inspection, and testing and evaluation for validation.

**Validating the assurance framework**

The workshops areconducted as focus groups designed to collect feedback data input to assess the feasibility of the assurance framework, along with an understanding of current preparedness of the UK ecosystem.

## Talk 2

**Title:** Emanations Protection

Ian Bryant,
James Flint'
Sharon Wiltshire

Weds 09 Sept
10:30 – 11:50

Cyber Security is an element of a wider set of Protections dealing with all aspects of Confidentiality, Integrity, and Availability Risks, and there are inevitable intersections between these various Dimensions. Defence is leading on work to define and explore the issue of Emanations Protection, which is a superset of what HM Government refers to as Technical Security, encompassing Signals of all varieties (Electromagnetic, Magnetic, Acoustic, and Quantum), and risk vectors (e.g. outbound interception, and inbound disruption). The briefing provides an introduction to Emanations Protection, to stimulate consideration and discussion of ways in which the intersections of interests can be further explored.

## Talk 3

**Title:** Zero Trust Architecture in Cloud-Native Robotics: Fragmentation, Gaps, and Opportunities

Abdullahi Arabo,

Jacob Dufossé

Weds 09 Sept

10:30 – 11:50

Modern robotic systems—including surgical assistants, autonomous logistics fleets, and infrastructure-monitoring drones—are increasingly deployed as components of distributed, cloud-native architectures. These systems are commonly orchestrated using Kubernetes, communicate via ROS 2 and its DDS middleware, and integrate with cloud services for telemetry, updates, and remote operation. Although this transition is not yet universal, the shift towards cloud-native robotics is well established and accelerating.

It also introduces an unprecedented attack surface spanning multiple layers, from cloud APIs and container registries to internal buses and physical sensors.

This work presents a Systematic Mapping Study examining how Zero Trust Architecture (ZTA) principles, as defined in NIST SP 800-207, are being applied to this emerging class of cyber-physical systems. The review synthesises evidence from 21 studies (10 primary and 11 secondary) and reveals a paradox of fragmented maturity. While individual ZTA components have been successfully demonstrated in isolation—such as certificate-based identity in SROS2, real-world zero-day detection via ROBOCOP, and Kubernetes-level anomaly detection using VISOR-ZT—no study assembles these elements into a coherent, end-to-end ZTA implementation suitable for robotic deployments.

The study contributes three outputs. First, an eight-pillar taxonomy that extends NIST ZTA to address robotic-specific constraints, including physical-layer security, DevSecOps integration, system resilience, and regulatory compliance. Second, a ZTA maturity matrix derived from the primary studies, classifying existing approaches across five readiness levels, from Deployed to Open Gap. Third, an inventory of ten open research gaps, including ZTA enforcement under disconnected operation, structural tensions between DDS multicast discovery and Kubernetes network isolation, and the complete absence of an integrated ZTA stack for cloud-native robotics.

This talk targets security practitioners and cloud architects working with connected robotic systems, offering a practical reference map for a domain not yet recognised as a distinct Zero Trust application area.

Talk 4

**Title:** Behind the Wheel: Trust, Risk and Vulnerabilities in Connected and Autonomous Vehicles

Adam Gorine,
Priya Gopal

Weds 09 Sept
10:30 – 11:50

Modern cars have advanced in technology, driving experience, and communication systems, offering drivers a new level of comfort and a better sense of connection to their surroundings.

This paper explores vulnerabilities and cyber threats in both in-vehicle and vehicle-to-everything communication. In-vehicle communication relies on the Controller Area Network (CAN) bus. The CAN bus enables efficient real-time communication between Electronic Control Units (ECUs), managing and coordinating critical subsystems such as braking, engine control and steering. However, despite its robustness and widespread use, the CAN bus remains vulnerable to attacks such as data injection, spoofing, and denial-of-service (DoS), which can compromise the safety and security of drivers and passengers. These in-vehicle vulnerabilities stem from a lack of authentication and encryption (Tanksale, V. 2024).

Vehicle-to-everything communication enables vehicles to interact with their surrounding environment. These technologies include wireless sensor communications such as LiDAR, radar, and cameras, as well as the Global Navigation Satellite System, allowing real-time data exchange and improved coordination with other vehicles and infrastructure. However, new threats have emerged that could undermine the functionality, safety, and security of these vehicles, including message spoofing, privacy leakage, denial-of-service attacks, and infrastructure compromise (Philipsen, S.G., 2022)

This research paper analyses cyber threats affecting the transport system and suggests new approaches to defend and secure modern and future autonomous vehicles. Ensuring trust behind the wheel requires not only technological innovation but also integrating cybersecurity as a fundamental component of the vehicle transportation systems.

# Technical Talks - Session 5: Cyber Risk, Deception and Societal Resilience

## Talk 1

**Title:** Cyber risk quantification: Markets, incentives and failures

Edward Cartwright,

Jacob Seifert,

Anna Cartwright

Weds 09 Sept

14:45 – 15:45

Cyber risk quantification is a fundamental process to help to analyze, understand, and manage cyber risk, whether at the level of an asset, organization, or wider society. In recent years there has been a rapid expansion in the number of cyber risk quantification tools and methods. These range from free, open-source options developed in academia to proprietary instruments developed in the private sector. There is also a growing market for consultancy services to advise on and implement cyber risk quantification.

In recent research we have surveyed the cyber risk quantification landscape and identified a number of potential failures in the current cyber risk quantification marketplace. We will focus the discussion on two inter-related challenges:

(a) Adverse selection is likely in the cyber risk quantification marketplace because of a lack of common information on how different quantification tools work. This can stifle innovation and lead to sub-optimal quantification tools dominating the market. This can be exacerbated by information cascades.

(b) Commonality across cyber risk quantification tools and methods can be exploited by cyber attackers resulting in substantial aggregate, potentially systemic, risk. This is because 'omissions' or 'bias' in the tools can be amplified by common use within and across sectors.

We will explore potential policy levers to alleviate these market failures. Crucial is a fuller, objective understanding of the limitations of different cyber risk quantification methods and tools and the differences between them. Policy needs to encourage innovation while also enabling more transparency within the marketplace.

Talk 2

**Title:** Disinformation Resilience among Cyber Security, AI and Information Warfare Experts: Findings and Call to Action

Allison Wyld,

Fabio James Petani

Weds 09 Sept

14:45 – 15:45

Experiences of disinformation from a recently published survey of professionals working in cyber security, artificial intelligence, and information warfare (military and civilian n=98)[1] are reported together with a call for future research. Resilience to disinformation is conceptualised here as the capacity to resist, detect, and mitigate false information, encompassing both cognitive and behavioural dimensions. Victimisation findings (n=81), showed 45.7% of respondents acknowledged being deceived into sharing information, 25.9% were uncertain of being deceived, and 28.4% had never been deceived (total = 100%).

The most prevalent countermeasure strategy was double-checking sources before sharing (53.1%), increased skepticism toward unknown sources (13.6%), use of fact-checking tools (11.1%), avoiding emotionally charged content (9.9%), peer validation (3.7%) other strategies (7.4%), and formal cognitive security training (1.2%) (total = 100%). Encounter frequency to disinformation, 69.2% of respondents, occasionally to very frequently (very frequently 14.8%, frequently 19.8%, occasionally 34.6%), while 30.8% encounter rarely or never. Detection confidence, 33.3% expressed high or very high confidence, 49.4% moderate confidence, and 17.3% reported neutral or low confidence.

These findings suggest that even domain experts self-report vulnerability to disinformation campaigns. The key contribution proposes a four-factor model of expert vulnerability, synthesising; information overload, trust network exploitation, adversarial sophistication, and confirmation bias. Implications highlight the need for enhanced cognitive security training and the deployment of institutional technical support mechanisms to counter disinformation. The escalating scale of threats from disinformation demands future scholarly attention. Promising directions include, addressing current limitations of self-reported data, increasing the sample sizes, including comparison groups[2] and progress to inferential analysis of key drivers to strengthen explainability.

[1], [2], references can be supplied.

## Talk 3

**Title:** Characterizing Cross-Border Scam-Driven Human Trafficking and Online Community Responses

Jingjie Li

Weds 09 Sept
14:45 – 15:45

A new form of human trafficking has emerged across Chinese borders, where individuals are lured to Southeast Asia with fraudulent job offers and then coerced into operating online scams.

By 2025, this crime ecosystem was estimated to involve more than 300,000 people trafficked into scam compounds and to cause annual losses of up to 64 billion US dollars. It has also become increasingly globalised, affecting large numbers of scam victims in Western countries, including the UK. Despite its massive economic and human toll, this scam-driven trafficking is underexplored in academic research.

In this talk, we will present our work and findings that uncover this topic. Through qualitative analysis of posts on RedNote, a major social media platform in China, we examined how Chinese online communities respond to this threat.

Our findings reveal that perpetrators exploit cultural ties to recruit victims for cybercriminal roles within self-sustaining compounds, using sophisticated manipulation tactics. Survivors face serious reintegration barriers, including family rejection, as the cultural values that enable trafficking also hinder their recovery.

While communities present protective strategies, efforts are complicated by doubts about the reliability of support and cross-border coordination.

We will further discuss key implications and collaboration opportunities for prevention, platform governance, and international cooperation that are relevant to the UK communities and stakeholders working to counter scam-driven trafficking.

# Technical Posters

## Poster Title: Promoting Cyber Security Awareness: Dice and Easy

**Steven Furnell, James Todd, Lucija Smid, Xavier Carpent and Simon Castle-Green**

Gamification is frequently suggested as a means of assisting cyber security awareness, with physical games having proven popular as a means of separating the lessons from the technology. However, many resulting games have also prove to be relatively time-consuming to learn and play, as well as often requiring some prior cyber knowledge on the part of the players.

Cyber Defence Dice is one of a series of game-based activities that the authors have developed as short-format provocations of interest that help to promote basic cyber security awareness. In this game, players (individually or in small groups) play the roles of attackers and defenders, with the threats (e.g. hackers, malware, and phishing) and safeguards (e.g. backups, updates, and secure configuration) represented as faces accompanying sets of coloured dice. The selected elements are based upon themes that are relevant and/or recognisable to individuals end users in both personal and workplace contexts, and so the game (and its resulting awareness-raising value) can be beneficial at several levels.

The game has been extensively play-tested with a range of audiences (including a variety of lay users, as well as more specialist groups including cyber students, professionals and educators). It has received positive feedback in all cases, with players considering it to have value as an engagement and awareness tool, while also being easy to understand and enjoyable to play. An expanded version of the game, in which players compete to protect and acquire a series of assets has also been developed for those that desire a longer game with additional security objectives.

The event will offer an opportunity to learn about the game, as well as a chance for some hands-on play.

## Poster Title: Mathematical Perspectives on Deception in AI-Mediated Cyber Security

**Abdullahi Arabo**

As AI systems increasingly interact directly with humans, deception is no longer only a social concern—it represents an emerging, quantifiable cyber-security risk. As artificial intelligence mediates an increasing proportion of digital interactions, the boundary between human cognition and computational decision-making is rapidly blurring. This poster addresses a timely and relevant challenge for the cyber-security community: understanding how deceptive behaviours can emerge in AI systems, particularly large language models (LLMs) and generative agents, and how these behaviours reshape contemporary attack surfaces.
The work adopts a technically grounded yet accessible approach, drawing on concepts from game theory, information theory, and adversarial machine learning to examine how deception can arise implicitly through optimisation processes, rather than explicit programming. We analyse representative cyber-security scenarios—including phishing detection evasion, synthetic media generation, and automated social-engineering simulations—to illustrate how AI systems may model, predict, and influence human behaviour by exploiting cognitive biases and social heuristics. These cases demonstrate a shift in cyber risk from predominantly technical infrastructure toward human trust and perception.

The poster further contributes original and interdisciplinary insight by framing deception and trust as properties that can be formally characterised and reasoned about within human–AI interaction. We critically discuss the strengths and limitations of mathematical abstractions in this context, and outline potential safeguards that could be integrated into learning and optimisation pipelines to mitigate misuse.

Overall, this work offers a clear, concise, and exploratory contribution that combines mathematical reasoning with practical cyber-security relevance, aiming to stimulate discussion on the detection, characterisation, and governance of AI-enabled deception in future cyber-defence strategies.

## Poster Title: A User-Centric Workbench for AI Auditing, Integrating Stakeholders in the Auditing Process

**Marios Aristodemou, Vyron Christodoulou and Siamak Shahandashti**

As AI systems increasingly impact society, recent legislation and frameworks demand rigorous auditing to ensure the safety, security, and trustworthiness of AI applications. However, current auditing practices are mostly confined to technical tests and hence they suffer from a blind spot where end user and domain experts' opinions are not being considered.

This approach may overlook real-world harms and biases of AI applications because it excludes the lived experiences, domain knowledge, and opinions of the very people affected such as the end-users and decision-subjects. For AI to be truly responsible and accountable, auditing must move beyond technical experts and actively integrate public perspectives. Thus, the concept of participatory AI auditing should be developed, and fill the gap of practical tools designed to help non-AI-experts (which may well be experts in other domains) systematically audit AI applications.

As part of the PHAWM project (https://phawm.org), an auditing methodology and a workbench have been co-designed to address this gap by exploring how end-users and decision-makers can be effectively integrated into the auditing process to uncover previously unknown risks. Our tool aims to provide actionable insights into how best to support end-users in the AI auditing process, ensuring that AI evaluations reflect the diverse voices of society.

Prior to the development of the tool, we have derived a comprehensive set of functional and non-functional requirements for participatory auditing tools based on empirical work with stakeholders. We present the user-centric design of the PHAWM Workbench, a novel open-source platform that empowers non-technical stakeholders to evaluate AI impacts and audit AI applications based on their domain knowledge. The Workbench serves as a step toward collaborative AI oversight, ensuring that the technology shaping our society is held accountable by the society it serves.