Cyber Science 2026

Cyber Science in the era of Artificial Intelligence

Royal Holloway University of London, London, UK

JUNE 3 – 5, 2026

@cmricorg #Cyberscience2026

www.c-mric.org

Our Sponsors

Cyber Science 2026

Cyber Science is the flagship conference of the Centre for Multidisciplinary Research, Innovation and Collaboration (C-MRiC) focusing on pioneering research and innovation in Cyber Situation Awareness, Social Media, Cyber Security and Cyber Incident Response.

Springer is the technical co-sponsor and publisher of the 2026 Cyber Science proceedings that will be published in the Springer Proceedings in Complexity book series.

The conference encourages participation and promotion of collaborative scientific, industrial and academic inter-workings among individual researchers, practitioners, members of existing associations, academia, standardisation bodies, and including government departments and agencies. The purpose is to build bridges between academia and industry, and to encourage interplay of different cultures. It is a platform for researchers and industry practitioners to present work encompassing principles, analysis, design, process, implementation, methods and applications.

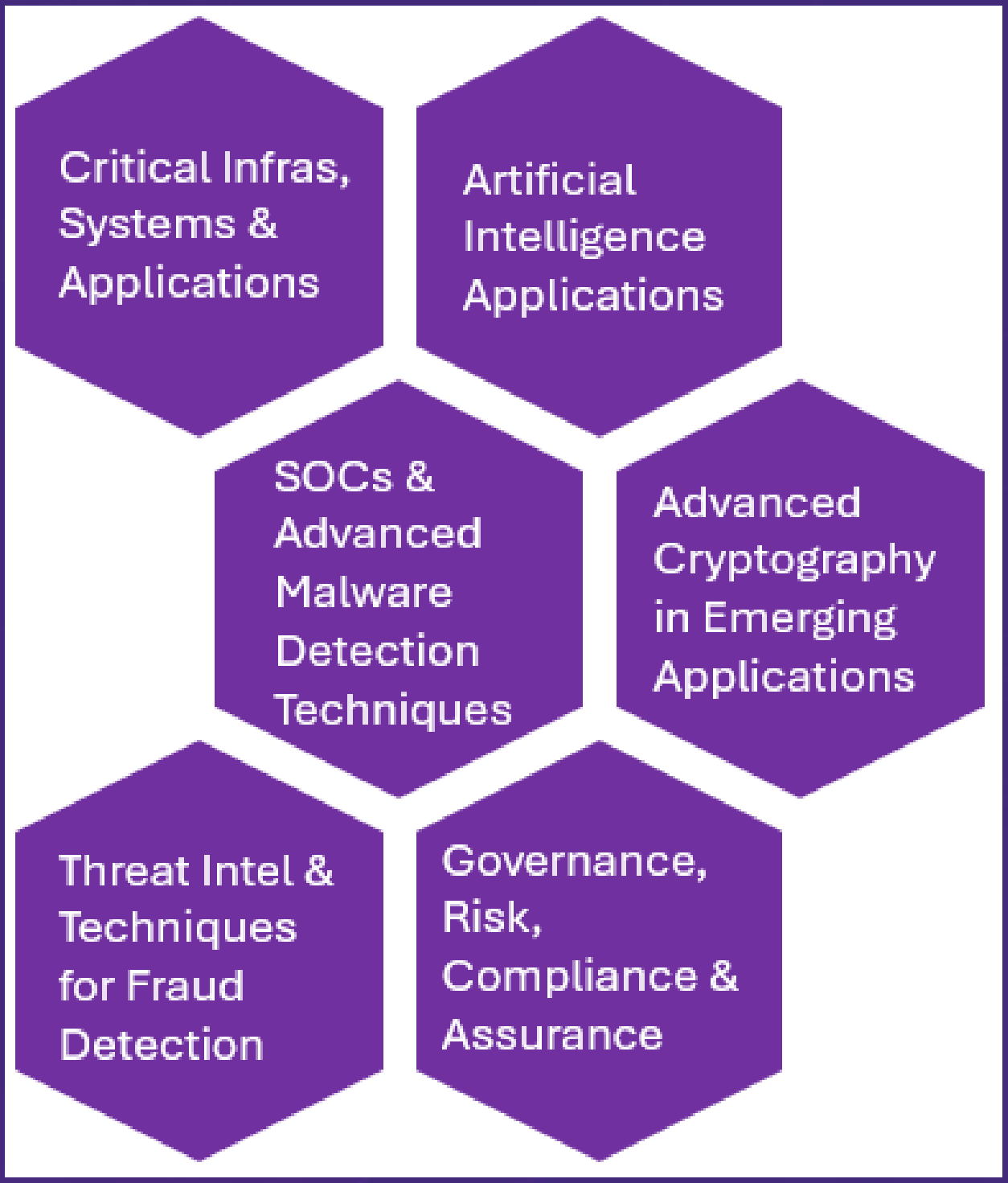
As a yearly event, it's held at various cities. The first three meetings (2015-2017) were in London, followed by Glasgow, Scotland in 2018, University of Oxford in 2019, Dublin City University, Dublin, Ireland in 2020 & 2021, Cardiff Metropolitan University, Wales, in 2022, University of Aalborg, Copenhagen, Denmark in 2023, Edinburgh Napier University, Scotland in 2024 and back again in London this year at the Royal Holloway University of London, London, UK.

Theme

Cyber Science in the ear of Artificial Intelligence

CyberScience 2026

Themes



Contents

CYBER SCIENCE 2026 CONFERENCE PRESENTATION TIMETABLE
TIME IN UK (GMT +1)

CYBER SCIENCE 2026 CONFERENCE INFORMATION

Wednesday, June 3rd - Friday, June 5th

Conference Venue

Royal Holloway University of London, **Stewart House via Senate House**, London, UK

At Royal Holloway, we're proud of our history. Brave from the beginning, we were founded over 170 years ago by two Victorian social pioneers who wanted to make a difference.

Royal Holloway is formed from two colleges:

- Bedford College, founded by **Elizabeth Jesser Reid** and
- Royal Holloway College, founded by **Thomas and Jane Holloway**.

These colleges were among the first places in Britain where women could access higher education, so our rich history is deeply rooted in providing equity in opportunity, transforming lives through education and creating positive change.

Bedford College in central London opened its doors in 1849. It was the first higher education college in the UK for women and counts the novelist George Eliot and the first female doctor, Dr Elizabeth Blackwell, among its early students.

In 1886, Royal Holloway College in Surrey was opened by **Queen Victoria**. By 1900, the colleges became part of the University of London, later merging in 1985 to form what is now Royal Holloway.

For more details https://www.royalholloway.ac.uk/about-us/our-history/

Location

Stewart House via Senate House

Malet St, London WC1E 7HU

United Kingdom

Conference Structure/ Organisation

- The conference is hybrid. In-person-only June 3-4 and Remote-only Friday June 5.
- There are a total of 3 keynote speakers and 32 paper presentations.
- The Conference proceedings will be opened by Dr Cyril Onwubiko, founder, Centre for Multidisciplinary Research, Innovation & Collaboration (C-MRiC) on Wednesday 3rd June 2026 at 09:30 GMT+1.
- The first keynote session will take place on 3rd June 2026 at 10:40 GMT+1 and will be presented by Dr Carolyn Swinney, Royal Air Force Communications & Electronics Eng. Officer, and Executive Fellow, University of Essex.
- The second keynote session to be delivered by Dr Konstantinos Mersinas, Associate Professor, Information Security Group, Royal Holloway University of London, UK at 09:00 GMT+1 on 4th June 2026.
- The third keynote session will be delivered by Dr Deepthi Ratnayake, Principal Lecturer, Cyber Security, Cybersecurity & Computing Systems Research Group, University of Hertfordshire, UK on Friday 5th June 2026.
- Wednesday and Thursday (June 3-4) are in-person ONLY, while Friday, June 5 is Remote-only conference (see Timetable, pp. 37).
- Dinner and drinks on Wednesday 3rd June evening, 18:00 – 21:00 at the Antalya Restaurant, London, UK https://antalyarestaurant.co.uk
- The Conference proceedings will be closed by Professor Martin Gilje Jaatun, SINTEF ICT, and Chair of the Cyber Science 2026 conference.

CYBER SCIENCE 2026 KEYNOTE SPEAKERS

Keynote Speakers

Dr Carolyn Swinney

Dr Carolyn Swinney is a Royal Air Force Communications & Electronics Engineering Officer and Executive Fellow at the University of Essex. She earned a BEng (Hons, First Class) in 2007 and an MSc (Distinction) in Electronics Engineering from the University of Essex in 2013, qualifying as an RAF Communications and Electronics Engineering Officer in 2014. In 2023 she completed a PhD in Electronics Engineering, specialising in deep-learning models for radio-frequency signal classification, with applications to counter drone systems. Carolyn serves in the RAF's cyberspace profession and is currently attending the Advanced Command and Staff Course at the Defence Academy of the United Kingdom. Her research interests include signal processing, autonomous systems, machine learning, and cyber security.

Dr Konstantinos Mersinas

Dr Konstantinos Mersinas, PhD, CISSP, is an Associate Professor at the Information Security Group, Royal Holloway, University of London, and Visiting Professor at Keio University, Tokyo, Japan. Konstantinos' research lies with human and behavioural aspects of cybersecurity, maritime security, and cybercrime. He has advised the UK All-Party Parliamentary Group (APPG) on Cybersecurity, the UK Fraud Act and Digital Fraud Committee, and a number of UK Government Departments. He co-founded the research group HIVE (Hub for Interdisciplinary research into Vulnerability to Exploitation). Konstantinos collaborates with the NATO Cooperative Cyber Defence Centre of Excellence (https://ccdcoe.org/) in Tallinn, Estonia and he is Director at the International Cyber Security Centre of Excellence (https://incs-coe.org), an international community founded between UK, USA and Japan, promoting cybersecurity research globally

Dr Deepthi Ratnayake

Dr. Deepthi Ratnayake is a Principal Lecturer in Computer Science (Cyber Security and Network), within the School of Physics, Engineering and Computer Science (PECS) She has nearly 30 years of experience in industry, defence and academia in the lines of cyber security, networking, and information systems management. She is a research active academic that enjoys cross-discipline collaborations and teamwork.

Researchers and PhD research students who are interested in exploring following areas to further develop the research to the next levels of novel methods and strategies are welcome; Network-based Intruder Detection Systems (NIDS). Host Based Intruder Detection Systems (HIDS). Application of Big Data Analytics Technologies for Intruder Detection. Cyber-Physical-System (CPS) based intrusion detection systems (includes robots and unmanned/self-driving vehicles) in are welcome. Please get in touch via d.ratnayake@herts.ac.uk.

Research profile:
https://researchprofiles.herts.ac.uk/en/persons/deepthi-ratnayake/

Organisers & Chairs

Professor Aunshul Rege

Dr. Aunshul Rege is a Full Professor in the Department of Criminal Justice at Temple University, where she directs the Cybersecurity in Application, Research, and Education (CARE) Lab. Her research has been funded by several National Science Foundation (CAREER, EAGER, CPS, SaTC EDU) and Department of Energy/Idaho National Lab grants. Her work focuses on critical infrastructure and cybersecurity, cyber adversarial decision-making and adaptation, ransomware, social engineering, and cybersecurity education. She is the organizer and host of the summer social engineering competitions for high school, undergraduate, and graduate students. Her cybersecurity awareness and training efforts extend beyond higher education to include working with youth, elderly, and previously incarcerated individuals via partnerships with local nonprofits.

Dr. Rege has a B.Sc. (2002) in Computer Science from the University of British Columbia and worked for two years as a software engineer. She also holds a B.A. (Hons.) (2006) and M.A. (2008) in Criminology from Saint Mary's University in Halifax, Nova Scotia. She completed her M.A. (2010) and Ph.D. (2012) in Criminal Justice from the Rutgers School of Criminal Justice. She has been featured on BBC World Service, WHYY/ PBS/NPR's Studio 2, the David Bombal's show, Technical.ly, and the BBC/CBC Podcast "Love, Janessa", and her work has been recognized in Security News, Dark Reading, and AARP to name a few. She currently serves as the Research Lead for the Social Engineering Community at Defcon. She also serves on the Advisory Board of Raices Cyber and Black Girls Hack

Dr Pierangelo Rosati

Dr Pierangelo Rosati is Associate Professor in Digital Business and Society at University of Galway and the Business Community Lead of the IEEE UK & Ireland Blockchain Group. Dr Rosati specialises in measuring the business value of digital technologies and his research has been published widely including European Accounting Review, Computers and Human Behaviour, New Media & Society, European Journal of Finance, International Review of Financial Analysis, Information Technology & People, JASIST and others. He is also a Series Editor on the Palgrave Studies in Digital Business & Enabling Technologies. Dr Rosati previously worked as an Assistant Professor in Business Analytics at DCU Business School where he was Co-Deputy Director of the Irish Institute of Digital Business (IIDB). Dr. Rosati holds a PhD in Accounting and Finance from the University of Chieti-Pescara (Italy), and an MSc and a BSc in Management and

Dr Hanan Hindy

Hanan Hindy, Ph.D. is currently a Lecturer (Assistant Professor) at the Computer Science Department, Faculty of Computer and Information Sciences, Ain Shams University in Egypt. Hanan received her doctorate from the Division of Cyber-Security at Abertay University, Dundee, Scotland. She received her bachelor's degree with honours (2012) and masters (2016) degrees in Computer Science from the Faculty of Computer and Information Sciences at Ain Shams University, Cairo, Egypt. Her research interests include Intrusion Detection Systems, Computer and Network Security, Artificial Intelligence, and Deep Learning.

Hanan won the IEEE Women in Engineering "Excellence in Engineering" Award in 2021.

You can reach Hanan at https://hananhindy.com

Dr Arnau Erola

Dr Arnau Erola is a Senior Security Consultant at the NCC Group, focusing on Web Application Security, Vulnerability Scanning, Cloud and Security Engineering.

Prior to joining the NCC Group in 2022, he worked at the Computer Science Department, University of Oxford, specialised in cyber security, data analytics, machine learning, data mining and information privacy.

He is a Research Fellow at CyberSecurity@Oxford at the University of Oxford, working on enterprise security, defence systems and better understanding the cyber-threat landscape. Within his portfolio, Arnau has engaged with several UK authorities, determining their needs and providing state of the art innovative solutions. Dr Erola holds a Ph. D., M. Sc. and B.Sc. in Computer Science from the Rovira i Virgili University of Tarragona (URV). He is author of several international journal articles on online privacy, anonymity protocols and intrusion detection mechanisms.

Dr Xavier Bellekens

Dr Xavier Bellekens is the CEO of Lupovis.io, a spinout company of the University of Strathclyde focusing on dynamic cyber-deception, a non-resident senior fellow with the Atlantic Council's Cyber Statecraft Initiative within the Scowcroft Center for Strategy and Security and a Chancellor's Fellow, Lecturer with the Department of Electronic and Electrical Engineering at the University of Strathclyde. He is also the Chair of the blockchain group and the Vice-Chair of the cyber-security group for IEEE UK and Ireland. He also has over 10 years of experience in consulting across public and private sector. His experience spans from cyber-defence, deception, deterrence and attribution of cyber-threats in critical infrastructures to cyber-situational awareness and cyber-diplomacy and frequently appears in the media to provide commentary to international press – on radio, tv and newspapers on major cyber-events

Professor Martin Gilje Jaatun

Professor Martin Gilje Jaatun is a Senior Scientist at SINTEF Digital in Trondheim, Norway and adjunct professor at the University of Stavanger. Formerly, he was Editor-in-Chief of the *International Journal of Secure Software Engineering* (IJSSE). Previous positions include scientist at the Norwegian Defence Research Establishment (FFI), and Senior Lecturer in information security at the Bodø Graduate School of Business. His research interests include software security, security in cloud computing, and security of critical information infrastructures. Dr. Jaatun graduated with a sivilingeniør degree in telematics from the Norwegian Institute of Technology, and received the Dr.Philos degree from University of Stavanger. He is Vice Chairman of the Cloud Computing Association, an IEEE Cybersecurity Ambassador, an IEEE Computer Society Distinguished Visitor, and a senior member of IEEE.

Dr Cyril Onwubiko

Dr Cyril Onwubiko is a cybersecurity executive, author and founder. He serves as Director and Business Information Security Officer at Barclays, where he leads cybersecurity for the Cross-Barclays Digital Platform (XDP), a bank-wide ecosystem of digital products and platforms. Before joining Barclays, he served as Senior Director of Enterprise Security Architecture at Pearson, a global leader in education. He has held senior leadership roles across financial services, telecommunications, government, and public-sector organisations.

Dr Onwubiko is an award-winning professional recognised for advancing global computing education, curriculum innovation, and academia–industry collaboration. He is a recipient of the 2025 Society and Council Professional Development Award, as well as the IEEE Computer Society Golden Core and Distinguished Contributor Awards, in recognition of his technical excellence and lifelong service.

CYBER SCIENCE 2026
ACCEPTED PAPERS

Accepted Papers

Cyber Science 2026 • 3–5 June 2026 • London, UK

32 accepted papers

1. Artificial Intelligence as a Force Multiplier in Social Engineering: An Experiential Learning Approach

Aunshul Rege (Temple University); Hwanhee Park (Temple University); John Snee (Temple University)*

Artificial Intelligence (AI) is reshaping social engineering (SE), enabling adversaries to generate convincing pretexts, phishing emails, and vishing scripts at scale. While technical defenses have advanced, the human element of cybersecurity remains underexamined. This paper reports on the 2025 Social Engineering Penetration Testing Competition, which integrated AI into a multiphase adversarial simulation spanning open-source intelligence (OSINT), persona and pretext development, bid proposal creation, phishing, and vishing. Forty teams across high school, undergraduate, and graduate levels participated under strict ethical and governance protocols. Findings highlight AI's dual role: it accelerates the production of realistic artifacts but requires human oversight for plausibility, accuracy, and responsible use. Using Experiential Learning Theory (ELT) as the pedagogical framework, this study illustrates how students engaged in all four stages of experiential learning—concrete experience, reflective observation, abstract conceptualization, and active experimentation—through the competition. Student reflections reveal both the technical and professional growth resulting from participation, as well as the challenges of integrating AI responsibly into adversarial simulations. Findings indicate that AI operated as a force multiplier for learning: it accelerated drafting and ideation, yet students' success hinged on human judgment, adaptation, and ethical reasoning. Participants demonstrated measurable growth in risk articulation, persona credibility, professional writing, and agile communication, and they developed AI literacy by editing, validating, and justifying model outputs. Mapping tasks to the NICE Cybersecurity Workforce Framework made workforce-relevant knowledge, skills, and tasks visible and transferable. The paper concludes with implications for designing experiential cybersecurity education and workforce development initiatives.

2. PhishShield: A Multi-Modal Gmail Security Extension Using Machine Learning and BERT for Phishing Detection

Tharun M R (K.S.Rangasamy College of Technology); Pugalendi S (K.S.Rangasamy College of Technology); Dinesh Kumar K (K.S.Rangasamy College of Technology)*

PhishShield, my project, is a program to identify phishing systems that are designed to protect users against the internet frauds. PhishShield works through a multi-modal approach that analyses URLs and screenshot of websites unlike the traditional methods that only test the URLs of websites [7]. It uses OCR together with BERT a transformer-based natural language processing model [1], to read and understand hidden or deceptive text in images, and machine learning to detect suspicious URL templates. A sophisticated decision engine that generates a score of confidence in detecting phishing is a combination of all these insights. Subsequently, the system assists the consumer in knowing why they have received a timely warning of the site; it alerts the consumer using explainable and real-time signals [4], [9] on the mobile

devices and browsers.

3. An Artefact-Driven Framework for Embedding Security by Design in the Systems Engineering V-Model

*Enoch Agyepong (Elbit Systems, UK)**

As engineered systems grow in complexity and interconnectivity, the need to demonstrate how security is embedded in the system development lifecycle has intensified. Traditional systems engineering frameworks such as the V-Model emphasize functional traceability and verification but fall short in evidencing security integration. This paper proposes a structured, artefact-driven framework that aligns security artefacts with each phase of the systems engineering V-Model. Drawing on academic literature and international standards such as the NIST SP 800-160, ISO/IEC 27034, ENISA, and OWASP SAMM, the paper consolidates 32 security artefacts and maps them to the engineering lifecycle. These artefacts can be used as verifiable and auditable evidence of security consideration and integration to support regulatory compliance and governance without disrupting established workflows.

4. Characterizing Vulnerabilities in the TCP/IP Protocol Stack Across Packet Semantics and Control Plane Architectures

*Omkar Bhalekar (RIT)**

The Internet Protocol (IP) protocol suite is a basic communication framework that constitutes today's most used networks; it was established, though, well before advanced levels of security measures were a particular necessity. This makes today's IP protocol suite highly insecure against a variety of different security attacks that might range from simple password "sniffing" attacks to entire DoS attacks. Another factor that has immensely increased levels of vulnerability for such security attacks is that tools for such attacks are widely available on the Internet. This paper will discuss some basic levels of vulnerabilities that exist within today's widely used TCP/IP protocol suite, mainly focusing on understanding how such vulnerabilities could be used by different attackers. Some of today's most basic well-known attacks against an IP protocol suite will also be discussed within this paper, including SYN flooding attacks, Ping of Death, that take place due to basic design vulnerabilities within these protocols.

5. WiFiPenTester: Towards Governed GenAI-Assisted Wireless PenTesting

Haitham Al-Sinani (RHUL); Chris Mitchell (Royal Holloway, University of London); Abdulaziz Al-Hosni (Diwan of Royal Court); Sultan Al-Harrasi (University of Technology and Applied Sciences)*

Wireless ethical hacking relies heavily on skilled practitioners manually interpreting reconnaissance results and executing complex, time-sensitive sequences of commands to identify vulnerable targets, capture authentication handshakes, and assess password resilience; a process that is inherently labour intensive, difficult to scale, and prone to subjective judgement and human error. To help address these limitations, we propose 'WiFiPenTester', an experimental, governed, and reproducible system for GenAI-enabled wireless ethical hacking. The system integrates large language models into the reconnaissance and decision-support phases of wireless security assessment, enabling intelligent target ranking, attack feasibility estimation, and strategy recommendation, while preserving strict human-in-the-loop control and budget-aware execution. We describe the system architecture, governance mechanisms, prompt-engineering methodology, and empirical experiments conducted across multiple wireless environments. The results show that GenAI assistance

improves target selection accuracy and overall assessment efficiency, while maintaining auditability and ethical safeguards.

6. The rise and rise of LLM-powered PenTesting systems: The State of the Art

Haitham Al-Sinani (Diwan of Royal Court & German University of Technology in Oman); Chris Mitchell (Royal Holloway, University of London); Abdulaziz Al-Hosni (Diwan of Royal Court)*

This survey examines the rapid emergence of AI-enhanced PenTesting (penetration testing) systems driven by Large Language Models (LLMs). It reviews research published between late 2022 and early 2026, analysing a broad range of proposed systems, tools, and benchmarking frameworks. Rather than documenting each work separately, the survey synthesises recurring architectures, operational scopes, capabilities, and evaluation approaches reported across the literature, highlighting representative design choices and limitations. By comparing these systems, the survey identifies common trends, research directions, and gaps, and discusses benchmark practices, ethical considerations, and open challenges. Given the complexity and rapid proliferation of LLM-powered PenTesting systems, which can be challenging for students and novices to navigate, this paper also aims to support learning for aspiring practitioners and newcomers. The goal is to provide researchers and practitioners with a concise overview of the state of the art and guidance for future development, and to encourage further research into understanding, evaluating, and improving LLM-driven PenTesting systems.

7. Operationalising Cyber Risk Management Using AI: Connecting Cyber Incidents to MITRE ATT&CK Techniques, Security Controls, and Metrics

*Emad Sherif (De Montfort University)**

The escalating frequency of cyber-attacks poses significant challenges for organi-sations, particularly small enterprises constrained by limited in-house expertise, insufficient knowledge, and financial resources. This research presents a novel framework that leverages Natural Language Processing to address these chal-lenges through automated mapping of cyber incidents to adversary techniques. We introduce the Cyber Catalog—a knowledge base that integrates CIS Critical Security Controls, MITRE ATT&CK techniques, and metrics. This integrated re-source enables organisations to connect threat intelligence directly to actionable controls and measurable outcomes. To operationalise the framework, we fine-tuned all-mpnet-base-v2, a highly regarded sentence-transformers model used to convert text into numerical vectors on an augmented dataset comprising 74,986 incident-technique pairs to enhance semantic similarity between cyber incidents and ATT&CK techniques. Our fine-tuned model achieved a Spearman correlation of 0.7894 and Pearson correlation of 0.8756, demonstrating substantial im-provements over top baseline models including all-mpnet-base-v2 ($\Delta\rho = 0.2042$), all-distilroberta-v1 ($\Delta\rho = 0.2118$), and all-MiniLM-L12-v2 ($\Delta\rho = 0.2309$). Fur-thermore, our model exhibited significantly lower prediction errors (MAE = 0.135, MSE = 0.027) compared to baseline models, confirming superior accuracy and consistency. The Cyber Catalog, datasets, trained model, and code made pub-licly available to facilitate further research and enable deployment in resource-constrained environments. This work bridges the gap between threat intelligence and operational security management, providing an actionable tool for systematic cyber incident response and evidence-based cyber risk management.

8. Security incident filtering theory – a model of log analysts' prioritizations

Teodor Sommestad (Swedish Defence Research Agency); Henrik Karlzén (Swedish Defence Research Agency); Max Landauer (Austrian Institute of Technology); Markus Wurzenberger (Austrian Institute of Technology); Florian Skopik (Austrian Institute of Technology)*

Many organizations invest considerable resources in monitoring security-related events in their computer networks, e.g., by setting up or hiring security operation centers. Best-practice guidelines and qualitative analyses based on interviews with practitioners suggest that it is complex and difficult to make decisions related to prioritizations in monitoring and subsequent analyses. In this paper, we introduce Security Incident Filtering Theory. This theory suggests that three variables influence the priority that an event receives in log analysis processes: the operational context the event occurs in, how unusual the circumstances of the event are, and if the event is believed to be malicious or not. The model is tested on hypothetical scenarios with answers from 57 practitioners. It is found that all three hypotheses hold and that the model explains 46% of the variance in event prioritizations.

9. Operational Technology in the Cloud

Karin Bernsmed (SINTEF Digital); Martin Gilje Jaatun (SINTEF ICT); Maren Istad (SINTEF Energy)*

The Norwegian power sector has come a long way when it comes to digitalization. In this paper we describe how the security of operational control systems can be ensured when functions critical to operational control depend on, or are located in, the cloud. Based on a literature review, we show how three other industries have started to use cloud services for IT/OT systems, and also peek at how the power industry outside Norway is approaching this issue. Furthermore, we review security requirements from current Norwegian regulations and relevant standards, and use this to sketch four different approaches to the secure use of cloud services for operational control systems. We conclude that cloud services are generally sufficiently secure for use in operational control systems, given that the supplier is willing and able to meet a number of necessary security requirements, but that it is currently not possible for Distribution System Operators to use such solutions due to current regulations.

10. Protecting Against Deepfake Voice Synthesis: A Gradient-Based Approach to Secure Voice Communications

Sanaz Kavianpour (Abertay University); Marc Kydd (Abertay University); Jordan O'Hara (Naked Energy)*

The rise of deepfake voice cloning technologies poses a significant threat to personal and corporate security, enabling real-time fraud, social engineering, and the evasion of authentication systems. Traditional reactive detection methods are often inadequate for safeguarding live communication systems, creating an urgent need for proactive defence strategies. We examined the effectiveness of a proactive defence system that uses gradient-based audio perturbations to interfere with the voice cloning process. A framework was created using a multi-objective evolutionary algorithm to optimise five key perturbation parameters: noise level, pitch shift, filter frequency, harmonic distortion, and amplitude modulation. The system's effectiveness was thoroughly tested against several state-of-the-art voice cloning backends, including the ElevenLabs API and the open-source XTTS-v2 model, with cloning success measured using ECAPA-TDNN speaker similarity. Experimental results demonstrated the system's effectiveness against commercial-grade voice cloning systems, with perturbations reducing voice similarity to a range of 0.37 to 0.52, which is below the threshold

for successful cloning. The framework consistently produced high-quality results, achieving an average fitness score of about 0.866 across multiple independent runs. Analysis of the parameters revealed that high-frequency filtering (in the 3.4-5.8 kHz range) and pitch shifting are the most critical parameters for effective disruption. A real-time implementation achieved a latency of around 32ms. These findings indicate that gradient-based perturbations, when adaptively optimised, serve as a viable and practical real-time defence against sophisticated voice cloning systems, which helps create a robust methodology for proactive voice security and provides a blueprint for deploying these protective systems in real-world communication environments.

11. WebShield: Real-Time Detection of Phishing Websites and Dark Patterns Using Machine Learning and Chrome Extension

Venkata Karthik Muppidi (SIDDHARTHA ACADAMY OF HIGHER EDUCATION); M. Vani Pujitha (SIDDHARTHA ACADAMY OF HIGHER EDUCATION); Jaswanth Potnuri (SIDDHARTHA ACADAMY OF HIGHER EDUCATION); Sandepogu Sudheer (SIDDHARTHA ACADAMY OF HIGHER EDUCATION); Parshav Meghan (SIDDHARTHA ACADAMY OF HIGHER EDUCATION)*

Real-time detection of phishing websites and manipulative dark patterns is essential for safe and informed browsing, particularly as cyber threats continue to grow in scale and sophistication. Website classification is achieved using the XGBoost ensemble model, aided by a 30-feature engineering pipeline spanning URL structure, domain metadata, HTML/JavaScript behavior, and abnormal interaction patterns. Hybrid filtering and intelligent caching provide strong input quality for realtime model inference. This work proposes WebShield, a Chrome browser extension that integrates machine learning-based phishing detection with dark pattern identification through a React.jsbased user interface. A dataset of 11,055 labeled websites (6,157 legitimate; 4,898 phishing) was collected from PhishTank, OpenPhish, and Alexa Top Sites to support robust model training. The system employs an XGBoost classifier taking website risk factors into account to predict threat levels depending on the browsing context. A companion extension called Time Tracker is developed to provide per-domain browsing analytics with complete ondevice privacy. Experimental outcomes on the proposed system achieved 97.06% cross-validation accuracy and 86.5% dark pattern detection rate, with a 1.4s end-to-end detection time and 99.4% system uptime. The proposed solution offers a scalable, explainable approach to browser security applicable to personal safety, digital well-being, and organizational cybersecurity.

12. Cybersecurity Economics, PQC and AI—25 Years Post Anderson (2001)

*Tim Williams (Bank of Ireland)**

In 2001, Ross Anderson published "Why Information Security Is Hard—An Economic Perspective," which established the foundational framework for understanding security failures through economic incentive analysis. Twenty-five years on, the technology-dependent world faces a significant challenge: transitioning global cryptographic infrastructure to quantum-resistant algorithms before cryptographically relevant quantum computers emerge. This challenge is complicated by simultaneous advances in artificial intelligence. This paper presents evidence that post-quantum cryptography (PQC) transitions face all five of Anderson's original perverse incentive categories simultaneously: an emergent scenario explained by the path dependence and technological momentum established when public-key infrastructure achieved global deployment in the 1990s. Drawing on Science and Technology Studies literature, we show why PQC constitutes the first fundamental paradigm change since

cryptographic technological momentum was established and consequently why economic incentive misalignments operate in uniquely concentrated form. We further identify and analyse ten PQC-specific perverse incentive mechanisms extending Anderson's taxonomy, organised into financial/accounting, information/discovery and temporal/governance thematic groups. The paper concludes with policy recommendations targeting each identified mechanism, demonstrating both the continued relevance of security economics 25 years after Anderson's founding contribution and the need for its extension to address the unforeseen challenges of cryptographic infrastructure transitions.

13. What Investigators Inherit: Public Corpora as Evidence States for Forensic Readiness Evaluation

Belinda Onyeashie (University of the West of England); Jonathan Lancelot (University of the West of England); Petra Leimich (Edinburgh Napier University)*

Incident response can restore service, yet post-incident investigation may remain inconclusive when required evidence does not exist at acquisition, cannot support corroborated claims, or does not address core investigative questions. Forensic readiness seeks to reduce this risk, but many readiness frameworks focus on policy and process instead of the evidential condition available after an incident. This study treats public forensic corpora as representative post-incident evidence states and examines readiness through the answerability of three core investigative questions: who initiated the activity, what systems or data changed, and whether data transfer occurred. A question-driven rubric assesses answerability through corroboration across independent evidence categories. The analysis examines eight public corpora spanning legacy disk-centric and modern log-centric acquisition approaches. Legacy corpora generally lack authentication logs, which limits attribution assessment regardless of disk preservation quality. Modern corpora commonly lack raw packet capture, which restricts data movement assessment except where application-layer logs provide alternative evidence. Only one corpus supports data movement assessment and only through HTTP request logging. The findings associate unanswered investigative questions with specific readiness controls referenced in ISO/IEC 27001 Annex A and the UK National Cyber Security Centre guidance. The study frames public corpora as evidence states for readiness evaluation and describes gap-to-control mappings intended to support preparedness decisions.

14. Impact of Import Hash on ML-Based YARA Rule Generation – A PoC Study

*Ferenc Leitold (Óbuda University)**

Machine learning-assisted YARA rule generation is increasingly used to support malware detection in Security Operations Center (SOC) environments. While string-based features are commonly employed in such workflows, the impact of incorporating structural information into ML-driven rule generation remains underexplored. This paper presents a proof-of-concept study evaluating the effect of adding a single structural feature, the Portable Executable (PE) import hash (imphash), to a string-based ML-driven YARA rule generation pipeline. Using the Sophos ML Toolkit and a logistic regression classifier, we trained models on a 2020 malware dataset and evaluated their performance on temporally separated malware corpora from 2022 and 2023, as well as on a Windows 10-based goodware dataset. Results show that while both configurations performed similarly on temporally close samples, the imphash-augmented model significantly improved forward temporal generalization. On the 2023 dataset, the average missed-malware ratio decreased from 8.32% (string-only baseline) to 1.06% when imphash was included. However, this improvement came at the cost of increased false positive detections on benign files. The findings demonstrate that even minimal structural feature

augmentation can materially influence ML-based YARA rule behavior. The proposed experimental framework provides a reproducible method for evaluating feature-level modifications in operational malware detection pipelines.

15. Optimal Spending on Cybersecurity Measures

*Tara Kissoon (University of London - Alumni Ambassador)**

"Met Police Officers at Risk After Serious Data Breach. London's Metropolitan Police Service is investigating a serious data breach that may have exposed names, ranks, and photographs for potentially all 47,000 personnel, including the identity of undercover officers." When an organization is attacked through a breach of information security controls, the law requires the organization to notify individuals that their personal identifiable information (PII) has been exposed. It has a privacy risk and may cause significant harm to the individual. When data is accessed and transferred to unauthorized individuals for the purpose of fostering criminal activity, PII data elements become extremely important. Criminals harvest and sell this information on areas of the internet known as the dark/deep web, to propagate further criminal activity.
This is the primary reason why funding safeguards in organizations is important for protecting organizations from data and privacy breaches that result in this type of impact on citizens. This session will utilize business-driven risk assessments to demonstrate adherence to data protection regulations. This session explores topics such as Geopolitical Risk, i.e., cyber threats, and how this may lead to identity takeover. One of the most important data protection legislations enacted to date is the General Data Protection Regulation (GDPR). Currently, there are more than one hundred and twenty countries that have enacted legislation to secure the protection of data and privacy. The European Union's (EU) General Data Protection Regulation, implemented in May 2018, brought data protection into public view and is considered a landmark privacy law with the introduction of new rights for individuals, such as the Right to be Forgotten and the Right to Portability. This session utilizes a case study to demonstrate the cybersecurity risk management framework and process, to address laws, government regulations, and industry standards.

16. CROFv2 – Moving beyond general cyber resilience framework

*Cyril Onwubiko (Research Series)**

In 2021 we published the cyber recovery operational framework (CROF), a general purpose operational cyber recovery and resilience framework, to be applied to any organisation for cyber recovery assessments. Since then, CROF has been adopted and used across many industries. Feedback, comments and insights from these organisations that have been using the framework have been invaluable. Incorporating these insights, we have extended the framework and developed a web app for it. It is now CROFv2, an adaptive, automated, robust framework that allows organisation to conduct cyber recovery assessment, execute immutable backups to prevent ransomware, compare and baseline their progress over time, compare their performance against peer organisations, and drive improvement and progress. In addition, the framework includes both domain-specific and sector-specific cyber recovery controls and objectives.

17. Alignment vs. Capability: Does Safety Removal Improve Cybersecurity Performance LLM?

*Vignesh Muraleedharan (SRMIST)**

Large language models are increasingly deployed in cybersecurity workflows for tasks such as vulnerability classification, static analysis review, and log triage. As these models gain capability, concerns have grown around the trade-off between safety alignment and task performance, especially safety mechanisms meaningfully restricting useful cybersecurity functionality. This paper presents a controlled empirical study comparing Qwen3-4B-Instruct-2507 (baseline aligned) against its abliterated variant produced via the Heretic framework, across four benchmark tasks: CWE vulnerability classification, SAST false positive reduction, security log triage, and a graded safety stress test. Results show that alignment removal produced no improvement in defensive task performance, the accuracy was identical or marginally lower in the decensored model across all three defensive benchmarks. Critically, the safety stress test revealed a complete collapse of refusal behaviour at operational and harmful severity levels: the aligned model refused 100% of severity-3 and severity-4 prompts while the abliterated model refused none. These findings challenge the common assumption that alignment meaningfully constrains cybersecurity utility, and provide quantitative evidence that safety mechanisms in this model operate as a targeted behavioural layer rather than a broad capability suppressor.

18. PolicyControls: A Web-Based Platform for Cybersecurity Control Assessment and Compliance Management

*Cyril Onwubiko (Research Series Ltd)**

PolicyControls is an open, web-based cybersecurity control assessment and compliance management platform designed to help organisations systematically identify, track, and remediate security control gaps across their technology estate. Built on a modern Python/Flask stack, the platform provides a structured catalogue of controls mapped to widely adopted regulatory frameworks including GDPR, ISO 27001, NIST CSF, PCI-DSS, HIPAA, and SOC 2. This paper describes the problem context, system design, key functional capabilities, a worked illustration using a real-world case study, and the broader importance of accessible compliance tooling for organisations of all sizes.

19. HDIT: A Cyber Situational Awareness Framework for Real-Time Behavioural Drift Detection in Autonomous AI Agents

*Wendi Soto (King's College London)**

Abstract. Autonomous AI agents are increasingly deployed in enterprise environments where they perform consequential tasks. Current governance models assume behavioural determinism, yet generative AI agents are stochastic systems whose behaviour evolves continuously. We present High-Dimensional Identity Tracking (HDIT), a cyber situational awareness framework that treats an AI agent's behavioural identity as a computable, time-varying signal. HDIT computes sixteen measurement vectors spanning semantic identity, operational patterns, and integrity indicators across three temporal horizons. When drift exceeds self-calibrating thresholds, a governance state machine transitions the agent through escalating restriction states. We formalize the mapping to Endsley's three-level situational awareness model, positioning agent monitoring as a novel cyberSA problem class. In controlled tests, the system produced clear governance separation between normal variation and induced drift conditions, with no false positives observed in the controlled same-model

condition evaluated here.

20. An Intelligent System for Identifying Fraudulent Accounts in Social Media Leveraging Deep Learning Techniques

*Sanyu Paul Reddy (Malla Reddy University); Sudha Kodi (Malla Reddy University)**

The growth of social media platforms has also led to an increase in the number of fraudulent activities including fake accounts, automated bots, and spam accounts which the users fall into a security trap. Current security systems often rely on mechanisms that can be bypassed by malicious users. To address these issues, this study introduces a detection framework that com-bines the behavioral analysis of accounts with deep learning mechanisms to determine whether a particular account is genuine or fraudulent. This is achieved by extracting 11 features which include all the major details of an account by which our system trains a Deep Neural Network (DNN) that helps in the identification of hidden patterns. The architecture is a web application featuring Flask backend and React frontend, providing a smooth functionality for the users along with maintaining platform integrity.

21. A Comparative Analysis of Situational Awareness Support in Cyber Incident Response Ticketing Systems

Liberty Kent (University College London); Nilufer Tuptuk (University College London); Ingolf Becker (University College London)*

This study aims to assess how effectively commonly used ticketing systems support situational awareness (SA) within Cyber Security Incident Response Teams (CSIRTs). To do this, we identified 38 CSIRT ticketing systems and conducted a multi-criteria decision analysis using measures developed from previous literature. We found that all systems showed sufficient support for Level 1 SA (perception) while support for higher-level SA (comprehension and projection) varied more across systems. Amongst these systems, Jira and BMC exhibited the highest scores, indicating a comparatively strong capacity to support SA. These systems utilised automation and visualisation effectively to support comprehension and projection, with the ability to identify similar tickets, represent information in contextualised visuals, provide insights into typical resolution paths, and triage according to past ticket data and operator workload/availability. The results highlight significant differences between common ticketing platforms, with a focus on perceptual-level function while lacking a similar regard for conceptualisation and projection. Our findings have implications for CSIRTs and their effectiveness, such as considerations for tool selection.

22. Know Your Unknowns: A Strategic Johari Game for Cyber Security Management

Eckhard Pfluegel (Kingston University); Neda Ahmadi (Kingston University); Rehan Usman (Kingston University)*

Cyber security teams must make joint decisions in dynamically changing environments while balancing two competing priorities: uninterrupted technical work and mutual communication for knowledge exchange. In practice, this balance is difficult to achieve as groups with different roles and incentives (e.g., security engineers conducting vulnerability assessment or security analysts focusing on risk management) may prefer different actions. To address this problem, we develop in this paper a novel two-player non-zero-sum coordination game that models the strategic choice between focusing (F) and communicating (C), inspired by the Johari Window approach. We contribute a dual-window game-theoretic formulation, a parameterised

payoff structure that informs the game design, and a complete analysis of its Nash equilibria. We show that the game has pure-strategy equilibria (F,F) and (C,C), as well as a mixed-strategy equilibrium. While the mixed equilibrium is not Pareto-optimal, it offers a valid compromise, as in practice, both focusing and communication must coexist in most team settings. Overall, our model offers a systematic decision-support framework for effective cyber security management.

23. Reactive Digitalization, Fragmented Security – Insights from Manufacturing Small and Medium-sized Enterprises

Henning Thomsen (Aalborg University); Christian Black Jørgensen (University College of Northern Denmark); Bent Rosenkilde (University College of Northern Denmark); Mikkel Graugaard Langdahl Antonsen (University College of Northern Denmark)*

Digitalization and cybersecurity are important elements in manufacturing, but Small and Medium-sized Enterprises (SMEs) struggle with these aspects. We tudy the challenges related to cybersecurity when manufacturing SMEs digitalize, and the measures taken. The results are obtained using interviews with SMEs, and we see that digitalization is mostly reactive and project-based, and cybersecurity is treated as an add-on motivated by external factors such as regulatory requirements. Also, the SMEs have limiting factors regarding employee knowledge, with only a few addressing this in terms of upskilling and courses. Our analysis shows that practical guidance and considering cybersecurity from the start are important, and that simplified models for digitalization for SMEs are warranted.

24. CypherTrap: Tracking Threats Through Integrated Honeypot-IDS Framework.

Amoghavarsha K (Jain university); Likith R (Jain University)*

The conventional security systems that use reactive detectives have been found to be inadequate as cyber threats transform into automated and high frequency campaigns. In this paper, a new intelligent and integrated system, named CypherTrap is presented by proactively identifying, analyzing and alerting on malicious activities based on a combination of deception technology and network monitoring. The system utilises Cowrie which is highly interactive honeypot that emulates vulnerable SSH environments and Suricata which is very powerful Intrusion Detection System (IDS) to inspect deep packets. CypherTrap can combine these different data streams into a single database in PostgreSQL in comparison with isolated legacy systems in which each stream of data is treated separately. This system has an automated responding engine that sends real time notifications through Gmail and Telegram API when there is a high severity event. The defenders have a centralized Flask-Tailwind dashboard, which allows them to have live visibility into attack vectors, geo-location traces, and system health. Experimental findings indicate that CypherTrap is a viable approach to bridge the correlation gap between the interfaces of low interaction and high fidelity network observation, as it provides an efficient means of cybersecurity, which is both proactive and scalable to the needs of enterprise and research settings.

25. Quantum Vulnerability as a Financial Impairment Triggering Event: Rethinking Capitalised Systems with Cryptographic Infrastructure Dependencies Under IAS 36, IAS 38, ASC 350-40 and ASC 360-10

*Tim Williams (Bank of Ireland)**

Post-Quantum Cryptography (PQC) migration is almost universally framed as a forward-looking security expenditure, but for or- ganisations carrying capitalised software and hardware assets this framing obscures a material financial reporting question that existing standards already require to be asked. Drawing on IAS 36 (Impairment of Assets), IAS 38 (Intangible Assets), ASC 350-40 (Internal-Use Software) and ASC 360-10 (Property, Plant and Equipment), we argue that the NIST fi- nalisation of post-quantum cryptographic standards in August 2024, read alongside national regulatory mandates and observable market signals, constitutes external indicators requiring organisations to assess whether capitalised assets dependent on quantum-vulnerable cryptographic proto- cols carry a recoverable amount below their current carrying value. By analogy with the IASB's treatment of climate-related matters, and more closely with the accounting precedent established during the London Inter-Bank Overnight Rate (LIBOR) discontinuation, we develop the concept of quantum-driven technological obsolescence as a triggering event and examine its implications for management, auditors and standard-setters. The paper identifies a structural gap between what existing standards require and what management and auditors currently do, and concludes with four structured recommendations for standard-setter, regulatory, audit firm and preparer action.

26. Hierarchical Detection of Malicious Command Lines Using Random Forest and Context-Aware LSTM

Qinzheng Hu (University of Glasgow); Yixian Jiang (University of Glasgow); Pavandeep Singh Baxi (University of Glasgow)*

Malicious command-line activity remains a key vector in host-based cyber attacks, particularly when obfuscation and multi-step execution are used to evade detection. Existing approaches either rely on rule-based methods that are sensitive to variation or apply machine learning to isolated commands without incorporating temporal context. This paper presents a hierarchical detection framework that integrates feature-based machine learning with sequence-aware modeling within a unified evaluation setting. At the single-command level, we compare rule-based detection, hybrid logistic regression, and random forest classifiers over a shared feature representation. At the sequence level, these models are extended using LSTM-based context modeling to capture temporal dependencies across command streams. Experimental results show that random forest achieves the strongest performance for single-command detection, while the RF+LSTM model provides the best performance for command-stream analysis. These findings demonstrate that effective command-line threat detection benefits from combining strong local classification with context-aware refinement, particularly when malicious behavior unfolds across multiple steps.

27. ToolProbe: A Two-Stage Evaluation Framework for LLM Agent Safety in MCP Tool-Calling Environments

Ramkumar Sundarakalatharan (Zerberus.ai); Sriram Gopalakrishnan (Zerberus.ai)*

As LLM agents are increasingly deployed with tool-calling capabilities via the Model Context Protocol (MCP), they become susceptible to adversarial attacks that cause harmful tool execution. Unlike text-based threats, tool-calling attacks span multiple architectural layers. Yet, existing benchmarks do not address this complexity, relying on single-pass classification that conflates safe refusals with incomplete reasoning. To address this, we present Tool-Probe, a two-stage evaluation framework operating on a three-layer attack taxonomy (20 types across Server, Host, and User layers) mapped to OWASP LLM Top 10 and MITRE ATLAS. Its two-stage LLM-as-Judge applies weighted criteria scoring for clear-cut cases, conditionally triggers a reasoning lifecycle audit for borderline cases, and includes independent score verification.

We construct 750 adversarial datapoints with tool definitions from 95 real Smithery.ai MCP servers and evaluate three frontier models: Claude Sonnet 4.5, GPT-5.1, and DeepSeek v3.1, yielding 2,250 assessments. Attacks succeeded in 47.8% of evaluations: DeepSeek exhibited 76.1% adversarial compliance, GPT-5.1 53.9%, while Claude demonstrated 86.7% resistance. Models failed to maintain majority resistance in 8 of 19 categories, with 69.6% inter-model disagreement and an inverse relationship between agentic activity and safety. These findings underscore the need for a multistage evaluation to assess the deployment readiness of LLM agents in tool-calling environments.

28. Mitigating Skin Tone Bias in AI-Based Melanoma Detection Using Neural Style Transfer and Adversarial Domain Adaptation

*Nyasha Samuel Makanza (University of Portsmoiyth); Intissar Ziani (University of Abdelhamid Mehri); Konstantina Kanta (Beiersdorf); Gueltoum Bendiab (University of Portsmouth); Aikaterini Kanta (University of Portsmouth)**

Skin cancer is a common and potentially deadly disease caused by ultraviolet (UV) radiation, with melanoma being the most aggressive form due to its high potential to metastasise if not detected early. Therefore, early diagnosis is crucial for effective treatment, and AI-based image detection and classification systems have shown significant promise in supporting this task. However, existing models often underperform on darker skin tones because melanoma can present differently in individuals with higher melanin levels, leading to biased systems that may misdiagnose or fail to detect the disease. To address this limitation, this paper generates synthetic melanoma images using style transfer, combining real lesions with diverse skin tones based on the Fitzpatrick scale. These synthetic images serve as the target domain in an Adversarial Domain Adaptation (ADA) framework with a Gradient Reversal Layer (GRL), enabling the model to reduce skin-tone bias by focusing on pathological features rather than background characteristics. Evaluation on an unseen synthetic test dataset demonstrated strong performance, achieving a peak AUC of 0.965 (Type IV) and maintaining values above 0.92 across all skin types, including Type VI. Overall, integrating synthetic data with adversarial learning provides a scalable and robust approach to improving fairness and diagnostic equity in melanoma detection.

29. Artificial Intelligence for Cyber Situation Awareness: Trust, Distortion and Human Judgement

Ibrahim Maniku (Queen Mary University of London); Maria Bada (Queen Mary University of London)*

Artificial intelligence is increasingly embedded within cyber security workflows as a decision-support and advisory mechanism. While prior work has primarily framed AI as a tool for enhancing detection and response capabilities, this paper argues that AI systems used in cyber decision-support settings should be understood not only as assistive tools, but also as cognitive mediators whose outputs may structure, narrow, or distort situational awareness in ways that matter for trust and judgement.

Drawing on concepts from human factors and cognitive science, this work examines the implications of AI-mediated decision support for trust, cognitive offloading, and confidence calibration. We argue that AI does not merely augment awareness but can also distort it through selective framing, overconfident outputs, persuasive errors, and potentially misleading or deceptive output patterns. These effects are particularly consequential in environments where users rely on AI-generated assessments to interpret threats and make time-sensitive decisions. The paper proposes a conceptual framework positioning AI as part of the situational-awareness problem itself, rather than an external support tool. We identify key risk factors, including over-reliance, reduced critical scrutiny, and misaligned mental models, and discuss their implications for organisational cyber resilience. By integrating perspectives from artificial intelligence and human cognition, this research contributes to a more nuanced understanding of human-AI collaboration in cyber security. It highlights the need for design, training, and governance approaches that explicitly account for the cognitive and behavioural impacts of AI-mediated decision making.

30. A Comparative Study of Machine Learning, Deep Learning, and Transformer-Based Models for Malware Detection with an Adaptive Mixture of Experts Framework

Rama Satya CH (Mahindra University); Raghu Kisore N (Mahindra University)*

Malware detection remains a challenging problem due to the heterogeneity, obfuscation, and rapid evolution of malicious software. This paper presents a comparative study of classical machine learning, deep learning, and transformer-based approaches for malware classification, with particular emphasis on the trade-offs between predictive performance, computational cost, and model adaptability. In addition to standard baselines, we investigated a Mixture of Experts (MoE) extension built on top of a transformer-based model to enable input-dependent expert specialization over contextual code representations.

Experimental evaluation on a Win32 malware dataset shows that ensemble-based classical models remain highly competitive in terms of raw accuracy, while deep and transformer-based models offer complementary advantages in recall, representation learning, and adaptability. Although the proposed MoE-based architecture does not outperform the strongest tree-based baseline, it demonstrates a balanced precision–recall trade-off and improved flexibility through sparse expert routing. These findings highlight that, in malware detection, adaptive transformer architectures should be evaluated not only by absolute accuracy but also by their ability to model heterogeneous malware behaviors and support extensible future designs.

31. Cyber Threat Intelligence Implementation in Critical Infrastructure: A Malaysian Perspective

NURUL NUHA BINTI ABDUL MOLOK (International Islamic University Malaysia (IIUM)); SHUHAILI BT. TALIB (International Islamic University Malaysia (IIUM)); NOOR HAYANI BINTI ABD RAHIM (International Islamic University Malaysia (IIUM)); ZAHIDAH BINTI ZULKIFLI (International Islamic University Malaysia (IIUM))*

The increased use of Artificial Intelligence (AI) today has dramatically changed the global cybersecurity threat landscape. While AI is being used in cybersecurity for control mechanisms, it also poses threats to organizations. Thus, large organizations are adopting cyber threat intelligence (CTI) as part of their cybersecurity practices. However, there are many organizations that have not adopted CTI, making this area underexplored in academic research. This paper explores global CTI implementation in general, focusing on its position in Malaysia, in particular. It examines the assignment of CTI roles within the cybersecurity function looking at the placement of a CTI unit, its reporting structure, the involvement of top management in CTI and members of the team. This study employs qualitative research approach through semi-structured expert interviews on CTI leaders and practitioners across 11 organizations as data collection instrument and thematic analysis as data analysis technique. Findings reveal three key themes of this exploratory study which are i) placement of CTI unit or department; ii) CTI reporting structure; and iii) CTI team members. It was highlighted that CTI was commonly a part of cybersecurity related departments or divisions of an organization, having a variety of reporting structures and roles. It is important to note CTI services are provided in-house, outsourced and hybrid (combination of in-house and outsourced) in Malaysian organizations under study. As the Malaysian Cyber Security Act 2024 emphasizes the collection, coordination and dissemination of CTI among National Critical Information Infrastructure (NCII) sector entities, this study does not only shed light on the implementation of CTI in Malaysian NCII entities but also to be adopted by any types of organizations that intended to beef up their cybersecurity posture in today's era.

32. The Epistemology of Confident Machines

*Carolyn Swinney (University of Essex)**

This paper argues that large language models produce Gettier cases not as occasional errors but as an architectural inevitability and that these cases are epistemically sterile in a way that human epistemic failures are not. It introduces a distinction between productive and sterile Gettier cases, with productive cases generating improved epistemic practice through consequence, backtracing, and adaptation, and sterile cases recuring indefinitely when none of these conditions are present. The paper develops this argument across three layers. First, the Gettier structure is shown to be inherent in next-token prediction, with recent research demonstrating that even perfectly calibrated models must hallucinate. Second, training time compression is shown to destroy provenance, making backtracing architecturally impossible. Third, the paper proposes three structural preconditions for epistemic caring, irreversibility, scarcity, and consequence, and argues that without them and the finitude they presuppose, no amount of calibration or scale can transform sterile Gettier cases into productive ones.

Cyber Science 2026 Conference

Timetable

Cyber Science 2026

Conference Timetable

3–5 June 2026

Stewart House via Senate House, Royal Holloway, University of London — London, UK

Theme — Cyber Science in the Era of Artificial Intelligence

c-mric.org

Day 1 — Wednesday, 3 June 2026

In-person only · UK Time Zone (GMT+1) · Moderator: Konstantinos Mersinas

Time	Session
Morning Session	
08:45 – 09:30	*Registration & Breakfast (45 minutes)* *Registration desk closes promptly at 09:30. All delegates are kindly asked to arrive in good time to complete registration before the session begins.*
09:30 – 09:50	**Conference Announcements & Opening of Proceedings** **Dr Cyril Onwubiko — Founder, Centre for Multidisciplinary Research, Innovation and Collaboration (C-MRiC)** **Cyber Science 2026 — the 11th International Conference on Cybersecurity, Situational Awareness and Social Media.**
10:00 – 11:00	**Keynote Session** **Dr Carolyn Swinney** **Royal Air Force Communications & Electronics Engineering Officer; Executive Fellow, University of Essex**
11:00 – 11:10	*Coffee & Tea Break (10 minutes)*
11:10 – 11:35	**Security Incident Filtering Theory — A Model of Log Analysts' Prioritizations** *Teodor Sommestad (Swedish Defence Research Agency)*; Henrik Karlzén (Swedish Defence Research Agency); Max Landauer (Austrian Institute of Technology); Markus Wurzenberger (Austrian Institute of Technology); Florian Skopik (Austrian Institute of Technology)*
11:35 – 12:00	**What Investigators Inherit: Public Corpora as Evidence States for Forensic Readiness Evaluation** *Belinda Onyeashie (University of the West of England)*; Jonathan Lancelot (University of the West of England); Petra Leimich (Edinburgh Napier University)*
12:00 – 13:00	*Group Photograph & Lunch Break (60 minutes)*
Afternoon Session	
13:10 – 13:35	**Impact of Import Hash on ML-Based YARA Rule Generation — A PoC Study** *Ferenc Leitold (Óbuda University)**
13:35 – 14:00	**CROFv2 — Moving Beyond a General Cyber Resilience Framework** *Cyril Onwubiko (Research Series)**
14:10 – 14:35	**A Comparative Analysis of Situational Awareness Support in Cyber Incident Response Ticketing Systems** *Liberty Kent (University College London)*; Nilufer Tuptuk (University College London); Ingolf Becker (University College London)*
15:00 – 15:25	**ToolProbe: A Two-Stage Evaluation Framework for LLM Agent Safety in MCP Tool-Calling Environments** *Ramkumar Sundarakalatharan (Zerberus.ai)*; Sriram Gopalakrishnan (Zerberus.ai)*
15:25 – 15:55	*Coffee & Tea Break (30 minutes)*
16:00 – 16:25	**The Epistemology of Confident Machines** *Carolyn Swinney (University of Essex)**
16:30 – 16:55	**An Artefact-Driven Framework for Embedding Security by Design in the Systems**

	Engineering V-Model *Enoch Agyepong (Elbit Systems, UK)**
17:00 – 17:25	**Quantum Vulnerability as a Financial Impairment Triggering Event: Rethinking Capitalised Systems with Cryptographic Infrastructure Dependencies under IAS 36, IAS 38, ASC 350-40 and ASC 360-10** *Tim Williams (Bank of Ireland)**
17:30 – 17:55	**Hierarchical Detection of Malicious Command Lines Using Random Forest and Context-Aware LSTM** *Qinzheng Hu (University of Glasgow)*; Yixian Jiang (University of Glasgow); Pavandeep Singh Baxi (University of Glasgow)*
Evening	
18:00 – 21:00	**Conference Dinner, Drinks & Photographs** **Antalya Restaurant, London, UK — https://antalyarestaurant.co.uk/**

Day 2 — Thursday, 4 June 2026

In-person only · UK Time Zone (GMT+1) · Moderator: Martin Gilje Jaatun

Time	Session
Morning Session	
08:30 – 09:00	*Breakfast — Tea, Coffee & Water (30 minutes)*
09:00 – 10:00	**Keynote Session** **Dr Konstantinos Mersinas** **Associate Professor, Information Security Group, Royal Holloway, University of London, UK**
10:00 – 10:10	*Break (10 minutes)*
10:10 – 10:35	**PolicyControls: A Web-Based Platform for Cybersecurity Control Assessment and Compliance Management** *Cyril Onwubiko (Research Series)**
10:35 – 11:00	**WiFiPenTester: Towards Governed GenAI-Assisted Wireless Penetration Testing** *Haitham Al-Sinani (RHUL)*; Chris Mitchell (Royal Holloway, University of London); Abdulaziz Al-Hosni (Diwan of Royal Court); Sultan Al-Harrasi (University of Technology and Applied Sciences)*
11:00 – 11:30	**Operational Technology in the Cloud** *Karin Bernsmed (SINTEF Digital); Martin Gilje Jaatun (SINTEF ICT)*; Maren Istad (SINTEF Energy)*
11:30 – 12:00	**Protecting Against Deepfake Voice Synthesis: A Gradient-Based Approach to Secure Voice Communications** *Sanaz Kavianpour (Abertay University)*; Marc Kydd (Abertay University); Jordan O'Hara (Naked Energy)*
12:00 – 13:00	*Group Photograph & Lunch Break (60 minutes)*
Afternoon Session	
13:10 – 13:35	**Know Your Unknowns: A Strategic Johari Game for Cyber Security Management** *Eckhard Pfluegel (Kingston University)*; Neda Ahmadi (Kingston University); Rehan Usman (Kingston University)*
13:35 – 14:00	**Artificial Intelligence as a Force Multiplier in Social Engineering: An Experiential Learning Approach** *Aunshul Rege (Temple University)*; Hwanhee Park (Temple University); John Snee (Temple University)*
14:10 – 14:35	**Cybersecurity Economics, PQC and AI — 25 Years Post Anderson (2001)** *Tim Williams (Bank of Ireland)**
15:00 – 15:25	**The Rise and Rise of LLM-Powered Penetration Testing Systems: The State of the Art** *Haitham Al-Sinani (Diwan of Royal Court & German University of Technology in Oman)*; Chris Mitchell (Royal Holloway, University of London); Abdulaziz Al-Hosni (Diwan of Royal Court)*
15:25 – 15:55	*Coffee & Tea Break (30 minutes)*
16:00 – 16:25	**HDIT: A Cyber Situational Awareness Framework for Real-Time Behavioural Drift Detection**

	in Autonomous AI Agents *Wendi Soto (King's College London)**
16:30 – 16:55	**Artificial Intelligence for Cyber Situation Awareness: Trust, Distortion and Human Judgement** *Ibrahim Maniku (Queen Mary University of London)*; Maria Bada (Queen Mary University of London)*
17:00 – 17:25	**Characterizing Vulnerabilities in the TCP/IP Protocol Stack across Packet Semantics and Control Plane Architectures** *Omkar Bhalekar (RIT)**
17:30 – 17:55	**Operationalising Cyber Risk Management Using AI: Connecting Cyber Incidents to MITRE ATT&CK Techniques, Security Controls and Metrics** *Emad Sherif (De Montfort University)**

Day 3 — Friday, 5 June 2026

Remote only · UK Time Zone (GMT+1) · Moderator: Cyril Onwubiko

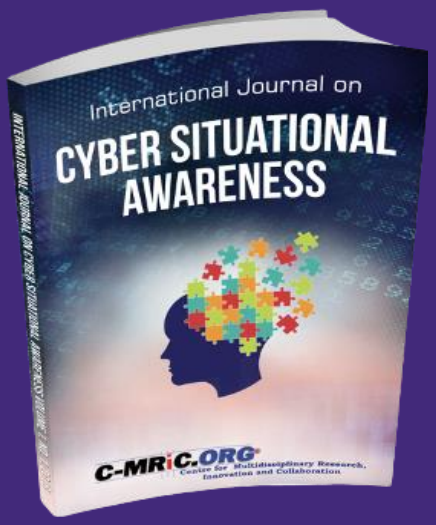
Time	Session
Morning Session	
08:30 – 09:00	*Breakfast — Tea, Coffee & Water (30 minutes)*
09:00 – 10:00	**Keynote Session** **Dr Deepthi Ratnayake** **Principal Lecturer in Cyber Security, Cybersecurity & Computing Systems Research Group, University of Hertfordshire, UK**
10:00 – 10:10	*Break (10 minutes)*
10:10 – 10:35	**Alignment vs. Capability: Does Safety Removal Improve Cybersecurity Performance of LLMs?** *Vignesh Muraleedharan (SRMIST, India)*
10:35 – 11:00	**Reactive Digitalization, Fragmented Security — Insights from Manufacturing Small and Medium-Sized Enterprises** *Henning Thomsen (Aalborg University)*; Christian Black Jørgensen (University College of Northern Denmark); Bent Rosenkilde (University College of Northern Denmark); Mikkel Graugaard Langdahl Antonsen (University College of Northern Denmark)*
11:00 – 11:30	**Mitigating Skin-Tone Bias in AI-Based Melanoma Detection Using Neural Style Transfer and Adversarial Domain Adaptation** *Nyasha Samuel Makanza (University of Portsmouth); Intissar Ziani (University of Abdelhamid Mehri); Konstantina Kanta (Beiersdorf); Gueltoum Bendiab (University of Portsmouth); Aikaterini Kanta (University of Portsmouth)**
11:30 – 12:00	**A Comparative Study of Machine Learning, Deep Learning and Transformer-Based Models for Malware Detection with an Adaptive Mixture-of-Experts Framework** *Rama Satya CH (Mahindra University)*; Raghu Kisore N (Mahindra University)*
12:00 – 13:00	*Group Photograph & Lunch Break (60 minutes)*
Afternoon Session	
13:10 – 13:35	**Cyber Threat Intelligence Implementation in Critical Infrastructure: A Malaysian Perspective** *NURUL NUHA BINTI ABDUL MOLOK (International Islamic University Malaysia (IIUM))*; SHUHAILI BT. TALIB (International Islamic University Malaysia (IIUM)); NOOR HAYANI BINTI ABD RAHIM (International Islamic University Malaysia (IIUM)); ZAHIDAH BINTI ZULKIFLI (International Islamic University Malaysia (IIUM))*
13:35 – 14:00	**WebShield: Real-Time Detection of Phishing Websites and Dark Patterns Using Machine Learning and a Chrome Extension** *Venkata Karthik Muppidi (Siddhartha Academy of Higher Education)*; M. Vani Pujitha (Siddhartha Academy of Higher Education); Jaswanth Potnuri (Siddhartha Academy of Higher Education); Sandepogu Sudheer (Siddhartha Academy of Higher Education); Parshav Meghan*

	(Siddhartha Academy of Higher Education)
14:10 – 14:35	**PhishShield: A Multi-Modal Gmail Security Extension Using Machine Learning and BERT for Phishing Detection** *Tharun M R (K.S. Rangasamy College of Technology)*; Pugalendi S (K.S. Rangasamy College of Technology); Dinesh Kumar K (K.S. Rangasamy College of Technology)*
15:00 – 15:25	**An Intelligent System for Identifying Fraudulent Accounts in Social Media Leveraging Deep Learning Techniques** *Sanyu Paul Reddy (Malla Reddy University); Sudha Kodi (Malla Reddy University)**
15:25 – 15:55	*Coffee & Tea Break (30 minutes)*
16:00 – 16:25	**CypherTrap: Tracking Threats Through an Integrated Honeypot–IDS Framework** *Amoghavarsha K (Jain University)*; Likith R (Jain University)*
16:30 – 16:55	**Optimal Spending on Cybersecurity Measures** *Tara Kissoon (University of London — Alumni Ambassador)**
17:00 – 17:15	**Conference Closing Remarks** **Prof. Martin G. Jaatun**

All times shown are UK Time (GMT+1). Programme subject to minor changes.

International Journal on Cyber Situational Awareness (IJCSA)

ISSN: (Print) 2057-2182 ISSN: (Online) 2057-2182, DOI: 10.22619/IJCSA



The International Journal on Cyber Situational Awareness (IJCSA) is a comprehensive reference journal, dedicated to disseminating the most innovative, systematic, topical and emerging theory, methods and applications on Situational Awareness (SA) across Cyber Systems, Cyber Security, Cyber Physical Systems, Computer Network Defence, Enterprise Internet of Things (EIoT), Security Analytics and Intelligence to students, scholars, and academia, as well as industry practitioners, engineers and professionals.

https://www.c-mric.com/journals/ijcsa

Editor-in-Chief: Dr. Cyril Onwubiko

C-MRiC Other Services

We provide a number of other and interrelated services, such as:

- Innovation, Research & Development ranging from national cyber security programmes, enterprise security management, information assurance, protection strategy & consultancy
- Customised & Professional Training
- Technology-inspired programmes, and undertake independent bespoke technology-based & survey-based research engagements
- Security Testing and Lab Experimentations
- Conference Organisation
- Printing and Publications
- Consultancy & Consortium-led collaborations

Contact Us

Centre for Multidisciplinary Research, Innovation and Collaboration (C-MRiC.ORG)

The Centre for Multidisciplinary Research, Innovation and Collaboration (C-MRiC) is a nonprofit non-governmental organisation.

The aim is to participate, encourage and promote collaborative scientific, industrial and academic inter-workings among individual researchers, practitioners, members of existing associations, academia, standardisation bodies, and including government departments and agencies.

The purpose is to build bridges between academia and industry, and to encourage interplay of different cultures.

C-MRiC is committed to outstanding research and innovation through collaboration, and to disseminate scientific and industrial contributions through seminars and publications. Its products range from conferences on advanced and emerging aspects of societal issues, ranging from Cyber security to environmental pollution, and from Health IT to Wearable, with the best of breeds of such contributions featuring in our journal publications.

C-MRiC is reliant on individual and corporate voluntary and free memberships to support its activities such as peer reviews, editorials, participating, organising and promoting conference and journal publications.

We collaborate with academia, industries and government departments and agencies in a number of initiatives, ranging from national cyber security, enterprise security, information assurance, protection strategy, climate control to health and life sciences.

We participate in academic and industrial initiatives, national and international collaborative technology-inspired programmes, and undertake independent bespoke technology-based & survey-based research engagements.

C-MRiC is free membership to both individuals and corporate entities; it is voluntary, open and professional.

Membership to C-MRiC entitles you free access to our publications, early sightings to research and innovations, and allows you to submit, request and pioneer research, conference or journal project through us. Members are selected based on expertise to support some of our activities on a voluntary basis, such as peer reviews, editorials, participating, organising and promoting conference and journal publications.

Address: C-MRiC.ORG
Email: submission@c-mric.org

Twitter: Follow @cmricorg

Web: http://www.c-mric.org