PhD projects (Cybersecurity)

- Evolving Cybercrime: Improving Threat Recognition and Response
- Designing Against Intimate Partner Violence in AI Companions
- Modelling Behavioural Change for Adaptive Biometrics
- Privacy in AI Companion / Assistant Technologies
- Scaffolding Student Learning through GenAI in Cybersecurity Education
- Validation of Cyber Threat Intelligence Leveraging Large Language Models

Evolving Cybercrime: Improving Threat Recognition and Response

Supervisor: Tasmina Islam / Kovila Coopamootoo

Project Description

Cybercrime continually evolves as offenders adapt their language, tactics and modes of interaction to new technologies and user behaviours. The growing use of generative AI may accelerate this process, making emerging threats harder both to detect computationally and to recognise by people.

This project will investigate how changes in malicious communications, such as phishing, scams and social engineering, influence threat recognition and response. Computational methods including natural language processing, machine learning, metadata analysis and temporal modelling will be used to identify how malicious narratives and behavioural patterns evolve over time. Human-centred studies will investigate how people recognise and interpret changing threats, the cues they rely on, common misconceptions, and factors that influence their security decisions.

The project will also investigate how people can be better prepared to recognise unfamiliar and changing threats. It will explore how computational findings about emerging threat patterns can inform cybersecurity learning and develop and evaluate approaches that support transferable threat-recognition and decision-making skills rather than reliance on known examples or fixed warning signs.

References

1. Islam, T., Becker, I., Posner, R., Ekblom, P., McGuire, M., Borrion, H., & Li, S. (2019). A socio-technical and coevolutionary framework for reducing human-related risks in cyber security and cybercrime ecosystems. In International Conference on Dependability in Sensor, Cloud, and Big Data Systems and
Applications (pp. 277-293).
2. Ng M, Bada M, Coopamootoo KPL (2023) . What We Do in the Shadows: How does Experiencing Cybercrime Affect Response Actions & Protective Practices? . Conference: 2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) vol. 00 , 659 - 672 .
3. J. Li and T. Islam, "Learncyber - Addressing Security and Privacy Concerns: Personalised Learning Tools," 2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), Yogyakarta, Indonesia, 2025, pp. 467-472,

Designing Against Intimate Partner Violence in AI Companions

Supervisor: Dr Ruba Abu-Salma and Dr Martim Brandao

Project Description

This project examines how sycophancy and aggression in AI companions are understood by users and by intimate partners who may be affected without using these systems themselves. It investigates how conversational patterns that affirm users' beliefs, simulate intimacy, or encourage harmful behavior may shape relationships and create risks that existing AI safety approaches overlook. Grounded in feminist standpoint epistemology, the research centers the knowledge of intimate partner violence survivors while attending to cultural differences and inequalities within each setting. The research will proceed through three connected phases. Interviews with users of AI companions and general-purpose models used for romantic conversations will explore interpretations of agreement, emotional validation, and influence. Workshops and interviews with survivors, recruited through trusted community organizations, will examine their understandings of these behaviors and the protections they consider meaningful. Findings will then inform co-design workshops with new user and non-user cohorts to develop a culturally differentiated corpus of harmful or contested interactions. Users, non-users, and AI safety researchers will collaboratively develop design recommendations and a working prototype, refined through iterative feedback. The project will contribute empirical accounts of relational harm, culturally grounded resources for safety evaluation, and practical design interventions addressing both user and non-user safety. Throughout, participation will prioritize informed consent, control over disclosure, and recognition of participants' knowledge and labor.

References

Position: Responsible AI for AI Companions Must Actively Combat Violence Toward Intimate Partners: https://openreview.net/forum?id=s6nxLbamu4

Modelling Behavioural Change for Adaptive Biometrics

Supervisor: Tasmina Islam

Project Description

Behavioural biometric systems identify individuals from interaction patterns such as keystroke and mouse dynamics. However, these behaviours evolve as people learn, develop proficiency and adapt to different tasks and contexts. Existing systems may treat such variation as behavioural drift, making it difficult to distinguish changes associated with legitimate skill development from contextual variation, identity change or potentially malicious behaviour.

This project will investigate how AI can disentangle relatively stable identity characteristics from behavioural changes associated with learning and developing proficiency. It will examine how behavioural signatures evolve during skill acquisition, which characteristics remain identity-discriminative, and whether learning-related changes persist or transfer across tasks.

The project will develop new computational methods for factorising sources of behavioural variation and determining when biometric adaptation is appropriate, drawing on representation and continual learning, uncertainty modelling and adversarial robustness. Existing behavioural-biometric datasets will be complemented by purpose-built, consented studies in which participants acquire unfamiliar digital skills over repeated sessions, enabling behavioural change to be related directly to developing proficiency. The research aims to advance understanding of how learning shapes digital behaviour while developing biometric systems that can accommodate legitimate human development without compromising identity assurance or security.

References

1. Nambiar, A. P., & Islam, T. (2025). Transformer-based multimodal biometric authentication using signature and EEG signals. In 2025 5th International Conference on Intelligent Cybernetics Technology and Applications (ICICyTA) (pp. 473—478). IEEE.
2. Wang, Y., & Islam, T. (2023). Addressing privacy and security concerns in online game account sharing: Detecting players using mouse dynamics. In ICPRAM 2023 — Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods, Volume 1 (pp. 864—871).
3. Subash, A., Song, I., Lee, I., & Lee, K. (2026). Adaptability of current keystroke and mouse behavioral biometric systems: A survey. Computers & Security, 160.

Privacy in AI Companion / Assistant Technologies

Supervisor: Kovila Coopamootoo

Project Description

AI systems are increasingly entering emotionally sensitive domains, including companionship, relationship advice, education guidance, mental-health support, and sustained personal conversation through agentic companions or assistants. Their integration of large language models (LLMs) facilitate coherent, human-like conversations, and are becoming more and more autonomous. However, they also raise significant privacy concerns, as well as trust, consent and agency challenges.

The PhD projects can include one or a combination, of the following objectives:

- Investigate data governance (and compliance) for AI companions/agents, and envision/design mechanisms to address gaps
- Investigate and design for privacy (/ trust / consent / agency)
o With a user centric perspective: Understand users, explore design possibilities, co-development
o Enforcement through privacy/security mechanisms implementation
- Assess risks factors and implications, e.g. to online safety, privacy, safety harms, in particular across human—AI interaction (HAI) and impacts on human—human relationships (HHI)
- Investigation of and calibrating human-AI decision-making
o Preferences, vulnerability, risk, design

Scaffolding Student Learning through GenAI in Cybersecurity Education

Supervisor: Hannah Cao / Ievgeniia Kuzminykh

Project Description

Scaffolding, a concept rooted in Vygotsky's Zone of Proximal Development (ZPD), refers to the instructional supports that help learners bridge the gap between their current abilities and learning goals. Traditionally, scaffolding has been provided by tutors, peers, or carefully designed instructional materials.
Generative AI (GenAI) tools such as ChatGPT are now transforming higher education. With the capacity to generate coherent text, explain complex concepts, and solve problems, these tools offer new opportunities for personalised learning. However, their adoption remains controversial. While they can enrich teaching and learning, concerns persist regarding academic integrity and the risk that students may bypass deeper cognitive engagement.

This project responds to these challenges by designing and evaluating a GenAI-driven scaffolding system for cybersecurity education. Rather than supplying direct answers, the system will prompt reflection, break down problems into manageable steps, and encourage exploration, drawing on established pedagogical approaches such as inquiry-based learning. By integrating advances in AI with the principles of effective scaffolding, this project aims to enhance cybersecurity education while contributing to broader debates about the ethical and pedagogical role of AI in higher education.

References

[1] Vygotsky, L.S.: Mind in Society: The Development of Higher Psychological Processes. Harvard University Press (1978)
[2] Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H., Wang, F.L.: Retrieval-augmented generation for educational application: A systematic survey. Computers and Education: Artificial Intelligence 8, 100417 (2025). https://doi.org/10.1016/j.caeai.2025.10041
[3] Liu, M., M'Hiri, F.: Beyond traditional teaching: Large language models as simulated teaching assistants in computer science. In: Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1. pp. 743—749. SIGCSE 2024, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3626252.3630789

Validation of Cyber Threat Intelligence Leveraging Large Language Models

Supervisor: Hannah Cao

Project Description

Cyber Threat Intelligence (CTI) is central to modern cybersecurity, equipping organisations to anticipate and mitigate evolving threats. Effective CTI enables proactive defence through timely incident response, patch management, and vulnerability awareness. However, the rapid advancement of Artificial Intelligence (AI), particularly Large Language Models (LLMs), has introduced new challenges. While LLMs offer powerful tools for processing the unstructured data that dominates CTI, they can also be exploited by malicious actors to generate fabricated intelligence at scale. Such disinformation threatens to mislead analysts, poison datasets, and erode the trust essential for collaborative defence.

Existing validation methods, which rely heavily on structured Indicators of Compromise (IoCs) or source reputation, are ill-equipped to handle nuanced, text-based intelligence of uncertain authenticity. This project addresses this critical gap by developing scalable methods to detect fabricated CTI. Specifically, it proposes leveraging LLMs themselves, guided by structured analysis and corroboration against external sources, to assess the likelihood of falsehood in unstructured reports.

The research will deliver a robust framework for validating CTI, strengthening resilience against AI-driven disinformation campaigns. By enhancing confidence in intelligence sharing, the project aims to reduce the risks of misallocated resources and flawed security postures, ultimately supporting more effective cybersecurity practices. Beyond technical validation, the project will also provide an educational platform for cybersecurity practitioners, equipping them with the knowledge and tools to recognise and counteract fabricated CTI in practice.

References

[1] P. Ranade, A. Piplai, S. Mittal, A. Joshi, T. Finin, Generating fake cyber threat intelligence using transformer-based models, 2021 International Joint Conference on Neural Networks (IJCNN), 2021, pp. 1—9. doi:10.1109/IJCNN52387.2021.9534192.

[2] S. Mitra, S. Neupane, T. Chakraborty, S. Mittal, A. Piplai, M. Gaur, S. Rahimi, LOCALINTEL: Generating organizational threat intelligence from global and local cyber knowledge, 2025. arXiv:2401.10036.

[3] Janku, S., Xiao, H., & Caris, T. (2025). An Educational and Validation Tool for Cyber Threat Intelligence Leveraging Large Language Models. In 3rd International Workshop on Cyber Security Education for Industry and Academia (CSE4IA 2025)