PhD projects in the Department of Informatics, AY 25-26 — Cybersecurity

The PhD projects listed below will be considered for 2025/26 studentships available in the Department of Informatics to start on 1 October 2025 or later during the 2025/26 academic year.

Please note that this list is not exhaustive and potential applicants can alternatively identify and contact appropriate supervisors outlining their background and research interests or proposing their own project ideas.

Each project is designated for a single student, meaning it can only be assigned to one successful applicant. Some projects come with allocated studentships, while others are eligible for "unallocated" studentships. Applicants who apply for projects with allocated studentships and are selected will be offered a full studentship. In the project list, these are marked as "studentship allocated." Applicants chosen for other projects will compete for the unallocated studentships.

We welcome applications from students who have secured, or are applying for, or plan to apply for other funding (within other studentships internal to the university or external schemes) and from self-funded students. See also this list of funding opportunities available at King's for post-graduate research in Computer Science.

PhD projects

- Leveraging Language Models for Contextual Vulnerability Identification
- Designing novel privacy IxD mechanisms in mobile health apps
- Inclusive and Accessible Cybersecurity Education
- Investigating LLM-based Generative AI Applications in Cybersecurity
- Privacy in the Internet of Things
- Designing and Developing a framework for responsible security and privacy practices for GenAI Tools

Leveraging Language Models for Contextual Vulnerability Identification

Supervisor: Maher Salem

Areas: Artificial Intelligence (symbolic AI, logic, etc.), Machine learning / Deep learning, Cybersecurity, Systems (software engineering, programming)

Project Description

As software systems grow increasingly complex, the need for effective vulnerability detection methods becomes paramount. Traditional static analysis tools often struggle to identify context-specific vulnerabilities due to their reliance on predefined patterns and rules. This research proposes leveraging advanced language models, such as transformers, to enhance the identification of vulnerabilities in software code by understanding its context. The central idea of this topic is to explore how large language models (LLMs) can be trained to analyze code not merely as isolated snippets but as part of a larger context. By fine-tuning LLMs on extensive datasets that include both vulnerable and secure code, the model can learn to recognize subtle patterns and interactions that indicate potential vulnerabilities. This approach aims to move beyond conventional methods by incorporating an understanding of how different code components interact with each other, thereby improving detection accuracy. The research will involve several key phases. First, a comprehensive dataset will be curated, containing various programming languages and a range of vulnerability types, such as SQL injection, cross-site scripting, and buffer overflows. This dataset will serve as the foundation for training the language models. Next, the study will focus on developing a framework that integrates the LLMs into an existing vulnerability detection pipeline, allowing for real-time analysis and feedback during the software development lifecycle. Furthermore, the research will explore the effectiveness of different model architectures and training techniques, including transfer learning and few-shot learning, to optimize performance. By evaluating the models against established benchmarks and real-world codebases, the study aims to quantify improvements in vulnerability detection rates compared to traditional static analysis tools. Another important aspect of this research is the interpretability of the model's predictions. It is crucial for developers to understand why a particular piece of code was flagged as potentially vulnerable. Therefore, the study will investigate methods to enhance the transparency of LLMs, providing explanations that can guide developers in addressing identified vulnerabilities. Ultimately, this research seeks to contribute to the field of cybersecurity by providing a novel approach to vulnerability detection that leverages the capabilities of modern AI. By harnessing the contextual understanding of language models, the goal is to create more robust and intelligent tools that can significantly enhance software security, helping developers proactively identify and mitigate vulnerabilities before they can be exploited.

References

DOI: 10.1145/3460318.3464820
DOI: 10.1145/3404835.3462831
DOI: 10.1145/3397670
DOI: 10.1109/TSE.2020.2986860

Designing novel privacy IxD mechanisms in mobile health apps

Supervisor: Ruba Abu-Salma

Areas: Cybersecurity, Human-centred computing (human-computer interaction)

Project Description

Mobile health (mHealth) apps provide a wide range of benefits. However, they collect a significant amount of sensitive user medical data, posing privacy risks to users. Whilst legislation exists to protect users' medical and health data, levels of protection vary across countries. Users may also overestimate these legal protections and, as a result, trust mHealth apps unduly with their health data---or may lose trust due to lack of transparency, and avoid mHealth despite its benefits. Efforts have been made to restrict mobile apps' data practices and improve transparency, but are not always efficacious at addressing problematic app behavior or improving users' understanding of app data practices. This project aims to design and evaluate interaction design (IxD) mechanisms that enable users to mitigate the privacy risks associated with the use of mHealth apps while allowing them to continue benefiting from such apps.

References

https://kclpure.kcl.ac.uk/ws/portalfiles/portal/251441290/chi24-626-21.pdf

Inclusive and Accessible Cybersecurity Education

Supervisor: Tasmina Islam

Areas: Cybersecurity, Human-centred computing (human-computer interaction), Education

Project Description

People are spending more time online due to the increasing digitisation of society. This also means the security measures need to be stronger and awareness of the cybersecurity risks, and implications is crucial, particularly for vulnerable and underrepresented groups, such as children, ethnic minority communities, and people with disabilities, who may face increased risks due to limited access to tailored resources and education. This project aims to address these challenges by developing adaptive learning environments that cater to individual needs, offering age-appropriate content for children and culturally relevant, accessible material for diverse communities. Learners will engage with realistic cybersecurity scenarios, equipping them with essential skills like identifying phishing attacks, managing privacy, and securing personal information. Using a combination of qualitative methods, such as interviews and focus groups, and quantitative surveys, the project will study the specific cybersecurity challenges faced by these populations. The knowledge gained from this research will inform the development of scalable educational tools, integrating AI for personalised learning and immersive technology to create engaging, hands-on experiences. By addressing the cybersecurity education gap, this initiative seeks to empower all learners with the knowledge and skills to safely navigate the digital world, enhancing digital safety, privacy, and security for all while promoting a more inclusive digital landscape. Prospective students can discuss options with the supervisor.

References

1. Hedges, M., & Islam, T. (2024). VirSec — Immersive Security Training within Virtual Reality. In 17th International Conference on Advanced Visual Interfaces: 2nd International Workshop on CyberSecurity Education for Industry and Academia (CSE4IA 2024)
2. Islam, T & Zou, Y 2023, ChildSecurity: A Web-based Game to Raise Awareness of Cybersecurity and Privacy in Children. in Cybersecurity Challenges in the Age of AI,Space Communications and Cyborgs.

Investigating LLM-based Generative AI Applications in Cybersecurity

Supervisor: Ievgeniia Kuzminykh/Hannan Xiao

Areas: Artificial Intelligence (symbolic AI, logic, etc.), Cybersecurity, Natural Language Processing, Human-centred computing (human-computer interaction)

Project Description

The rapid development and deployment of GPT-based agents in cybersecurity mark a significant leap forward in approaching digital security challenges from a practitioner standpoint. Under this topic you can explore the ways generative AI is impacting the cybersecurity industry, from both sides, such as gen AI for security and security of gen AI. From one side, malicious attackers are seizing the potential of generative AI to launch cyber attacks that are harder to detect and defend against. OWASP top 10 for Generative AI [1] lists out the top 10 vulnerabilities impacting the applications usilising LLM. Prompt Injection, Insecure Output Handling and Data Poisoning take the top 3 spots and are also the root causes for the other type of vulnerabilities (Overreliance, Insecure Plugins etc) as shown in research paper [2]. From another side, Gen AI is also helping make security teams more accurate, efficient, and productive in defending their organisations. Examples of utilising of generative AI for security operations could be [3] : - Supplementing understaffed security teams - Detecting threats in real time - Improving incident response. The potential topics in this project area could include but not limited to: 1. Optimisation of prompts for security related topics. Through clever prompt engineering (called jailbreaking [4]), LLMs can be made to reveal internal mechanisms, share private data, produce offensive speech, or perform unintended workloads. LLMs thus pose a security risks [1, 5, 6]. 2. Prompt injection detection mechanisms. 3. Ensuring online safety using LLM. AI seems like the perfect response to the growing challenges of content moderation on social media platforms: the immense scale of the data, the relentlessness of the violations, and the need for human judgments without wanting humans to have to make them. The paper [7] elaborates on the topic of prompt/response classifiers. The prompts and answers could be classified into groups such as safe and harmful. Typical examples of a harm would be Child Safety, Exfiltrating PII/SPII, Sexually Explicit Content, Malicious/Dangerous content. 4. Content moderation using LLM. Similar to previous but can be extended to the detection of harassment and throlling [8]. 5. Understanding Generative AI for Cloud Security. Generative AI can make new data from existing patterns. For cloud security, this means it has the potential to: 6a. Simulate Threat Scenarios: Generative AI can create realistic threat scenarios, allowing security teams to test and validate their Cloud infrastructure's resilience. By simulating potential attack vectors, organizations can proactively identify vulnerabilities and take steps to ensure they are protected against them before they are exploited. 6b. Optimize Security Configurations: AWS offers a number of services, each with its own set of security configurations. With Generative AI, we can analyze existing configurations, simulate various combinations, and ask Generative AI to provide recommendations based on our specific needs. 6c. Enhance Monitoring and Alerts: By training on historical security logs and events, Generative AI can predict potential security breaches or anomalies. The key word here is "potential." Knowing what "could" happen allows security teams time to prepare and allows for more rapid action to be taken. 6. Understanding Generative AI for firewall optimisation Generative AI could simulate web traffic patterns based on your historical log data and compare that to your existing WAF or firewall rules, ensuring that malicious requests are blocked while legitimate traffic flows seamlessly. 7. 10. Understanding of Gen AI for Qualitative audit of security policies Each organisation is governed by a security policy, which technically or conceptually specifies a number of guidelines for ensuring IT security. You will investigate whether GenAI can be employed to translate a security policy for wider staff [9, 10]..

References

[1] https://owasp.org/www-project-top-10-for-large-language-model-applications/

[2] Knowledge Bases and Language Models: Complementing Forces. https://link.springer.com/chapter/10.1007/978-3-031-45072-3_1

[3] https://secureframe.com/blog/generative-ai-cybersecurity

[4] Liu, Y., et al.: Jailbreaking ChatGPT via prompt engineering: an empirical study. arXiv preprint arXiv:2305.13860 (2023)

[5] https://simonwillison.net/2023/May/2/prompt-injection-explained/

[6] https://www.geeksforgeeks.org/what-is-jailbreak-chat/

[7] Llama Guard: LLM-based Input-Output Safeguard for Human-AI Conversations https://arxiv.org/html/2312.06674v1

[8] Content moderation, AI, and the question of scale https://journals.sagepub.com/doi/full/10.1177/2053951720943234

[9] Introducing Rules Genie: Generative AI for Automating Policy Creationhttps://www.opsmx.com/blog/introducing-rules-genie-generative-ai-for-automating-policy-creation/

[10] The LLM Police: Between Firewalls and Policies for Generative AI https://www.fairly.ai/blog/the-llm-police-between-firewalls-and-policies-for-generative-ai

Privacy in the Internet of Things

Supervisor: Maribel Fernandez

Areas: Cybersecurity, Systems (software engineering, programming)

Project Description

Data Collection policies are used to restrict the kind of data transmitted by devices in the Internet of Things (e.g., health trackers, smart electricity meters, etc.) according to the privacy preferences of the user. The goal of this project is to develop cloud/IoT architectures with integrated data collection and data sharing models, to allow users to specify their own policies and trade data for services. For this, new data collection and data sharing models will have to be developed, with appropriate user interfaces, policy languages, and policy enforcement mechanisms. An important aspect of the project is the development of policy recommendation systems that can suggest/create policies based on user profiles, making privacy an integral part of the system (according to the privacy-by-design" IoT paradigm).

References

A Privacy-Preserving Architecture and Data-Sharing Model for Cloud-IoT Applications, Maribel Fernandez; Jenjira Jaimunk; Bhavani Thuraisingham IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 4, pp. 3495-3507, 1 July-Aug. 2023, doi: 10.1109/TDSC.2022.3204720.

Designing and Developing a framework for responsible security and privacy practices for GenAI Tools

Supervisor: Maher Salem

Areas: Cybersecurity, Human-centred computing (human-computer interaction), Artificial Intelligence (symbolic AI, logic, etc.)

Project Description

Generative Artificial Intelligence (GenAI) technologies have transformed human life, impacting areas such as healthcare, education, and social interactions. While GenAI tools offer creative content generation, data synthesis, and automation benefits, they also introduce significant challenges that may negatively impact individuals and communities, particularly vulnerable groups. These challenges include privacy and ethical issues, legal risks, bias and discrimination, misinformation, and inaccurate outputs. Despite the importance of these issues, there is limited empirical research on users' experiences and views regarding GenAI security and privacy. This project aims to apply both qualitative and quantitative methods to investigate how users interact with GenAI tools, the reasons behind their use, and how these experiences shape perceptions of GenAI. A key component of the study will involve understanding users' mental models of the benefits and risks of GenAI, the educational resources they use to assess potential risks, and the protective measures they adopt. By examining users' learning processes and resources, this research will provide insights into gaps in GenAI literacy. The study's educational goals include developing targeted resources and practical guidelines to improve GenAI literacy among diverse groups. Empirical insights from this research will guide the design of safeguards for GenAI technologies and inform curriculum and policy recommendations, enabling institutions to equip students, educators, and users with the skills to navigate GenAI responsibly and securely.