Candidate Information

Position:	Research Fellow in Agentic AI Security
School/Department:	School of Electronics, Electrical Engineering and Computer Science
Reference:	25/112782
Closing Date:	Monday 18 August 2025
Salary:	£41,519 per annum
Anticipated Interview Date:	Monday 25 August 2025
Duration:	25 months

JOB PURPOSE:

Assurance and verification offerings for AI are fragmented and target specific areas of AI lifecycle (e.g. model assurance or prompt injections) without offering a unified view of the security of an AI System. The role involves performing research to create a configurable AI agentic framework that 1) integrates with the full security lifecycle of a system whilst 2) providing innovative verification and assurance options.

Specifically, the role will involve investigating:

• Excessive Agency and discovery of data and other misconfigured attack vectors an AI system can exploit.

• Misuse, by analysing access and system logs using anomaly detection.

• Sensitive Data Exposure, by helping identify memorised data in finetuning settings, misconfigured access control, or leaks in RAG settings.

• Supply chain by integrating existing MLSecOps, DevSecops and advance techniques for 3rd party model assessment as part of provenance.

MAJOR DUTIES:

1. Undertake high-quality and novel research in the areas of AI agentic systems, excessive agency, misuse, sensitive data exposure, and supply chain assurance.
2. Design, develop, and refine experiments to evaluate anomaly detection performance.
3. Carry out analyses, critical evaluations, and interpretations using methodologies and other techniques appropriate to AI security.
4. Present regular progress reports on research to members of the research group or to external audiences to disseminate and publicise research findings.
5. Prepare, often in consultation with the line manager, material for publication in national and international journals and presentations at international conferences such as IEEE Security and Privacy, ACM in Computers and Communication Security.
6. Produce high-quality research outputs consistent with project aims and commensurate with the career stage.
7. Undertake supplementary duties relevant to the success of the project including administrative duties and additional training and development activities as required.

ESSENTIAL CRITERIA:

1. Have or be about to obtain a PhD in computer science, engineering, mathematics or physical sciences area.
2. Recent high quality research experience in machine learning/AI or cybersecurity, or both, as evidenced by a strong track record of publications in leading journals and conferences in relevant areas.
3. Software programming skills.
4. A consummate team player who is open-minded and is prepared to work closely with other members of a large multidisciplinary research and development team, as well as with industrial collaborators.
5. Able to visit collaborative partners and to attend meetings and conferences nationally and internationally as requested.
6. On-site presence required in accordance with QUB policy (currently for a minimum 3 days per week).

DESIRABLE CRITERIA:

1. Normally have or be about to obtain a PhD in the area of machine learning/AI, cybersecurity or cyber AI.

2. Hands-on experience in clustering and anomaly detection methods.
3. Experience in deep learning for natural language processing.
4. Experience in AI security and/or Trustworthy AI.
5. Experience in AI agentic systems.
6. Experience in initiating and developing research plans.
7. Experience in collaborating with industry.
8. Proficient in Python.
9. Experience using PyTorch, Keras, and/or TensorFlow. Jupyter Notebook.
10. Ability to design, train and test machine learning/AI systems using appropriate methodologies and datasets.
11. Familiarity with the theory of Machine Learning fundamentals (Statistics, Optimization, Linear Algebra, Partial Derivative Equations, etc.)
12. Demonstrable commitment to the purpose and objectives of CSIT and Cyber-AI Hub.
13. Ability to assess and organise resources.
14. Evidence of being a strong communicator with excellent oral and written communication skills.
15. Demonstrate resilience and the ability to work in a fast-changing environment with competing priorities.
16. Demonstrates a high degree of integrity, honesty and openness in professional conduct.