

Candidate Information

Position: Research Assistant in Agentic AI Security
School/Department: School of Electronics, Electrical Engineering and Computer Science
Reference: 26/113642
Closing Date: Monday 5 October 2026
Salary: £36,366 to £41,727 per annum
Anticipated Interview Date: Thursday 8 October 2026
Duration: Fixed Term - Full Time; available for approx. 12 months or until 30 September 2027, whichever is soonest

JOB PURPOSE:

AI coding assistants and automated code-review agents are being adopted rapidly across industry and government, introducing new attack surfaces — source code crafted to evade AI reviewers, and poisoned or backdoored LLMs embedded in code-generation and review pipelines — that can propagate directly into the software supply chain. The post-holder will help the project team discover, characterise and mitigate these vulnerabilities. Typical duties include implementing proof-of-concept attacks and defences against coding and review agents, curating vulnerable-code datasets, and contributing to papers, demos and industry outreach. The post is ideal for an MSc graduate (or exceptional BSc Hons) who wants to gain deep, publishable experience before a PhD or industrial R&D career.

MAJOR DUTIES:

1. Carry out literature & GitHub reconnaissance on code-agent attacks, LLM code security and software-supply-chain threats.
2. Build attack harnesses that craft adversarial / evasive source code and probe coding and code-review agents.
3. Investigate poisoned or backdoored LLMs in code-generation and review pipelines, and how triggers are embedded in code context.
4. Evaluate both OSS and proprietary models / agents via APIs, respecting T&Cs.
5. Prototype lightweight defences and detectors (e.g. review-time checks, provenance / attestation, anomaly detection) and quantify overheads.
6. Maintain reproducible code & datasets, including curated vulnerable-code corpora, in the group repo.
7. Present weekly progress; contribute figures and ablation tables for papers.
8. Co-author papers / blogposts; help prepare demo videos for industry partners.

ESSENTIAL CRITERIA:

1. 2:1 or above BSc in Computer Science, Cyber-security, AI or related.
2. Experience working in an academic research environment.
3. Significant Python programming experience.
4. Familiarity with at least one LLM framework or API (Hugging Face, llama.cpp).
5. Solid software-engineering and version-control (Git) practice.
6. A consummate team player who is open-minded and prepared to work closely with other members of a large multidisciplinary research and development team, as well as with industrial collaborators.
7. Might require travelling to meet with partners.

DESIRABLE CRITERIA:

1. Sc or MSc with thesis on AI / software security, program analysis or secure software engineering.
2. Experience in AI security, software-supply-chain security and/or secure code review.
3. Knowledge of program analysis (static/dynamic), compilers or code-security tooling.
4. Familiarity with code-generation / review agents (e.g. Copilot-style).
5. Evidence of being a strong communicator with excellent oral and written communication skills.

6. Demonstrates a high degree of integrity, honesty and openness in professional conduct.