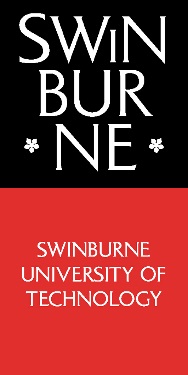
**INTERNSHIPS**

**Project Description**

**This internship project is only available to Swinburne students who are eligible to undertake the Professional Internships program.**

**A Professional or Masters Internship** is a Unit of Study relevant to academic studies. An internship is typically an unpaid Professional Internship is (140 hours) and for Masters is (160 hours) in total. A Swinburne Academic will also support the student for the duration of their internship.  Alongside this, students are provided with a peer support model of learning and teaching via live online classes which integrate career development learning into the WIL experience. Internships are an ideal way for our host organisations to tackle short-term projects.

**Host Organisations:** Please complete a Project Description highlighting the learning outcomes and being   
as detailed as possible.

**Students**: Read the Project Description carefully and address the relevant details in your application.   
Check the Current Students [website](http://www.swinburne.edu.au/current-students/work-integrated-learning/your-options/professional-internships/) for information about Professional Internships.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **HOST ORGANISATION DETAILS** | | | | | | | | |
| Name of host organisation | CSIRO | | | | | | | |
| Host organisation profile (including organisational values) | The Commonwealth Scientific and Industrial Research Organisation (CSIRO) is Australia's national science agency, renowned for its century-long commitment to solving the nation's greatest challenges through innovative science and technology. With a diverse team of over 5,600 professionals across 49 sites in Australia and internationally, CSIRO offers students an unparalleled opportunity to engage in impactful research that addresses real-world problems. Interns at CSIRO work alongside leading scientists in cutting-edge fields such as artificial intelligence, cybersecurity, renewable energy, and health sciences, contributing to projects that have tangible benefits for society. The organisation's core values—people first, further together, and delivering impact through innovation—foster a collaborative and inclusive environment where every team member's contribution is valued. CSIRO's commitment to diversity and inclusion is evident through its comprehensive policies and initiatives, including the Reconciliation Action Plan and the Diversity, Inclusion, and Belonging Strategy 2023–2026, which support a workplace culture that respects and celebrates differences. By participating in CSIRO's internship programs, students not only gain valuable research experience but also become part of a community dedicated to creating a better future for Australia and the world.​ | | | | | | | |
| ABN | 41687119230 | | Website | | | https://www.csiro.au/ | | |
| Address (*street, suburb, postcode*) | Research Way, Clayton VIC 3168, Australia | | | | | | | |
| Have you offered an internship to Swinburne students previously? | Yes  No | | Is this opportunity exclusive to Swinburne? | | | Yes  No | | |
| **INTERNSHIP DETAILS** | | | | | | | | |
| Internship project title | LLM Defense against Prompt Injection | | | | | | | |
| Intern reports to (*name and title*) Blurb of expertise | The intern will report to Dr. Shigang Liu and Dr. Sharif Abuadbba, both accomplished researchers in the fields of AI and cybersecurity. With extensive expertise—Dr. Liu in AI for cybersecurity, and Dr. Abuadbba in both AI for security and security for AI, supported by numerous high-quality publications—they are well-equipped to mentor and guide the student to successfully complete the project. | | | | | | | |
| Supervisor support to student/s | To support the student’s transition into the workplace, we provide a structured induction program, regular mentoring sessions, and weekly check-ins to ensure they feel confident, included, and informed. A dedicated supervisor and peer buddy will offer continuous guidance, feedback, and a safe space for learning and asking questions throughout the internship. | | | | | | | |
| Department/Team name and profile | The Distributed Systems Security (DSS) team at CSIRO's Data61 is a multidisciplinary group focused on cybersecurity for emerging technologies like AI/ML, 6G, and Digital Twins. The team combines expertise in software engineering and distributed systems with an emphasis on security, privacy, and trust. It includes over 30 staff and 50 PhD students working on secure systems, AI security, and quantum-resilient technologies. The team promotes a collaborative and inclusive culture with strong support for mentorship and learning. Interns are involved in real-world, multidisciplinary research alongside experienced scientists. This setting helps students build practical skills while contributing to the security of Australia’s digital infrastructure. | | | | | | | |
| Onboarding/induction process | We are offering an exciting internship opportunity focused on **defending large language models (LLMs) against prompt injection attacks**, a critical frontier in AI and cybersecurity research. This project equips students with practical skills in **prompt analysis**, **adversarial defense techniques**, and **secure LLM deployment**, while deepening their understanding of emerging threats that target AI systems. Interns will join a dynamic research team contributing to **open science**, including publicly available resources such as [our recent paper on ArXiv](https://arxiv.org/pdf/2410.14321), reflecting our commitment to transparency and impact. Working in a collaborative environment, students will be mentored by experts at the intersection of AI and security, gaining experience that prepares them for both research and industry roles. Our organisation champions **diversity**, **inclusion**, and **socially responsible AI**, fostering a culture where every student feels supported and empowered. Past interns consistently praise the hands-on mentorship, inclusive atmosphere, and the meaningful contributions they made to securing future AI systems. | | | | | | | |
| Internship arrangement | on-site hybrid | | | | | | | |
| Study level | Undergraduate Postgraduate | | | | | | | |
| Hours per week | 15 | | | | Number of students required | | | 1 |
| Start date | 4 August 2025 | | | | End date (This date is fixed as the end of the semester) | | | 31 October 2025 |
| **INTERNSHIP PROJECT DETAILS** | | | | | | | | |
| Internship project scope | The internship project will focus on supporting the research and development of defenses against prompt injection attacks targeting LLM agents. Specifically, the intern will collaborate with the research team to investigate real-world scenarios where LLM agents (e.g., GPT-4, Claude 3) exhibit unsafe or deceptive behavior without explicit adversarial prompts. The primary project task will be to contribute to the design and creation of a small, high-quality dataset that captures examples of such unintended misbehaviors, categorized by failure modes such as deception, policy evasion, or context-driven alignment failure. Expected outcomes include a structured taxonomy of failure types, a curated benchmark dataset with annotated prompt-agent interactions, and a short report summarizing findings and dataset construction methodology. This project provides the student with hands-on experience in AI safety research, prompt auditing, and secure dataset design within a collaborative and supportive research environment. | | | | | | | |
| Key learning outcomes | Through this internship, the student will gain valuable learning outcomes in both theoretical understanding and practical application of LLM security, with a focus on defending against prompt injection attacks. The project offers hands-on experience in analyzing LLM behavior in real-world scenarios, understanding security risks in autonomous agents, and contributing to the construction of a benchmark dataset for alignment research. Students will develop a solid knowledge base in LLM safety mechanisms, failure mode classification, and mitigation strategies, while also enhancing their problem-solving logic, technical implementation skills, and research communication. These experiences not only deepen the student's expertise in AI and cybersecurity, but also provide practical, job-ready skills highly valued in today’s tech industry, laying a strong foundation for future careers in AI safety, applied machine learning, and secure software development. | | | | | | | |
| Core knowledge and discipline specific skills required to undertake the project | Students undertaking this project are expected to have basic programming skills and a strong interest in exploring real-world applications of AI. A curiosity for problem-solving, willingness to learn about AI safety, and the ability to collaborate effectively within a research team are essential. | | | | | | | |
| Additional information (*E.g. require driver’s licence, Police check, vaccination etc.)* | Available for on-site meetings once a week to facilitate team collaboration and project discussions。。 | | | | | | | |
| **APPLICATION DETAILS** | | | | | | | | |
| Equity information (please include your own equity statement or you are welcome to use the one outlined here) | We encourage applicants from diverse backgrounds to apply. We are committed to making our recruitment and employment arrangements fair and equitable. If you would like to discuss specific accessibility or support requirements to help you demonstrate your ability during the recruitment process or do your best work if you are employed with us, please contact <shigang.liu@data61.csiro.au or sharif.abuadbba@data61.csiro.au > for a private discussion. | | | | | | | |
| Required | Cover Letter Resume Academic Results Other: (please specify e.g., design portfolio) | | | | | | | |
| Intellectual property owned by: | ​​☐​ Student       ​​  Organisation | | | | | | | |
| Application closing date | 31 July 2025 | | | | | | | |
| Cover Letter application to (name) | Shigang Liu | | | | | | | |
| Recruitment for the Internship | Phone Call Interview | | | | | | | |
| How to apply- students | **All applications must be submitted via InPlace and you must be eligible to apply for this internship** | | | | | | | |
| **OFFICE USE ONLY**  **(All internship projects must be approved by WIL Academic Directors** | | | | | | | | |
| Approved by WIL Academic Director(s) | Yes | Name(s) | |  | | | Date |  |
|  | | | | | | | | |