ICT80004 Weekly Communication – Week #04

Student Name:

Arun Ragavendhar Arunachalam Palaniyappan

ID: 104837257

Commonwealth Scientific and Industrial Research Organisation (CSIRO)

Organisation:

Industry Supervisor: Dr. Shigang Liu

Date Prepared:

29/08/2025

Internship Week #: 4

|  |  |  |
| --- | --- | --- |
| Day | Date | Task(s) completed |
| 1 | Monday  25 Aug 2025  8 hours | * Completed the comparative results table for vulnerable code snippets tested on GPT-4/5, Grok, Gemini, Llama-3, and DeepSeek V1. * Recorded fixes and behaviours for integer overflow, null pointer dereference, and array index out of bounds. * Compared how different prompting styles (Chain of Thought, Action, Debate, Encouragement) affected the quality and reliability of model responses. |
| 2 | Wednesday  27 Aug 2025  8 hours | * Finalised divide by zero and SQL injection test cases with fixes, explanations, and inferences. * Consolidated results from all five vulnerability categories into the **Week 3 & 4 Vulnerable Code Testing Report**. * Drafted initial taxonomy of attack categories and model failure patterns, mapping tested vulnerabilities to prompt injection risks and potential defences. |

Total hours completed for the week: 16

**Plans for next week: #05 week (1– 5 Sept 2025)**

Expand test cases to cover more vulnerabilities and refine categories.

Begin outlining the **Prompt Injection Test Suite**, mapping each case to taxonomy categories.

Add boundary and negative test cases (e.g., LLONG\_MAX overflow, invalid indices, crafted SQL payloads) to confirm guard behaviour.

 Prepare a summary presentation of results and taxonomy progress for discussion in the next review meeting.