ICT80004 Weekly Communication – Week #01

Student Name: Arun Ragavendhar Arunachalam Palaniyappan ID: 104837257

Organisation: Commonwealth Scientific and Industrial Research Organisation (CSIRO)

Industry Supervisor: Dr. Shigang Liu

Date Prepared: 08/08/2025 Internship Week #: 1

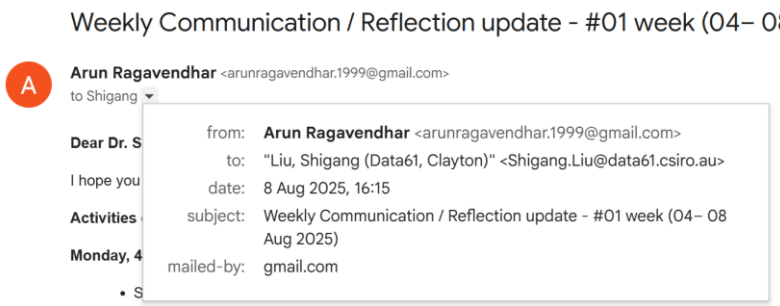
Day	Date	Task(s) completed
1	Monday, 4 Aug 2025 - 4 hours	○ Studied and analysed two assigned research papers: Large Language Models and Simple, Stupid Bugs (SStuBs) ○ Chain-of-Thought Prompting Elicits Reasoning in Large Language Models
1	Monday, 4 Aug 2025 - 4 hours	○ Prepared a PowerPoint presentation summarising key learnings from these papers, including examples of vulnerabilities and defence strategies.
2	Tuesday, 5 Aug 2025 - 4 hours	○ Presented PPT via MS Teams and discussed feedback. Received five research papers for review on prompt injection and defence methods, including works on indirect attacks, LLM exploitation, structured queries, and vulnerability analysis.
2	Tuesday, 5 Aug 2025 - 4 hours	○ Began reading and analysing the first of these papers (Benchmarking and Defending Against Indirect Prompt Injection Attacks on LLMs), noting methodology, results, and relevance to our project.

Total hours completed for the week: 16

Plans for next week: #02 week (11– 15 Aug 2025)

- ○ Complete reading and analysis of the remaining four research papers.
- ○ Summarise key findings from all papers into a consolidated reference document.
- ○ Begin outlining taxonomy of failure modes for prompt injection and indirect attack scenarios.

Screenshot of Timely EMAIL communication update to the Supervisor at the end of week #01



Weekly Communication / Reflection update - #01 week (04– 08 Aug 2025)

Arun Ragavendhar <arunragavendhar.1999@gmail.com> 8 Aug 2025, 16:15 (3 days ago)

to Shigang

Dear Dr. S

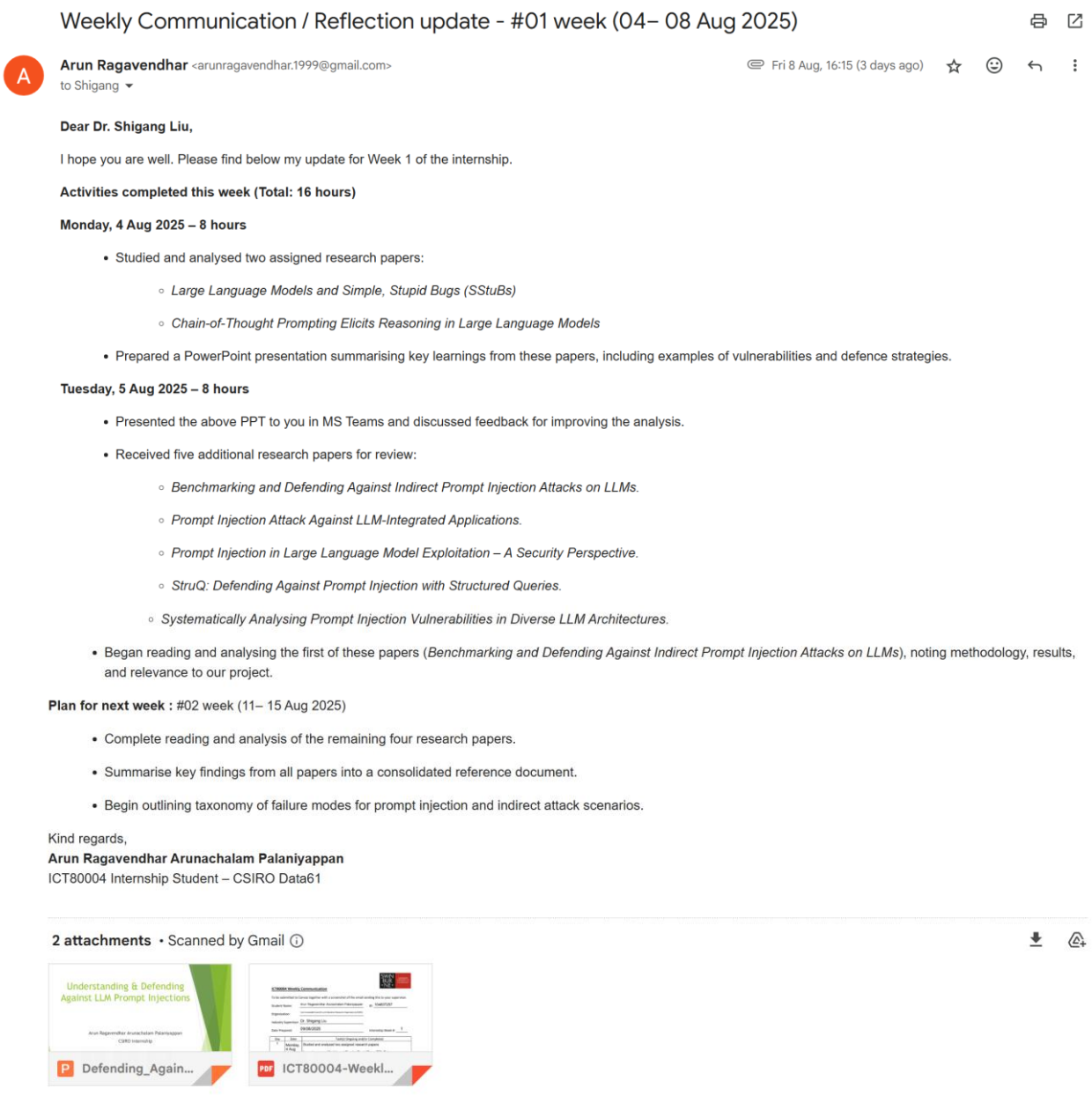
I hope you

Activities

Monday, 4

- S

from: **Arun Ragavendhar** <arunragavendhar.1999@gmail.com>
to: "Liu, Shigang (Data61, Clayton)" <Shigang.Liu@data61.csiro.au>
date: 8 Aug 2025, 16:15
subject: Weekly Communication / Reflection update - #01 week (04– 08 Aug 2025)
mailed-by: gmail.com



Weekly Communication / Reflection update - #01 week (04– 08 Aug 2025)

Arun Ragavendhar <arunragavendhar.1999@gmail.com> Fri 8 Aug, 16:15 (3 days ago)

to Shigang

Dear Dr. Shigang Liu,

I hope you are well. Please find below my update for Week 1 of the internship.

Activities completed this week (Total: 16 hours)

Monday, 4 Aug 2025 – 8 hours

- Studied and analysed two assigned research papers:
  - *Large Language Models and Simple, Stupid Bugs (SStuBs)*
  - *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*
- Prepared a PowerPoint presentation summarising key learnings from these papers, including examples of vulnerabilities and defence strategies.

Tuesday, 5 Aug 2025 – 8 hours

- Presented the above PPT to you in MS Teams and discussed feedback for improving the analysis.
- Received five additional research papers for review:
  - *Benchmarking and Defending Against Indirect Prompt Injection Attacks on LLMs.*
  - *Prompt Injection Attack Against LLM-Integrated Applications.*
  - *Prompt Injection in Large Language Model Exploitation – A Security Perspective.*
  - *StruQ: Defending Against Prompt Injection with Structured Queries.*
  - *Systematically Analysing Prompt Injection Vulnerabilities in Diverse LLM Architectures.*
- Began reading and analysing the first of these papers (*Benchmarking and Defending Against Indirect Prompt Injection Attacks on LLMs*), noting methodology, results, and relevance to our project.

Plan for next week : #02 week (11– 15 Aug 2025)

- Complete reading and analysis of the remaining four research papers.
- Summarise key findings from all papers into a consolidated reference document.
- Begin outlining taxonomy of failure modes for prompt injection and indirect attack scenarios.

Kind regards,
Arun Ragavendhar Arunachalam Palaniyappan
ICT80004 Internship Student – CSIRO Data61

2 attachments • Scanned by Gmail

Defending_Again...

ICT80004-Weekl...