PROJECT STATUS REPORT

*WIL Project Milestone 1*

## **Group ID**

WIL Project 41

## **Project Aim Statement**

As an RMIT Student I want to use a chatbot to quickly get answers from RMIT Connect services So that I do not have to wait long for emails or spend time searching for information

## **Table of Group Members**

|  |  |  |
| --- | --- | --- |
| Student ID | Name | Skills/Strengths |
| s4119966 | Patricia Celon | Data Visualization  (Python, R, Power BI), Data Scraping, EDA, Cloud Computing (Azure, AWS), SQL (T-SQL, Spark SQL) |
| s4134044 | Natchaya Chanbanchee | EDA, Data Visualisation (Python, R, Power BI), SQL |
| s3943743 | Justin Nguyen | Machine learning (Python – scikit-learn, NLP, embeddings, classification), Data visualisation (Python – matplotlib, seaboard; R – ggplot2, plotly), SQL (Azure, T-SQL) |
| s4133194 | Khwar Nyo Thinn | - Data Wrangling and cleaning  - Data Visualization |
| s4144857 | Brandon Kok Zheng Tan | Data Visualisation/Analysis (Python, R,) |
| s4057112 | Joseph Khai Zhe Lee | SQL, Machine learning (Python), Data visualisation(Python), JavaScript |

## **Current Progress**

* Held initial team meetings (scheduled recurring time: Tuesdays 9am)
  + Drafted team availability schedule for coordination
* Agreed on potential project directions (e.g., RMIT Connect Chatbot, Healthcare/Mental Health)
  + Started exploring how to get data from RMIT Connect Chatbot
* Conducted a literature review of RAG approaches, focusing on the Walert framework
  + Explored Github and tried reproducing Walert
* Assigned preliminary tasks and shared OneDrive folder access for collaboration

## **Next Steps (3 Week Plan)**

* Finalize project domain selection
* Conduct reproducibility study of Walert RAG pipeline
* Collect and clean preliminary dataset(s) for chosen project
* Begin EDA and summarise findings
* Draft evaluation framework for measuring chatbot performance (effectiveness, coverage, correctness).
  + Tested the existing VAL chatbot with sample queries (e.g., course prerequisites) and confirmed its limitations.

## **Issues or Risks**

* Dataset availability for chosen domain
* Access limitations for scraping RMIT website content
  + RMIT website uses dynamic content (via JavaScript), which makes automated scraping more complex.
* Team scheduling conflicts (different availability)
* Technical complexity of RAG pipeline setup
  + Could be fixed by beginning with small-scale Walert replication before scaling.

## **References**

## **[1] D. Spina. 2024. Walert: Your Open Day FAQ Buddy. Retrieved August 22, 2025 from https://www.damianospina.com/project/walert/**

## **[2] RMIT IR Group. 2024. Walert: Code and data for the Walert large language model-based chatbot. Retrieved August 22, 2025 from https://github.com/rmit-ir/walert**

## **Appendices -** Walert Reproduction

**Progress**

* Enviromental setup:
* Indexing: Built the BM25 index on the FAQ dataset, all passages were handled without error
* Evaluation: Ran eval.py script on both known and inferred topics using qrels.txt and target/runs/. Outputs were stored in target/trec\_eval\_results/ and combined into a full report text file.
* Baseline reproduction: Obtained consistent NDCG scores that matched the expected Walert baselines for intent-based, dense FAISS and BM25 models.

**Appendix A – Reproducibility of Walert**

**Environmental Setup**

- OS: Mac OS

- IDE: Visual Studio Code

- Environment: conda create –n walert python=3.10

- Installed dependencies with pip install –r requirements.txt

**Indexing**

./index-bm25.sh

**Evaluation on Known Topics:**

python src/retrieval/eval.py known data/qrels.txt \

target/runs/rag-bm25.txt \

target/runs/rag-dense-faiss.txt \

target/runs/walert-intent.txt\

>> target/trec\_eval\_results/full\_report.txt

**Evaluation on Inferred Topics:**

python src/retrieval/eval.py inferred data/qrels.txt \

target/runs/walert-intent.txt \

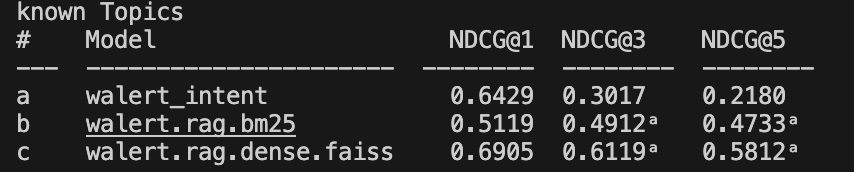
target/runs/rag-bm25.txt \

target/runs/rag-dense-faiss.txt \

> target/trec\_eval\_results/full\_report.txt

**Preliminary Results**

The reproduced evaluation results for known topics:



The reproduced evaluation results for inferred topics:

