**Medical Chatbot with Retrieval-Augmented Generation (RAG)**

**MSDS692 – Data Science Practicum II**  
Weekly Progress Reports  
**Jonish Bishwakarma**

**Week 2 – Proposal Revision and Evaluation Design**

**Planned Work for the Week**

Revise the proposal and strengthen references. Define evaluation framework in more detail.

**Project Timeline**

* Week 1 – Project definition and proposal submission (DONE)
* Week 2 – Proposal revision and evaluation design (DONE)
* Week 3 – Collect medical documents and build knowledge base
* Week 4 – Implement retrieval pipeline and integrate with LLM
* Week 5 – Develop evaluation script and run baseline testing
* Week 6 – Debugging and refinement
* Week 7 – Build chatbot interface
* Week 8 – Final evaluation, analysis, and presentation

**Progress for the Week**

* Added stronger references, including *The Gale Encyclopedia of Medicine* (Guha, 2007) and a recent research paper on RAG faithfulness.
* Clarified evaluation metrics:
  + **Accuracy** – correct answers to in-scope medical questions.
  + **Faithfulness** – whether the answer is grounded in retrieved sources.
  + **Privacy** – system response to out-of-scope questions.
* Created initial design of compute\_scores() function for automatic scoring.

**Roadblocks/Issues**

Balancing realism with simplicity: fully automated evaluation of “faithfulness” is complex, so I had to simplify by linking it to correctness and presence of retrieved evidence.

**Plan for Next Week**

Collect medical resources and prepare the knowledge base for retrieval.

**Resources for the Week**

* Guha, M. (2007). *The Gale Encyclopedia of Medicine (3rd edition).* Detroit: Thomson Gale.