CONCLUSION OF CAPSTONE PROJECT PHASE - 1

In this first phase of the project, we have successfully defined the scope and methodology for building a **Cost-Effective, Proactive Hallucination Routing System**. Our work to date has established the following:

- **Problem Validation:** Through an extensive literature survey, we validated that current hallucination mitigation strategies are largely reactive and computationally expensive, often attempting to fix errors after they occur.
- **Proposed Innovation:** We have designed a solution that improves upon the "self-aware" base paper by making it "context-aware." We introduced a novel **External Reality Check** layer to assess risks such as Context Scarcity, Context Conflict, and Domain Mismatch.
- **Feasibility Confirmation:** We conducted a feasibility study confirming that the system is computationally viable. By utilizing smaller LLMs (360M parameters) and open-source tools, the project can be implemented using standard university hardware.
- **Strategic Roadmap:** We have outlined a clear path for Phase 2, which includes creating a custom labeled dataset from QA benchmarks and training the confidence module to dynamically route queries between Direct Generation, RAG, and Human Review.

Final Summary: This project bridges the gap between model efficiency and reliability. By shifting from post-hoc correction to pre-generation risk assessment, we aim to create a system that allows LLMs to be safely deployed in high-stakes domains like healthcare and finance without the inefficiencies of "always-on" RAG.