**Development Report**

**Report Period:** February 04, 2024 – March 16,2024

**Project Title:** ARA (AI-powered Research Assistant)

**Guide:** Dr. D. J. Chaudhari

**Report Prepared By:**  Kaustubh Warade

Aditya Deshmukh

Devansh Parapalli

Yashasvi Thool

**Executive Summary:** The ARA (AI-powered Research Assistant) project has successfully completed its coding phase from February 4, 2024 to March 16, 2024. Key accomplishments include implementing the frontend using Sveltekit, designing the database schema with Supabase and PostgREST, setting up version control, and integrating state-of-the-art models from Cohere and Mistral for retrieval augmented generation, token generation, and response generation. Challenges included running resource-intensive models on the cloud and creating a custom communication pipeline for serialization/deserialization. Next steps are testing, verification/validation, and performance measurement. Privacy/security, bias mitigation, and knowledge drift were identified as risks requiring robust data handling, model fine-tuning/filtering, and continuous learning mechanisms.

**Current Phase Progress:**

|  |  |
| --- | --- |
| Task Description | Coding phase |
| Scheduled Completion | March 16,2024 |
| Actual Completion | March 16,2024 |
| Status | Completed |
| Remarks |  |

**Accomplishments:**

* Implemented frontend components using Sveltekit.
* Designed the database schema to store user accounts, LLM Runs, citations, sources, and other relevant data using Supabase and PostgREST
* Set up version control using Git and a repository on a platform like GitHub.
* Implementation of Retrieval Augmented Generation, Token Generation and Response Generation were done using SoTA models from Cohere and Mistral.
* A custom communication pipeline was created to ensure efficient serialization and deserialization of requisite data.

**Challenges & Mitigation:**

* Machine Learning models, including the ones used here require powerful computers. To avoid costs associated with infrastructure management, the models were run on the Cloud.
* Models when running on different platforms require a communication pipeline to communicate with the application, such pipelines must deconstruct and reconstruct objects across a text medium. A custom communication pipeline was created, to handle the large volume of serialization and deserialization operations.

**Planned Activities for Next Phase:**

* Testing of all functionalities of the application, including unit testing, integration testing and system testing.
* Verification of all functionalities, content generated, and prompts used for the models. Verification and Validation of the requirements as per Analysis Phase.
* Measure response times for different actions across the entire platform, such as Time to First Token, Page Load Criteria etc.

**Financial Summary:**

|  |  |
| --- | --- |
| Description | Cloud Computing Costs |
| Budgeted Amount | Rs. 5000 |
| Amount Spent This Week | Rs 2049 (USD $25) |
| Total Expenditure | Rs. 2050 |

**Risk Assessment:**

* Handling and processing sensitive data from various sources raise privacy and security concerns. Implementation of robust data anonymization, encryption and access control measures was done.
* The language models used can exhibit biases, leading to unfair, incorrect, or discriminatory outputs. Model fine-tuning, output filtering and grounding are implemented to counteract bias.
* Knowledge used by the models, may become outdated or drift from reality. Implementation of continuous learning, auto source validation, citations and grounding were done to counteract model drift and staleness.

**Conclusions and Recommendations:**

ARA (AI-powered Research Assistant) represents a significant step forward in the development of autonomous research assistants capable of generating comprehensive notebooks of information across various domains. By leveraging SoTA techniques such as Retrieved Augmented Generation (RAG) and large language models (LLMs), ARA aims to provide researchers and knowledge workers with a powerful tool for efficient knowledge acquisition and synthesis.

**Approval:**

Dr. D. J. Chaudhari

Project Guide

Assistant Professor, CSE Department

Government College of Engineering Nagpur

Sector-27, Mihan Rehabilitation Colony

Khapri, Nagpur

441108

Date: April 01, 2024