**Testing Approval Phase**

**Report Period:** March 17, 2024 - March 23, 2024  
**Project Title:** ARA (AI-powered Research Assistant)

**Guide:** Dr. D. J. Chaudhary  
**Report Prepared By:** Kaustubh Warade

Aditya Deshmukh

Devansh Parapalli

Yashasvi Thool

**Executive Summary:** The AI-powered Research Assistant (ARA) project successfully completed the testing phase, accomplishing full test coverage, bug fixes, and verifying large language model outputs against references. Challenges included inconsistent authentication and verifying random model outputs. Next phase plans include deployment with cloud resources, continuous integration/deployment, monitoring, and reinforcement learning for model improvement. Total expenditure remains $2,050. Key risks are ensuring robustness against edge cases, reproducible model outputs, and adequate load testing - with mitigations in place. ARA leverages cutting-edge techniques to provide autonomous research assistance capabilities.

**Current Phase Progress:**

|  |  |
| --- | --- |
| **Task Description** | Testing phase |
| **Scheduled Completion** | 23rd March 2024 |
| **Actual Completion** | 23rd March 2024 |
| **Status** | Completed |
| **Remarks** |  |

**Accomplishments:**

* Completion of Testing Phase.
* Identification of bugs and problems with code, and fixing such bugs and problems
* Ensuring full test coverage for code.

**Challenges and Mitigation:**

* Inconsistent authentication state across multiple devices. Fixed by creating a new policy of one authenticated device per user account.
* LLMs contain randomness, accurately verifying their output is not possible using conventional testing methods. Manual Testing against books and cited sources was used to ensure correctness.

**Planned Activities for Next Phase:**

* Creation of deployment plan, including cloud resources.
* Implement Continuous Integration and Continuous Deployment with deployment policies to ensure fixes and features reach users as fast as possible.
* Implement monitoring mechanisms to track ARA’s performance post-deployment.
* Establish Reinforcement Learning from Human Feedback to ensure continual improvement to the base models.

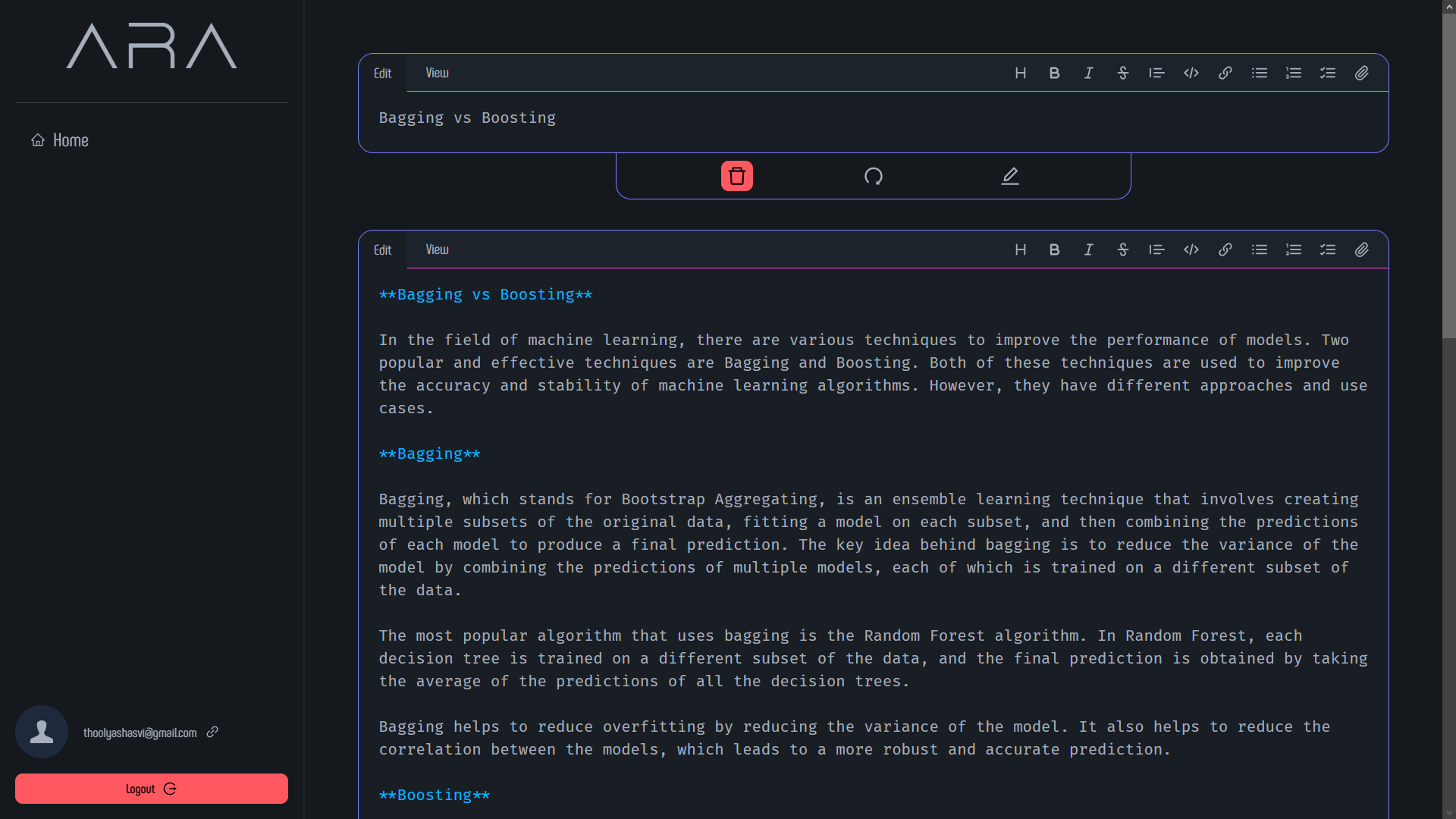
**Financial Summary:**

|  |  |
| --- | --- |
| Description | Testing |
| Budgeted Amount | 0 |
| Amount Spent This Phase | 0 |
| Total Expenditure | 2050 |

**Risk Assessment:**

* Testing the system’s robustness against edge cases, corner cases, and unexpected inputs or scenarios can be difficult, given the vast number of possible inputs and the inherent variability of language models. To mitigate this risk, model output is passed through a filter that checks for potentially illegal, incoherent, unethical and obscene responses.
* Ensuring reproducible and consistent outputs from language models across different runs or scenarios could be a challenge, potentially affecting the trustworthiness of the project. To mitigate this risk,
* Current load testing of ARA’s core systems might not truly indicate problems with scalability in the future. Continuous monitoring and load testing is required.

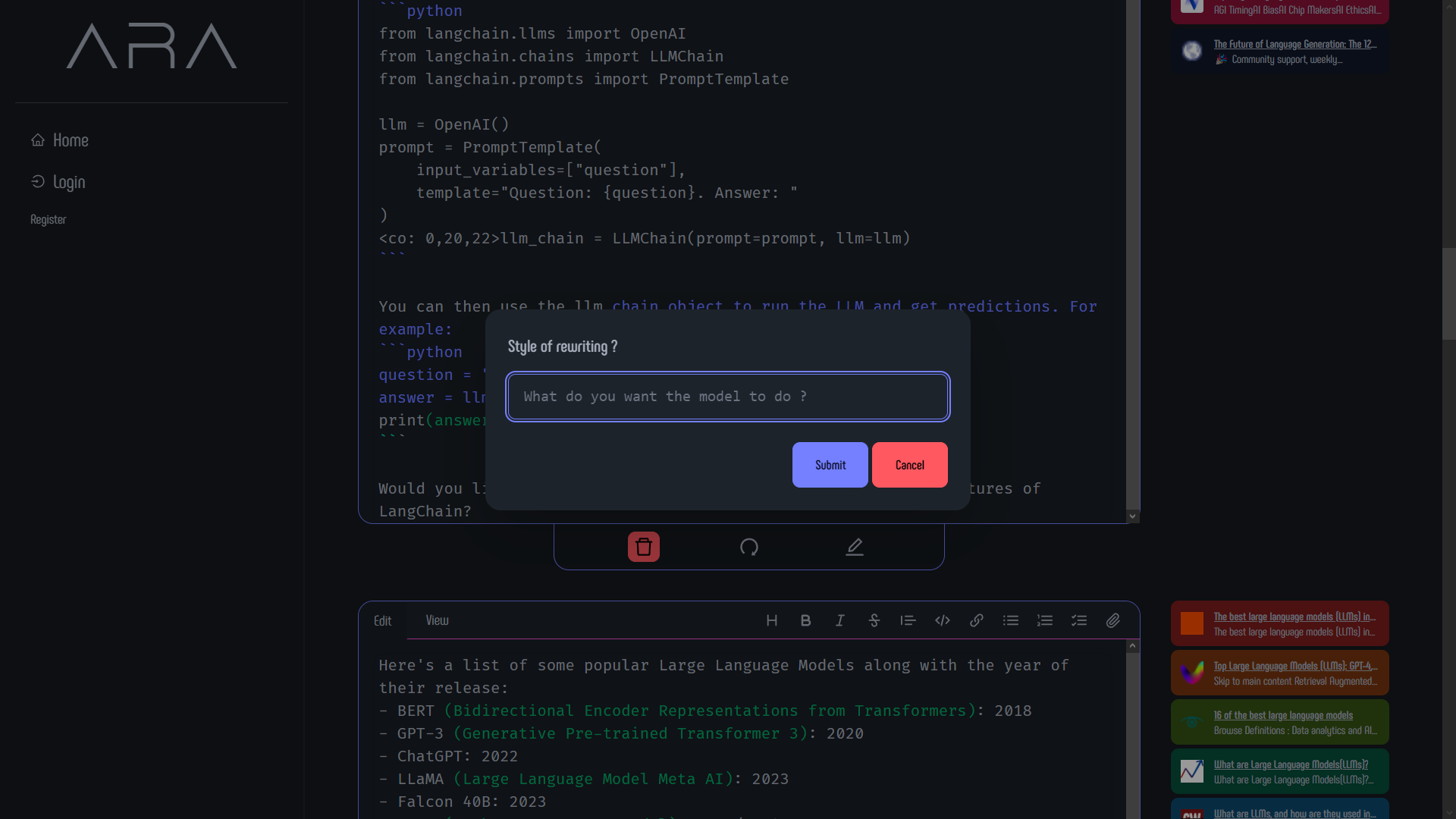
**Image 1:**



**Image 2:**

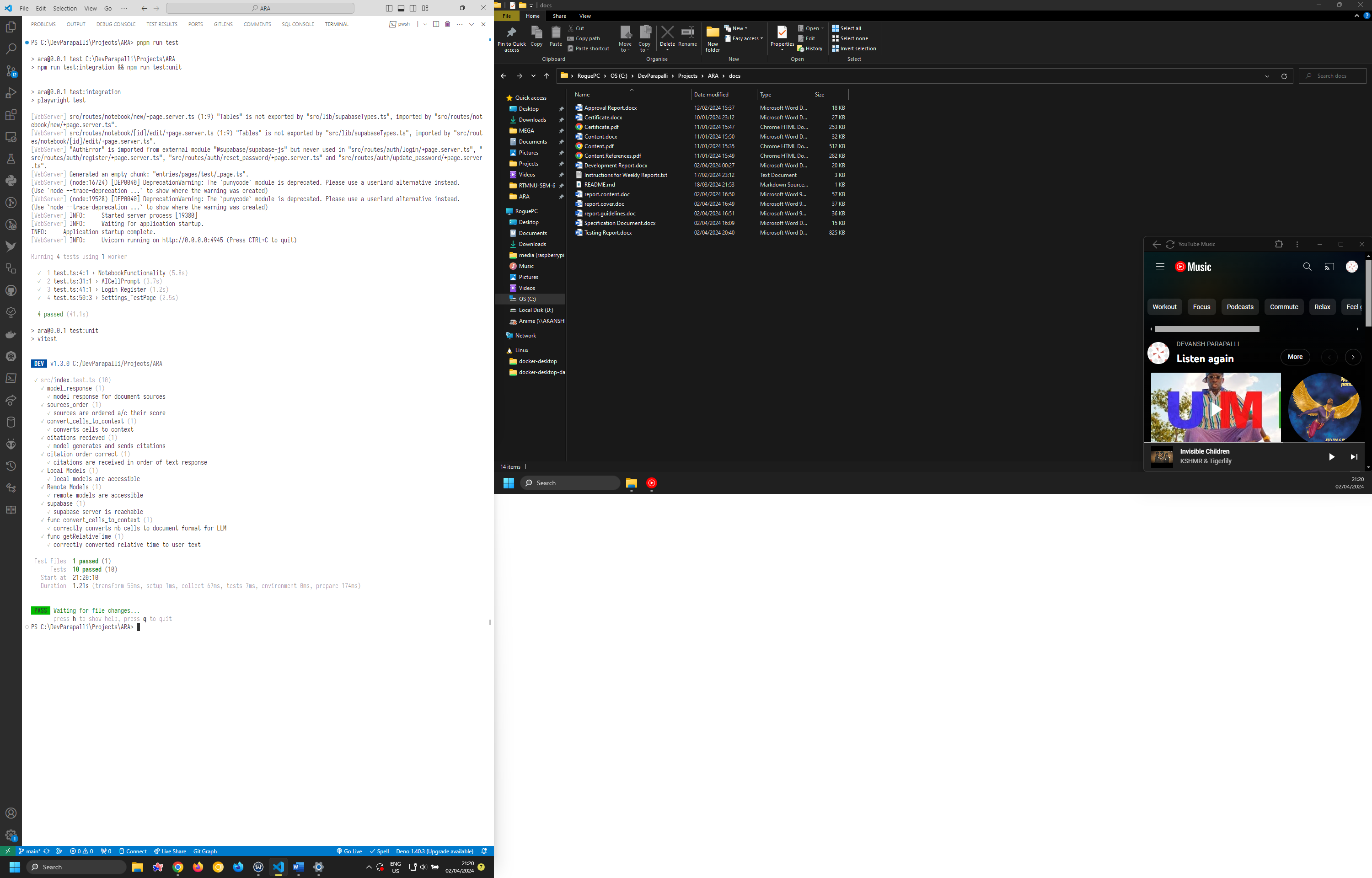


**Image 3:**



**Testing Output:**

|  |  |
| --- | --- |
| A screenshot of a computer  Description automatically generated | A screenshot of a computer  Description automatically generated |

****

**Conclusions and Recommendations:**

ARA represents a significant step forward in developing autonomous research assistants capable of generating comprehensive information across various domains by leveraging state-of-the-art techniques like Retrieved Augmented Generation and large language models. It is recommended to proceed with deploying ARA, implementing the planned continuous integration/deployment pipeline, monitoring mechanisms, and reinforcement learning components. This will allow ARA to provide powerful knowledge acquisition and synthesis capabilities to researchers and knowledge workers, while ensuring the system stays up-to-date, scalable, and continually improves its base models through user feedback. Adequate risk mitigation strategies should be maintained, especially around ensuring robustness, reproducibility, and scalability as real-world usage ramps up.

**Approval:**

Dr. D. J. Choudhary  
Project Guide,

Government College of Engineering,

Khapri, Nagpur,

Maharashtra, India

441108

Date: April 03, 2024