**Assignment**

1. **Introduction**: This assignment is meant to give you (i) a sense of the kind of technical work that you may work on as well as (ii) a way to make your thinking visible to us. We want to ensure that we are as much a fit for you as much as we would want you to be a fit for the organization.
2. **Important things to keep in mind**:
   * Once shortlisted, we would do a deep dive with you trying to understand how you think.
   * Please feel free to use any and all online resources, including any assistance from LLMs. If you use an LLM, we would love to understand how you used it to complete the assignment.
   * Use any python libraries you find fit.
3. You are expected to complete **Parts 1 and 2** of the assignment from below. Bonus points for completing **Part 3**. This is to test your familiarity with the subject matter.
4. **Assignment Part 1:** Building a RAG system
   * RAG systems are one of the most widely used patterns which is powering a lot of AI applications. The basic idea is using an external data source in a vector database along with an LLM. You have to build a RAG system which works on a medium sized dataset. In this case you would be working with NCERT PDF text.
   * Build a RAG system and serve it using a FastAPI endpoint. You should be able to send a query and get a response back.
   * You can use any frontend you see fit to showcase this endpoint.
5. **Assignment Part 2:** Building an Agent that can perform smart actions based on the user’s query. Extend the service and host another endpoint for the agent.
   * The agent should be reliably able to decide when to call the VectorDB and when to not. Ex: Hello -> should not call the VectorDB.
   * Secondly, introduce at least ONE more **action** / **tool** in your system that the Agent can invoke based on the user’s query. Bonus points for more creative actions!
6. **Assignment Part 3:** Give a Voice to your Agent. This is **OPTIONAL** that you can attempt for bonus points!
   * Use [Sarvam’s APIs](https://docs.sarvam.ai/api-reference-docs/introduction) to add voice to your Agent.
7. **Dataset: NCERT PDF Data:** [[link](https://drive.google.com/file/d/17jqGIlgT5yJVfIeM_-4yCMFO5Ojd8H_9/view?usp=sharing)]
8. Please push your code on Github as a **private repository** and share with Github IDs: meenakshi-sarvam and atishay-sarvam and share the repo link on [this form](https://forms.gle/F6AAktwVh6UhftFV7).