

Assignment 1

Due: Sept 27th 03:59 pm

Submission:

1. GitHub Repo Link with
 - a. Diagrams
 - b. A fully documented codelabs
 - c. Video of the submission (5 minutes)
 - d. Link to working application
 - e. GitHub project

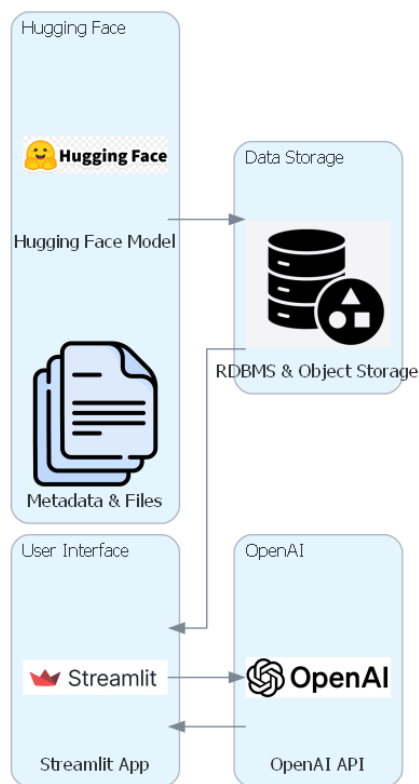
Additional notes:

1. Required attestation and contribution declaration on the GitHub page:
WE ATTEST THAT WE HAVEN'T USED ANY OTHER STUDENTS' WORK IN OUR
ASSIGNMENT AND ABIDE BY THE POLICIES LISTED IN THE STUDENT HANDBOOK
Contribution:
 - a. member1: 33%
 - b. member2: 33%
 - c. member3: 33%
2. Make sure you do not push anything to your GitHub after submission date
3. Create a Code lab document describing everything you did. In your GitHub you should have a readme.md files which would tell what all things are there in this GitHub repository.
4. Keep your repository private until the deadlines. In case of any plagiarism cases both the teams which be equally held responsible

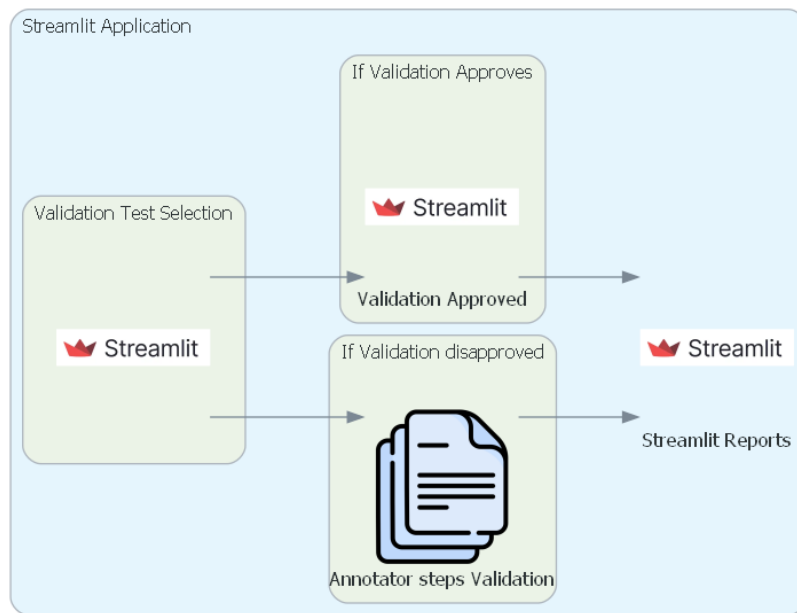
Instructions:

Your team is tasked with building a tool with Streamlit for Model evaluation team to select a validation test case from GAIA dataset [1] and evaluate OpenAI model [2] against the test case.

1. Your tool will be built using streamlit [3]
2. A user would select a specific test case available from the **validation** test file (metadata.jsonl)
3. You should design a system to send the context data and question to OpenAI model to answer it for you.
4. Give an option to user to compare the answer from OpenAI model to the final answer from the metadata file, if the answer from the OpenAI model is incorrect, you should give option to user to modify the Annotator steps present in the Metadata file and evaluate the Model again. (Note: Review the steps and ensure you are not giving away the answer to the model)
5. Record the feedback/response of user on each evaluation which user perform and generate the reports and visualization [4] for the results on streamlit applications.



Assignment 1-A



Assignment 1-B

Notes:

1. Deploy the streamlit app on the Streamlit public cloud
2. Clearly indicate your contributions and time it took to complete your part (individual) in the submission

References:

1. <https://huggingface.co/datasets/gaia-benchmark/GAIA>
2. <https://openai.com/api/>
3. <https://docs.streamlit.io/>
4. <https://blog.streamlit.io/crafting-a-dashboard-app-in-python-using-streamlit/>