Problem Statement

Introduction to GenAI and Simple LLM Inference on CPU and finetuning of LLM Model to create a Custom Chatbot
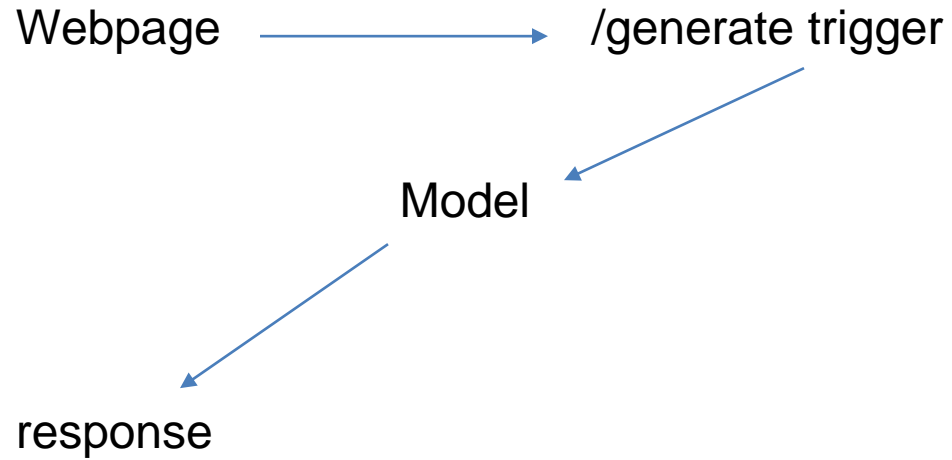
Unique Idea Brief (Solution)

The idea is to use a pretrained model(we chose LlaMa 2 model) for the fine tuning and inference purpose, as it was an open source and easily accessible model. We tune some hyperparameters and run the model locally using flask to deploy the model locally with the UI, after which we can run model in inference mode for finetuning the model.

Features Offered

- Code Generation

- Text generation

- Code correction

- Summarisation

Process flow

Webpage → /generate trigger

Model

response

Architecture Diagram

Technologies used

- HTML

- CSS

- JavaScript

- Flask for deployment

- Python

Team members and contribution:

Afsal Ahamad A – Solo
Contributor

Conclusion

Hence we can experience GenAI through the simple project with various tasks like code generation, and fine tuning of the LLMs.