**CONTENTS**

Chapter 1: Problem statement

Chapter 2: Unique Idea Brief (solution)

Chapter 3: Features offered

Chapter 4: Process flows

Chapter 5: Technologies used

Chapter 6: Team members and contribution

Chapter 7: Conclusion

**Chapter 1: Problem statement**

Introduction to GenAI and simple LLM Inference on CPU and finetuning of LLM model to create a custom chatbot.

**Chapter 2: Unique Idea Brief (solution)**

* Create a chatbot that uses LLM and finetuning the LLM model that understand customer queries and provides personalized solutions.
* Design a chatbot that uses GenAI’s dialogue management capabilities to engage customers in a conversational interface.
* Utilize GenAI’s natural language generation capabilities to generate responses an
* Id knowledge based articles.

**Chapter 3: Features offered**

1. **Conversational GenAI** : Uses natural language processing (NLP) to understand and respond to customer queries.
2. **Conversational Flow** : Manages the conversationl flow to ensure a smooth and natural conversation.
3. **Real-time Analytics** : Provides real-time analytics and insights on customer interactions, sentiments and feedback.
4. **Intent Identification** : Identfies the intent behind customer queries to provide accurate solutions.
5. **Sentimate Analysis** : The chatbot analyzes the customer’s sentiment and emotions to respond appropriately.

**Chapter 4: Process flow**

* First Create our virtual environment and installed the required packages.

**conda create -n itrex python=3.10 -y**

**conda activate itrex**

**pip install intel-extension-for-transformers**

**git clone https://github.com/intel/intel-extension-for-transformers.git**

**cd ./intel-extension-for-transformers/intel\_extension\_for\_transformers/neural\_chat/**

**pip install -r requirements\_cpu.txt**

**pip install -r requirements.txt**

**huggingface-cli login**

**##install jupyter and ipykernel**

**python3 -m pip install jupyter ipykernel**

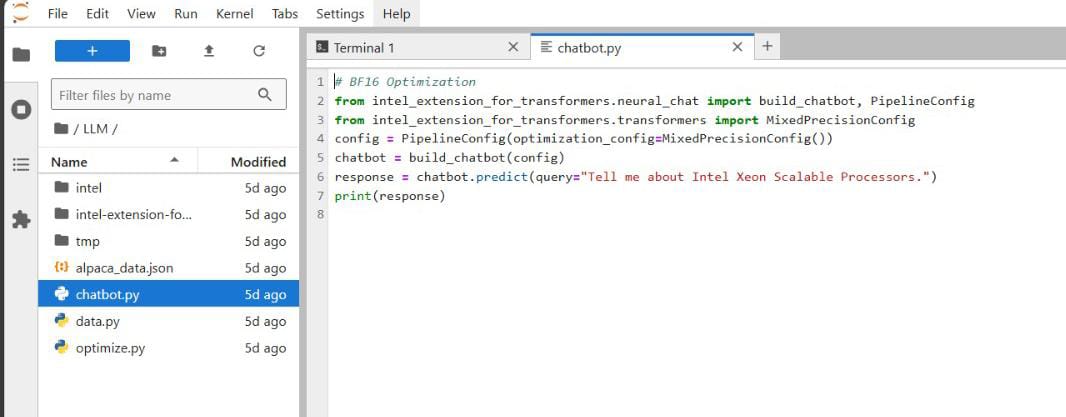
**##Add kernel for its environment**

**python3 -m ipykernel install –nam**

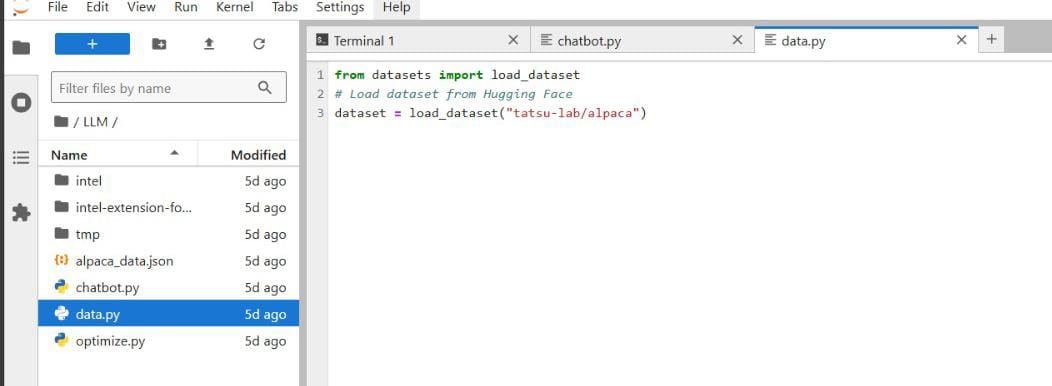
* Required packages:
* Intel Extension for Transformers : It provides a set of optimizations and performance improvements tailored to Intel CPUs and GPUs.
* r requirements\_cpu.txt : The package integrates seamlessly with popular deep learning libraries such as PyTorch ans TensorFlow.
* r requirements.txt : It helps in extracting entities like names, dates, and location from text.
* Hugging Face ID : It refers to the unique identifier used to access and manage user accounts on the Hugging Face Platform.
* Load the dataset. In this model we use alpaca dataset.
* Alpaca dataset : The dataset consists of instruction-response pairs thar covers a wide range of topics and tasks.
* These pairs are designed to help the model learn how to respond accurately and appropriately to different types of user queries.
* It helps in finetuning large language models.

**Chapter 5 : Technology Used**

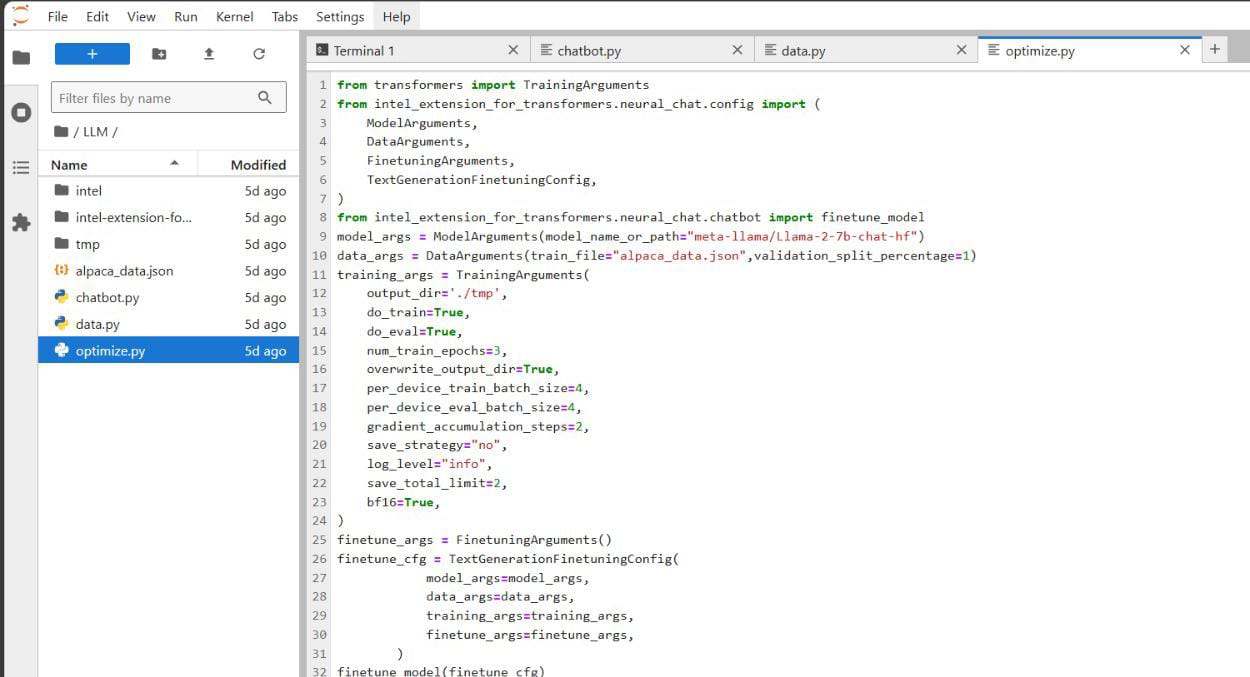
* Intel developer cloud console

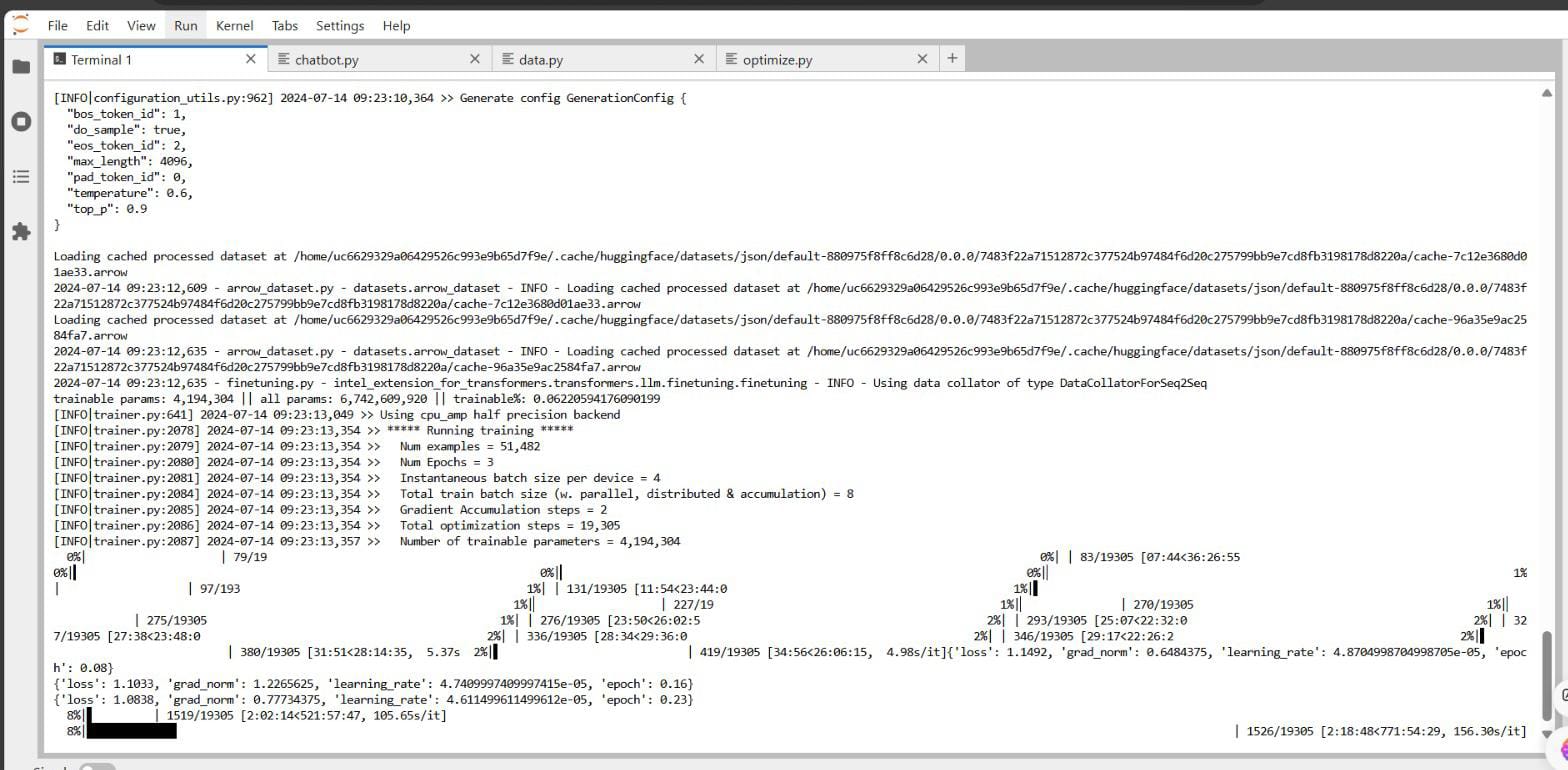


* Import Alpaca dataset



* Finetuning : Finetuning is further trained on a specific dataset to adapt it to particular needs or tasks. For example, a language model trained on general text might be fine-tuned on medical literature to improve its performance on healthcare-related queries.





**Chapter 6: Team members and contribution**

Poorna N. Lutimath : Executing the code and finetuning the model.

Sanjana K.S : Done the report part.

**Chapter 7 : Conclusion**

This project is cover the basis of GenAI, followed by a hands-on demonstration of how to perform simple inference tasks using LLMs on a CPU. We learn how to adapt pre-trained LLMs to their specific use cases, tailoring the model to their desired application or domain. This knowledge and skills can be applied to various applications, such as building conversational interfaces, language translation, or text summarization