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)

A chatbot has been created in-order to help common people to understand the banking sector and the financial services offered by the economical institutions.

A custom dataset was created for training in a .json file in alpaca format
https://github.com/mudxssir/projectintel/tree/main/dataset

For finetuning, A huggingface model has been used:
https://huggingface.co/meta-llama/Llama-2-7b-chat-hf

Features Offered

- Banking and Financial Service expertise and guidance
- Contextual/Non-contextual interaction

Process flow

1. Create a python3 pip environment
2. Activate the Environment
3. Clone GitHub repository:
   https://github.com//intel/intel-extension-for-transformers
4. Change Directory to:
   cd ./intel-extension-for-transformers/intel_extension_for_transformers/neural_chat
5. Install required files as “requirements-cpu.txt” and “requirements.txt”
6. Login to huggingface.co and access “meta-llama/Llama-2-7b-chat-hf” model
7. Finetune with a custom dataset of alpaca format of a .json file
8. Import fine-tuned model to the chatbot interface.

Architecture Diagram

https://github.com/mudxssir/projectintel/tree/main/architecture

Technologies used

- Python pip environment
- Huggingface - https://huggingface.co/meta-llama/Llama-2-7b-chat-hf
- Custom iPykernel named neural_chat_1
- Transformers, requirement files

Team members and contribution:

1) Mudassir Alam N (Team Lead)
2) Alex Kurian
3) Dhaksha D

Conclusion

Basically we learn about Generative AI and Large Language Models and their working on the platform of Intel Developer Core Cloud and the Xeon Scalable GPUs

We ran codes and and obtained outputs for multiple inputs for a pre-trained model and the outputs have been provided in pdf format in the github repository.

The fine-tuning has been performed which took quite a considerable time on a custom dataset created for a domain specific sector to help out regular people to understand the terms and information about the banking and financial services