# REPORT

**Problem Statement:**
Introduction to GenAI and Simple LLM Inference on CPU and Fine-Tuning of LLM Model to Create a Custom Chatbot.

## Table Of Contents




**Name-Bhavyaa Goyal**
**Institute-Manipal Institute of Technology**
**Branch-Data Science**
**College Mentor-Dr. Rashmi Laxmikant Malghan**

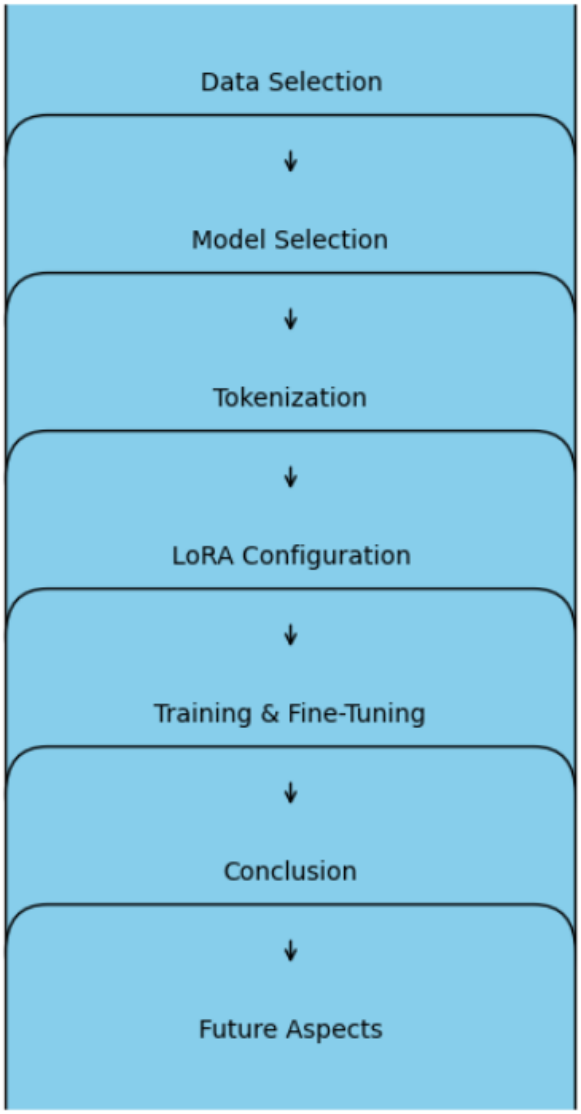
## Objective

- Generate professional email drafts with accurate and contextually appropriate communication for various professional scenarios.
- Ensure improved accuracy and relevance in the generated email drafts.


## Abstract

In this report, we present the process of refining and fine-tuning a Large Language Model (LLM) to generate professional email drafts. The FLAN-T5 model was chosen due to its robust performance in natural language processing tasks and flexibility in fine-tuning for specific applications. Our methodology demonstrates superior results compared to existing models by effectively adapting the model to professional communication contexts. The report details data selection, model selection, tokenization, and fine-tuning processes.

## Methodology




Methodology Workflow

**1) Data Selection**

The dataset comprises approximately 300 emails, meticulously curated from various open-source datasets and AI chatbots. The emails are categorized to cover a range of professional topics, including:

- Vacation Requests
- Requests for Personal Leave
- Annual Leave Requests
- Medical Leave Requests
- Weekly Status Updates
- Project Status Reports
- Milestone Achievement Reports
- Progress Overviews
- Performance Review Meeting Requests
- Performance Evaluation Discussion Requests
- Requests for Salary Increase Reviews
- Requests for Additional Resources
- Resource Allocation Inquiries
- Changes in Resource Allocation

**2) Model Selection**

The FLAN-T5 model was selected due to its strong performance in natural language processing tasks and its adaptability for fine-tuning to specific applications. This choice ensured improved accuracy and relevance in the generated email drafts due to:

- Enhanced understanding of context-specific language.
- Ability to fine-tune effectively for professional communication scenarios.

**3) Tokenization**

Tokenization is a crucial step as it translates text into numerical representations that the model can process, enabling effective learning from the provided data. We initialized the tokenizer using the HuggingFace AutoTokenizer and created tokenizing functions to convert words into vectors, mapping each email into a format suitable for training the model.

**4) LoRA Configuration**

The Low-Rank Adaptation (LoRA) configuration was set up for fine-tuning the model. LoRA efficiently adapts pre-trained language models to specific tasks by adding a small number of trainable parameters, reducing computational requirements and speeding up the training process while maintaining high performance.

**5) Setting Up the Training Arguments and Training/Fine-Tuning the Model**

The training arguments were configured, and the training/fine-tuning process was initiated. The training was completed in **12 minutes for 5 epochs** thanks to the optimized setup and the Intel Developer Cloud CPU. The parameters included settings for learning rate, batch size, number of epochs, weight decay, and strategies for evaluation, saving, and logging. These settings ensured efficient and effective training while optimizing performance and monitoring progress.

## 6)Conclusion

This project successfully demonstrates the fine-tuning of the FLAN-T5 model to generate accurate and contextually relevant professional email drafts. Through this process, I gained valuable insights into data selection, tokenization, and model adaptation using LoRA. The model's performance exceeded expectations, consistently producing high-quality email drafts. This achievement highlights the potential of fine-tuned models in enhancing automated professional communication, setting a foundation for future advancements in this field.

## 7)Future Aspects

a. **Multilingual Support**: Expand the chatbot to support multiple languages, allowing users to generate emails in their preferred language. This can be achieved by fine-tuning the model on multilingual datasets.
b. **Domain-Specific Email Generation**: Train the model on domain-specific datasets to cater to various industries such as healthcare, finance, education, and more. This will allow the chatbot to generate emails with industry-specific terminology and context.
c. **Enhanced Personalization**: Integrate user profiles to personalize email content based on user preferences, past interactions, and specific requirements. This can improve user satisfaction and relevance of the generated emails.

## Reference Links

Source Code- https://github.com/goyalbhavya2/ChatBot
Website Link- https://email-pro-draft.streamlit.app/
GoogleDriveVideo- https://drive.google.com/drive/folders/1JvdTZ3uy8_2Y9f3bbRfmCe6zWNEPHu7g?usp=sharing