

Phase-1 Submission

Student Name: Mohamed Irfan A

Register Number: 712523205040

Institution: PPG Institute Of Technology

Department: Information Technology

Date of Submission: 16-05-2025

1.Problem Statement

Customer support is a critical aspect of any business, but traditional methods often suffer from inefficiencies such as long response times, repetitive queries, and high operational costs. This project aims to address these challenges by developing an intelligent chatbot capable of providing automated, real-time assistance to customers. By leveraging natural language processing (NLP) and machine learning, the chatbot will enhance customer experience, reduce workload on human agents, and lower operational costs. Solving this problem is crucial for businesses seeking to improve customer satisfaction while optimizing resource allocation.

2.Objectives of the Project

The primary objectives of this project are:

- Develop an intelligent chatbot to handle common customer queries and automate repetitive tasks.*
- Analyze customer interactions to improve chatbot performance and provide insights into customer behavior.*

Key outcomes include:

- *A functional chatbot prototype for FAQs, troubleshooting, and user guidance.*
- *Insights into customer pain points and predictions on scaling support through automation.*

3.Scope of the Project

The scope of the project includes:

- *Build and train a chatbot using NLP and machine learning.*
- *Test reliability and scalability in simulated environments.*

Limitations and constraints:

- *Focus on text-based interactions; no voice support initially.*
- *Deployment limited to web apps or platforms like Slack/Messenger.*

4.Data Sources

- *This dataset contains customer support conversations designed for testing LLM-based chatbots. It includes a diverse set of customer-agent interactions covering various queries and scenarios, making it suitable for evaluating chatbot performance in real-world customer service settings.*
 - <https://huggingface.co/datasets/Victorano/Bitext-customer-support-llm-chatbot-testing-dataset-seed42-4k-4.5k>

5.High-Level Methodology

Data Collection

- *Gather data from public repositories or generate synthetic customer-agent interaction data.*

Data Cleaning

- *Clean missing values, duplicates, and standardize text formats; filter out irrelevant or noisy data.*

Exploratory Data Analysis (EDA)

- *Analyze query frequency, types, and visualize trends like peak times and common issues.*

Feature Engineering

- *Tokenize text and extract features using TF-IDF or embeddings; add sentiment or urgency-based features.*

Model Building

- *Experiment with rule-based, ML, and deep learning models; fine-tune pre-trained models like BERT/GPT.*

Model Evaluation

- *Evaluate using metrics like accuracy, F1-score, and BLEU; test usability and user satisfaction.*

Visualization & Interpretation

- *Use dashboards and visualizations (e.g., bar charts, heatmaps) to present key metrics and insights.*

Deployment

- *Deploy as a web app using Streamlit/Flask; optionally integrate with platforms like Slack or Messenger.*

6.Tools and Technologies

Programming Language

- *Python*

Notebook/IDE

- *Google Colab or Jupyter Notebook*

Libraries

- *Data Processing: pandas, numpy*
- *Visualization: matplotlib, seaborn, plotly*
- *NLP & Modeling: NLTK, spaCy, scikit-learn, TensorFlow, Hugging Face Transformers*
- *Deployment: Streamlit, Flask, FastAPI*

Optional Tools for Deployment

- *Dialogflow or Rasa for advanced conversational AI capabilities.*

7.Team Members and Roles

Name	Role	Responsibilities
Ram Kishore N	<i>Project Manager</i>	<i>Oversee project progress, coordinate tasks, and ensure timely delivery.</i>
Harsen K	<i>Data Scientist/NLP Scientist</i>	<i>Perform EDA, feature engineering, text preprocessing, intent recognition, and initial model building.</i>
Anand V	<i>Developer</i>	<i>Handle deployment, API integrations, and front-end development for the web app.</i>
Mohamed Irfan A	<i>Quality Assurance Tester</i>	<i>Test the chatbot for bugs, usability issues, and performance bottlenecks.</i>