

Project Report

Project Idea: Intelligent Customer Support Chatbot for Real-Time Issue Resolution

By Akhil Juvvanapudi

1. Project Overview

This project deals with the development of an Intelligent **Customer Support Chatbot** which is capable of understanding customer queries, retrieving relevant information from a knowledge base, and providing accurate real-time responses.

2. Problem Statement

The objective of this project is to develop an AI-powered chatbot that:

- Understands customer queries
- Retrieves accurate and context-aware answers from historical customer service data
- Responds in real time
- Escalates issues that require human intervention

3. Objectives

- Fine-tune a Large Language Model (LLM) on customer support data
- Implement Retrieval-Augmented Generation (RAG) to retrieve information from knowledge bases for factual accuracy
- Design intent-aware query handling
- Build an interactive graphical user interface for real-time customer interaction
- Evaluate the chatbot's performance using validation metrics and real-world queries

4. Dataset Description

The dataset used for this project is **Bitext Customer Support LLM Chatbot**

Training Dataset. Source: Hugging Face

<https://huggingface.co/datasets/bitext/Bitext-customer-support-lm-chatbot-training-dataset>

Dataset Characteristics:

- Contains real-world customer support queries and responses
- Covers multiple sections such as order tracking, refunds, account management, subscriptions, and complaints
- Includes intent labels that help in intent-aware retrieval

This dataset is good enough for fine-tuning sequence-to-sequence models for customer support agent use cases.

5. System Architecture

The system consists of the following major components:

5.1 Large Language Model (LLM)

- **Model used:** FLAN-T5 Base
- Fine-tuned on customer support interactions
- Generates natural language responses

5.2 Retrieval-Augmented Generation (RAG)

- Knowledge base created using customer support responses
- Sentence embeddings generated using `all-MiniLM-L6-v2`
- FAISS used for efficient similarity search
- A cross-encoder reranker (`ms-marco-MiniLM-L-6-v2`) to improve retrieval accuracy

5.3 Intent-Aware Retrieval

- Rule-based intent detection using keywords
- Intent-specific document filtering before retrieval
- Ensures relevant policies are prioritized

5.4 Escalation Agent

- Automatically escalates sensitive issues (e.g., legal threats)
- Provides human-agent handoff messaging

5.5 User Interface

- Built using Gradio
- Chat-based interface for real-time interaction
- Lightweight and easy to deploy

6. Methodology

Step 1: Data Collection

- Historical customer service interactions extracted from the dataset
- Responses indexed as knowledge base documents

Step 2: Model Fine-Tuning

- FLAN-T5 fine-tuned using supervised learning
- Input: customer instructions
- Output: agent responses
- Padding tokens masked to avoid loss distortion

Step 3: Knowledge Base Construction

- Sentence embeddings created for policy documents
- FAISS vector database used for similarity search

Step 4: Query Handling Pipeline

1. User query received
2. Escalation rules checked
3. Intent detection performed
4. Relevant documents retrieved
5. Reranking applied
6. Final response generated using LLM + retrieved context

Step 5: UI Integration

- Gradio Blocks-based UI
- Single-line input box for compact chat experience
- Real-time response rendering

7. Model Training and Evaluation

Training Configuration:

- Epochs: 3
- Batch size: 8
- Gradient accumulation: 2
- Precision: bfloat16
- Optimizer: AdamW (Default optimizer used by Trainer)
- 3000 samples were used in which 2700 samples used for training and remaining 300 samples for validation (test_size=0.1)

Training Results:

Epoch	Training Loss	Validation Loss
1	0.791900	0.668034
2	0.679100	0.614224
3	0.626500	0.597773

Observations:

- Validation losses consistently decreased across consecutive epochs
- Indicates effective learning and decent generalization
- No signs of overfitting observed

8. Manual Evaluation

Query	Expected	Chatbot	Escalation	Remarks
	Intent	Response		
	Quality			
Where is my order?	Track Order	Correct & Relevant	No	Tracking response
Cancel my order	Cancel Order	Correct	No	Clear steps provided
Renew my subscription	Subscription	Correct	No	Redirected to support
I want to sue you	Complaint	Correct	Yes	Escalated properly

Intelligent Customer Support Chatbot

Hi, I'm AI Assistant. How can I help you?

Chatbot

Where is my order?

We apologize for the inconvenience you are experiencing while trying to correct your order. To assist you further, could you please provide us with the order number? This will allow us to better understand your situation and find a suitable solution for you.

Replace my order

We appreciate your inquiry about changing your order. To assist you with this, please provide us with the order number and the specific changes you would like to make. This will help us process your request promptly and ensure your order is exactly as you desire.

Type a message... Submit

Retry Undo Clear

Examples

Where is my order? Cancel my order Replace my order Renew my subscription Change delivery Address I want to esclate this issue

Use via API · Built with Gradio

Intelligent Customer Support Chatbot

Hi, I'm AI Assistant. How can I help you?

Chatbot

Cancel my order

I'm on it! I understand that you're seeking information about canceling an order. To cancel your order, follow these steps: 1. Sign in to Your Account: Access your account on our website by logging into your account on our website. 2. Locate Your Order: Look for the order you wish to cancel and click on it. 3. Initiate Cancellation: Once you've logged in, navigate to the "Order History" or "My Orders" section. 4. Click on the order to view the history. 5. Verify the Cancell

Renew my subscription

At the moment, I can help with newsletter subscriptions. For renewing paid subscriptions, please contact our associate.

Type a message...

Submit

Retry

Undo

Clear

Examples

Where is my order?

Cancel my order

Replace my order

Renew my subscription

Change delivery Address

I want to esclate this issue

Use via API  · Built with Gradio 

Intelligent Customer Support Chatbot

Hi, I'm AI Assistant. How can I help you?

Chatbot

Change delivery Address

We appreciate your inquiry about changing your delivery address. To assist you with this, please log in to your account on our website and navigate to the "My Account" section. From there, you will be able to find an option to update your shipping address. If you need further assistance, feel free to reach out to our customer support team for further guidance.

I want to esclate this issue

Sorry for the inconvenience. Our Customer Support Executive will be contacting you shortly.

Type a message...

Submit

Retry

Undo

Clear

Examples

Where is my order?

Cancel my order

Replace my order

Renew my subscription

Change delivery Address

I want to esclate this issue

Use via API  · Built with Gradio 

9. Results and Discussion

The chatbot successfully:

- Answered common customer queries such as order tracking, cancellation, refunds, and address changes
- Maintained factual accuracy using retrieval-based responses
- Escalated legal and sensitive queries to human support
- Provided consistent and context-aware responses
- Subscription renewal queries are handled using a **predefined business rule** that redirects users to human support due to the absence of such cases in the training data.

The combination of fine-tuning and RAG significantly improved response quality compared to a base LLM.

10. Limitations

- Intent detection is rule-based and may miss ambiguous queries
- The chatbot relies on the coverage of the existing knowledge base
- Real-time learning or feedback incorporation is not implemented

11. Future Enhancements

- Replace rule-based intent detection with a trained intent classifier
- Add multilingual support
- Integrate CRM or ticketing systems
- Implement continuous learning from user feedback
- Train model on a larger set of data

12. Conclusion

This project focuses on building an intelligent customer support chatbot by applying practical natural language processing techniques. The system was developed by fine-tuning a large language model and integrating it with a retrieval-augmented framework to provide relevant and timely responses. Intent-aware logic and a cross-encoder-based reranking approach were used to improve the accuracy of retrieved information and ensure that the most suitable knowledge base content was selected. As a result, the chatbot is able to handle customer queries efficiently, minimize manual support effort, and appropriately escalate complex or sensitive issues, making it suitable for real-world customer service environments.