A Query Resolving Chatbot Using Specialized RNN Approach

1.PROBLEM STATEMENT

One of the major challenges in customer support is efficiently resolving queries. Existing chatbots are often limited to specific domains, leading to inefficiencies in handling diverse queries. This project aims to develop a universal chatbot that can resolve queries across multiple domains quickly and efficiently.

2. SETTING UP THE ENVIRONMENT

To begin, you'll need to set up your development environment. We will use Google Colab for coding, which provides free access to GPUs and an interactive coding environment.

Python Libraries: Install necessary Python libraries like TensorFlow, Keras, Numpy, Pandas, Matplotlib, and NLP tools like NLTK or SpaCy

```
import tensorflow as tf
import numpy as np
import pandas as pd
import json
import nltk
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.layers import Input, Embedding, LSTM, Dense, GlobalMaxPooling1D, Flatten
from tensorflow.keras.models import Model
import matplotlib.pyplot as plt
import string
import seaborn as sn
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import classification_report
```

3.DATA COLLECTION

Gather a dataset with Frequently Asked Questions (FAQs) and their corresponding answers from various domains. Ensure the dataset is in JSON format and properly labeled with questions and answers.

AS PER DATASET PROVIDED:

```
data1 = {
  "intents": [
    {
      "tag": "greetings",
```

```
    "input": ["hello", "hi there", "nice to meet you", "hi", "hey there", "hi there"],
    "responses": [
      "hi, welcome",
      "this is spark, how can I help you",
      "hello, how may I help you",
      "hello, my friend, how can I help you"
    ]
  },
  {
    "tag": "whoareyou",
    "input": ["who are you?", "what is your name?", "are you a chatbot?", "what can I call you?"],
    "responses": [
      "I am a virtual assistant",
      "my name is spark, I am a virtual assistant",
      "you can call me spark"
    ]
  },
  {
    "tag": "ai",
    "input": ["what is artificial intelligence?", "what is ai?", "artificial intelligence?", "ai"],
    "responses": [
      "The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision making, and translation between languages.",
      "Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems.",
```

"AI is the ability of a machine to display human-like capabilities such as reasoning, learning, planning, and creativity. AI enables technical systems to perceive their environment, deal with what they perceive, solve problems, and act to achieve a specific goal."

]

},

{

"tag": "ml",

"input": ["what is machine learning?", "what is ml?", "tell me about machine learning?", "ml"],

"responses": [

"Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.",

"Machine learning is a subfield of artificial intelligence, which is broadly defined as the capability of a machine to imitate intelligent human behavior. Artificial intelligence systems are used to perform complex tasks in a way that is similar to how humans solve problems.",

"Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values."

]

},

{

"tag": "computer vision",

"input": ["what is computer vision?", "tell me about computer vision?", "computer vision"],

"responses": [

"A field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos, and other visual inputs.",

"Computer vision is a field of artificial intelligence that trains computers to interpret and understand the visual world. Using digital images from cameras and videos and deep learning models, machines can accurately identify and classify objects — and then react to what they see.",

"Computer vision is a field of AI that trains computers to capture and interpret information from image and video data. By applying machine learning (ML) models to images, computers can classify objects and respond—like unlocking your smartphone when it recognizes your face."

]

}

]

}

4.DATA PROCESSING

The preprocessing step involves cleaning and preparing the text data for model training.

Steps involved in preprocessing:

1. Text Cleaning: Remove unnecessary symbols, punctuations, and numbers.

2. Tokenization: Break down sentences into words or tokens.

3. Lemmatization: Convert words to their base forms.

4. Vectorization: Convert the text data into numerical format using word embeddings.

```
# Pre-processing
# Removing punctuations
data['inputs'] = data['inputs'].apply(lambda wrd: [ltrs.lower() for ltrs in wrd if ltrs not in string.punctuation])
data['inputs'] = data['inputs'].apply(lambda wrd: ''.join(wrd))

# Tokenize the data
tokenizer = Tokenizer(num_words=200000)
tokenizer.fit_on_texts(data['inputs'])
train = tokenizer.texts_to_sequences(data['inputs'])
```

5.MODEL DESIGN AND TRAINING

The core of the chatbot is built using an LSTM (Long Short-Term Memory) network, which is a type of RNN suitable for handling sequences and long-term dependencies.

Steps:

1. Build an RNN architecture using LSTM layers.

2. Train the model on the preprocessed dataset.

3. Use categorical cross-entropy as the loss function and Adam optimizer.

4. Evaluate the model using accuracy metrics on the test set.

6. MODEL EVALUTION

Model Summary: Check the structure and parameters of the trained model.

Accuracy and Loss: Plot and analyze training accuracy and loss.

	precision	recall	f1-score	support
ai	0.938462	0.910448	0.924242	67.000000
aws	0.986301	1.000000	0.993103	72.000000
cpu	0.941176	0.984615	0.962406	65.000000
dbms	0.942857	0.956522	0.949640	69.000000
java	0.710000	0.986111	0.825581	72.000000
ml	0.968750	0.516667	0.673913	60.000000
nlp	1.000000	0.975000	0.987342	80.000000
python	1.000000	0.666667	0.800000	3.000000
sql	1.000000	1.000000	1.000000	72.000000
accuracy	0.923214	0.923214	0.923214	0.923214
macro avg	0.943061	0.888448	0.901803	560.000000
weighted avg	0.936374	0.923214	0.919238	560.000000

Confusion Matrix: Assess model performance using a confusion matrix.

ai	61	0	1	3	0	0	0	0	0
aws	0	72	0	0	0	0	0	1	0
cpu	2	0	64	0	0	0	2	0	0
dbms	4	0	0	66	0	0	0	0	0
java	0	0	0	0	71	29	0	0	0
ml	0	0	0	0	1	31	0	0	0
nlp	0	0	0	0	0	0	78	0	0
python	0	0	0	0	0	0	0	2	0
sql	0	0	0	0	0	0	0	0	72

7.IMPLEMENTING CHATBOT IN PYTHON

Implement the chatbot's user interface and integrate the trained model to provide real-time query resolution. The chatbot should preprocess input queries, pass them through the model, and fetch responses from the knowledge base.

```
 "tag":"skin cancer",
 "input":["what is skin
 cancer","cancer"], "responses":[
 "Skin cancer is the out-of
control growth of abnormal cells in the epidermis, the outermost
 skin layer, caused by unrepaired DNA damage that triggers.",
 "Skin cancer that forms in melanocytes (skin cells that make
 pigment) is called melanoma. Skin cancer that forms in the lower
 part
 of the epidermis (the outer layer of the skin) is called basal
 cell carcinoma."]
 },
 {
 "tag":"malaria",
 "input":["what is
 malaria","malaria"], "responses":[
 "Malaria is a disease caused by a parasite. The parasite is
 s pread to humans through the bites of infected mosquitoes."]
 },
 {
 "tag":"virus",
 "input":["what is
 virus","virus"], "responses":[
 "A virus is an infectious microbe consisting of a segment
 of nucleic acid (either DNA or RNA) surrounded by a protein coat",
 "A virus is a submicroscopic infectious agent that
 replicates only inside the living cells of an organism."]
 },
 {
 "tag":"population",
 "input":["population of india","india
 population"], "responses":["140.76 crores"]
 },
 {
 "tag":"largest country",
 "input":["which is the largest country in the world","largest c
 ountry in the world"],
 "responses":["russia"
```

```
 "tag":"oops",
 "input":["what is
 oops","oops"], "responses":[
 "Object
oriented programming is based on the concept of objects In object
oriented programming data structures, or objects are defined, each
 wi th its own properties or attributes"
 ]
 },
{
 "tag":"pillars of oops",
 "input":["four pillars of oops","pillars of
 oops"], "responses":[
 "Abstraction ,Inheritance ,Polymorphism ,Encapsulation"
 ]
 },
 {
 "tag":"cancer",
 "input":["what is a
 cancer","cancer"], "responses":[
 "Cancer is a disease caused when cells divide uncontrollably
 and spread into surrounding tissues"
 ]
 },
 {
 "tag":"causes of cancer",
 "input":["main causes of cancer","cause of
 cancer"], "responses":[
 "Leading risk factors for preventable cancers are smoking, ge
 tting too much ultraviolet (UV) radiation from the sun or tanning
 bed s, being overweight or having obesity, and drinking too much
 alcohol. "
 ]
 },
 {
 "tag":"cardiology",
 "input":["what is cardiology","cardiology","What is the role of
 cardiology"],
 "responses":[
 "Cardiology is a medical specialty and a branch of internal
 m edicine concerned with disorders of the heart.",
 "Cardiology is the study of the heart. In medicine,
 cardiolog y is a branch of medicine that deals with disorders of the
 heart and the cardiovascular system.",
```

8.DEPLOYMENT AND OUTPUT

Deploy the chatbot on a cloud service such as AWS, Google Cloud, or Heroku. Create an API using Flask or Django to serve the chatbot responses.

```
Input: What is skin cancer?
Output: Skin cancer that forms in melanocytes (skin cells that make pigment) is called melanoma. Skin cancer that forms in the lower part of the epidermis (the outer layer of the skin)
is called basal cell carcinoma.

Input: What is malaria?
Output: Malaria is a disease caused by a parasite. The parasite is spread to humans through the bites of infected mosquitoes.

Input: What is a virus?
Output: A virus is an infectious microbe consisting of a segment of nucleic acid (either DNA or RNA) surrounded by a protein coat.

Input: What is the population of India?
Output: 140.76 crores.

Input: Which is the largest country in the world?
Output: Russia.
```

Testing:

Service: Use your EC2 instance's public IP or domain name to access and test your application directly in a web browser or through API tools like Postman.

Monitoring:

Service: AWS CloudWatch

- Set up CloudWatch alarms and logs to monitor application performance and resource utilization.

Scaling:

Service: AWS Auto Scaling

- Configure Auto Scaling groups to automatically adjust the number of EC2 instances based on demand.

Running:

Service: EC2 Instance

- Manage and keep your application running using the EC2 instance.

Configuring Web Server:

Service: Nginx or Apache

- Configure Nginx or Apache to serve your application and handle web traffic.

Setting Up Domain and SSL:

Domain: AWS Route 53

- Configure domain name settings and DNS records.

SSL: AWS Certificate Manager

- Obtain and manage SSL certificates for secure HTTPS connections.