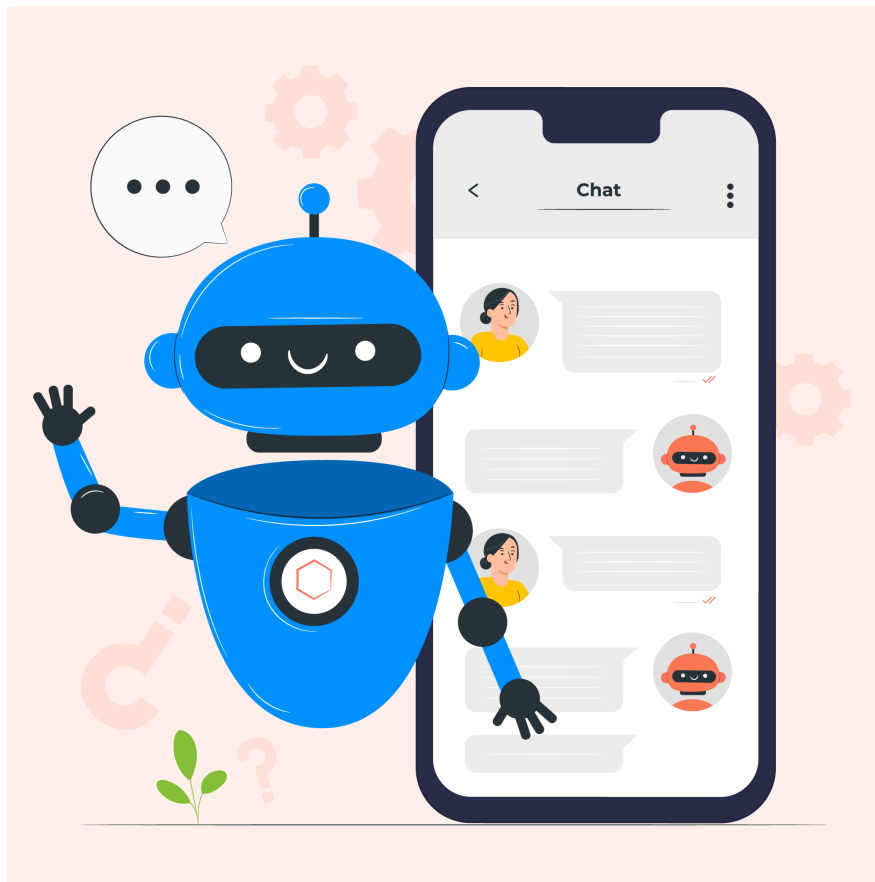


# Implementation of Chatbot using NLP



## Problem Statement

Develop a chatbot capable of understanding and responding to user inputs by identifying intents and extracting entities. The chatbot should provide relevant responses based on trained intents, assisting in creating seamless user interactions.

## 1. Aim

The objective of this project is to develop an intents-based chatbot leveraging Natural Language Processing (NLP) techniques and a Logistic Regression model. The chatbot will be able to process user input, classify it into defined intents, and respond with relevant information. This project aims to provide a conversational interface for users, improving the interactivity and efficiency of communication.

## 2. Project Requirements

### 2.1 Hardware Requirements:

- **Processor:** Intel Core i3 or equivalent (minimum requirement)
- **RAM:** 4GB (Recommended: 8GB or more)
- **Storage:** 500MB of free space (Recommended: 1GB for efficient processing)
- **Operating System:** Windows, macOS, or Linux

### 2.2 Software Requirements:

- a. **Python Version:** 3.6 or higher version
- b. **Required Libraries:**
  - **nltk** - Natural Language Toolkit for text processing tasks such as tokenization and lemmatization
  - **streamlit** - A framework for developing the user interface of the chatbot
  - **scikit-learn** - For machine learning model training, specifically Logistic Regression and TF-IDF vectorization
  - **numpy** - For numerical operations required during vectorization and model training
  - **scipy** - For scientific computing and optimizations
- c. **Development Environment:** Python IDE such as Visual Studio Code, Jupyter Notebook, or PyCharm.

## 3. Knowledge of Data Modelling

Data modelling in this project focuses on transforming **user inputs** (in text form) into a **structured format** that can be processed by a machine learning model.

- **Data Preprocessing:** Raw user inputs (patterns) are tokenized into words, lemmatized, and vectorized using **TF-IDF** (Term Frequency-Inverse Document Frequency) to capture the importance of words in the context of the dataset. This transformation ensures the model can interpret the textual data effectively.
- **Modeling:** The core machine learning model used is **Logistic Regression**, a classification algorithm that is trained on the preprocessed dataset. The model classifies each user input into

one of several predefined intent categories (e.g., greeting, query, farewell). The classifier is trained using a labeled dataset containing various user queries and their associated intents.

- **Model Evaluation:** After training the Logistic Regression model, it is evaluated on a test set to assess its performance and accuracy in predicting the correct intent. Performance metrics such as accuracy and F1-score can be used to measure the model's effectiveness.

## About the Project

This project aims to develop a chatbot that can identify user intents and generate appropriate responses. Using **Natural Language Processing (NLP)** techniques, such as tokenization and lemmatization, along with **Logistic Regression** for intent classification, the chatbot processes user input and provides relevant information. The chatbot is deployed with an interactive user interface built using **Streamlit**, allowing users to engage in seamless conversations with the system.

The scope of this project can be expanded with more sophisticated NLP techniques, larger datasets, and improved response generation methods. This project lays the foundation for building more advanced conversational agents and interactive systems.

## Week 1 Task Source Code

Link:[https://colab.research.google.com/drive/1GDD\\_KznF99CPdFdg\\_mtzxubc7JgtLo4M:](https://colab.research.google.com/drive/1GDD_KznF99CPdFdg_mtzxubc7JgtLo4M)