Phase2

INNOVATION

Project Name : CREATE A CHATBOT IN PYTHON

Team ID : 8932

Introduction:

In this phase, we outlined steps involved , Natural language Processing(NLP) techniques and Architecture for creating ChatBot using Python.

Steps involved in designing:

Tokenization: The input text is broken down into smaller units called tokens. Tokens can be words or even subwords, depending on the language and model used.

Text Preprocessing: The text is preprocessed to remove noise, like punctuation, capitalization, and stopwords (common words like "the," "and," "in" that don't carry much meaning).

Extraction: The chatbot extracts features from the tokens to represent the input text. These features can include word embeddings, which map words to numerical vectors, making it easier for the model to understand the text's meaning.

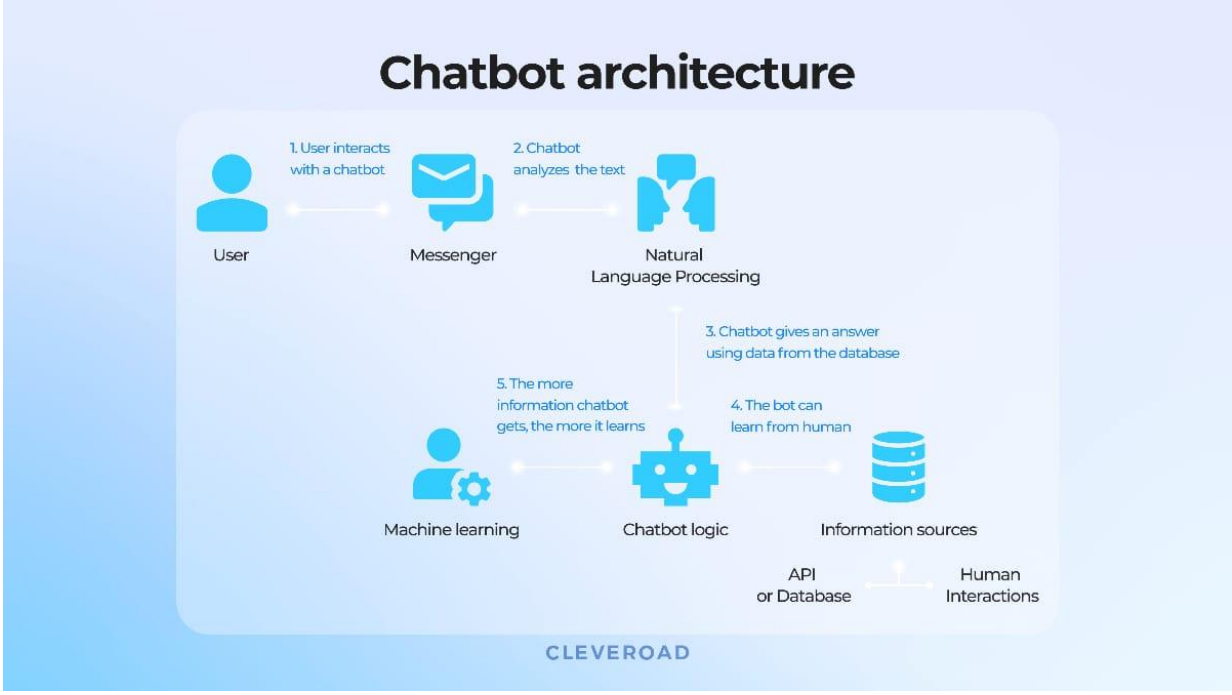
NLP Model: The chatbot uses a machine learning or deep learning model, often based on neural networks, to analyze the input. Common models include recurrent neural networks (RNNs), convolutional neural networks (CNNs), or transformer-based models like GPT (Generative Pre-trained Transformer).

Understanding: The model processes the input text and tries to understand the user's intent. This involves recognizing entities (like names, places, or dates) and determining the context.

Matching: The chatbot compares the user's input to predefined patterns or rules to identify the most appropriate response. This might involve searching a database of responses or using predefined conversational templates.Generating

Response: Based on the understanding and matching results, the chatbot generates a response. This can be done using template-based responses, rule-based responses, or by generating text using the model itself.

Architecture:



Innovation techniques:

Various Natural Language Processing Algorithms are used for reading and replying in ChatBot.

Named Entity Recognition (NER): Identifies and classifies entities like names, dates, and locations in text.

Text Classification: Categorizes text into predefined classes (e.g., sentiment analysis, spam detection).

Text Clustering: Groups similar documents or sentences together based on their content.

Information Retrieval: Retrieves relevant documents from a large corpus in response to a query.

Machine Translation: Translates text from one language to another.

Word Embeddings: Represent words in a continuous vector space, e.g., Word2Vec, GloVe.

Seq2Seq Models: Used in tasks like machine translation, chatbots, and text summarization.

Keyword Extraction: By definition, keyword extraction is the automated process of extracting the most relevant information from text using AI and machine learning algorithms.

Tools Used:

1.SpaCy

2.ChatterBot

3.NLTK

4.PyNLPL

Conclusion:

The Chatbot is developed furtherly innovated in more steps and ways that are been developed with the Questions and Answers.