

CREATE A CHATBOT USING PYTHON

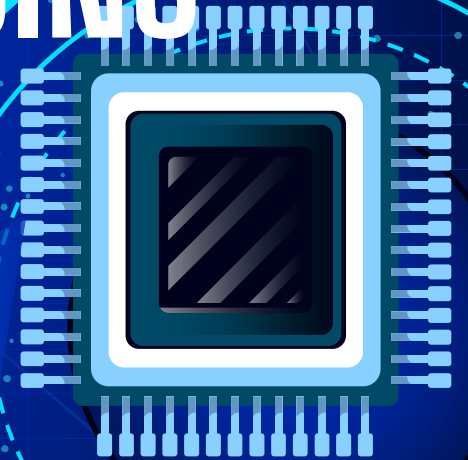


PRESENTED BY,
SIVASUBRAMANIAN TJ

01. INTRODUCTION

02. DATA PREPROCESSING

03. CONCLUSION



INTRODUCTION

- In this phase we will cover data preprocessing for chatbot creation using python.
- Data preprocessing is the fundamental stage in which the raw, often noisy data is transformed into a structured and refined format, making it suitable for the chatbot's understanding and response generation.

DATA PREPROCESSING

```
In [2]: import pandas as pd

# Read the dataset into a DataFrame
df = pd.read_csv('dialogs.txt', sep='\t', names=['user_message', 'bot_response'])

# Remove any rows with missing values
df.dropna(inplace=True)

# Remove leading and trailing whitespaces in the text data
df['user_message'] = df['user_message'].str.strip()
df['bot_response'] = df['bot_response'].str.strip()

# Convert text to lowercase for consistency
df['user_message'] = df['user_message'].str.lower()
df['bot_response'] = df['bot_response'].str.lower()

# Optionally, you can remove duplicate dialog pairs
df.drop_duplicates(inplace=True)

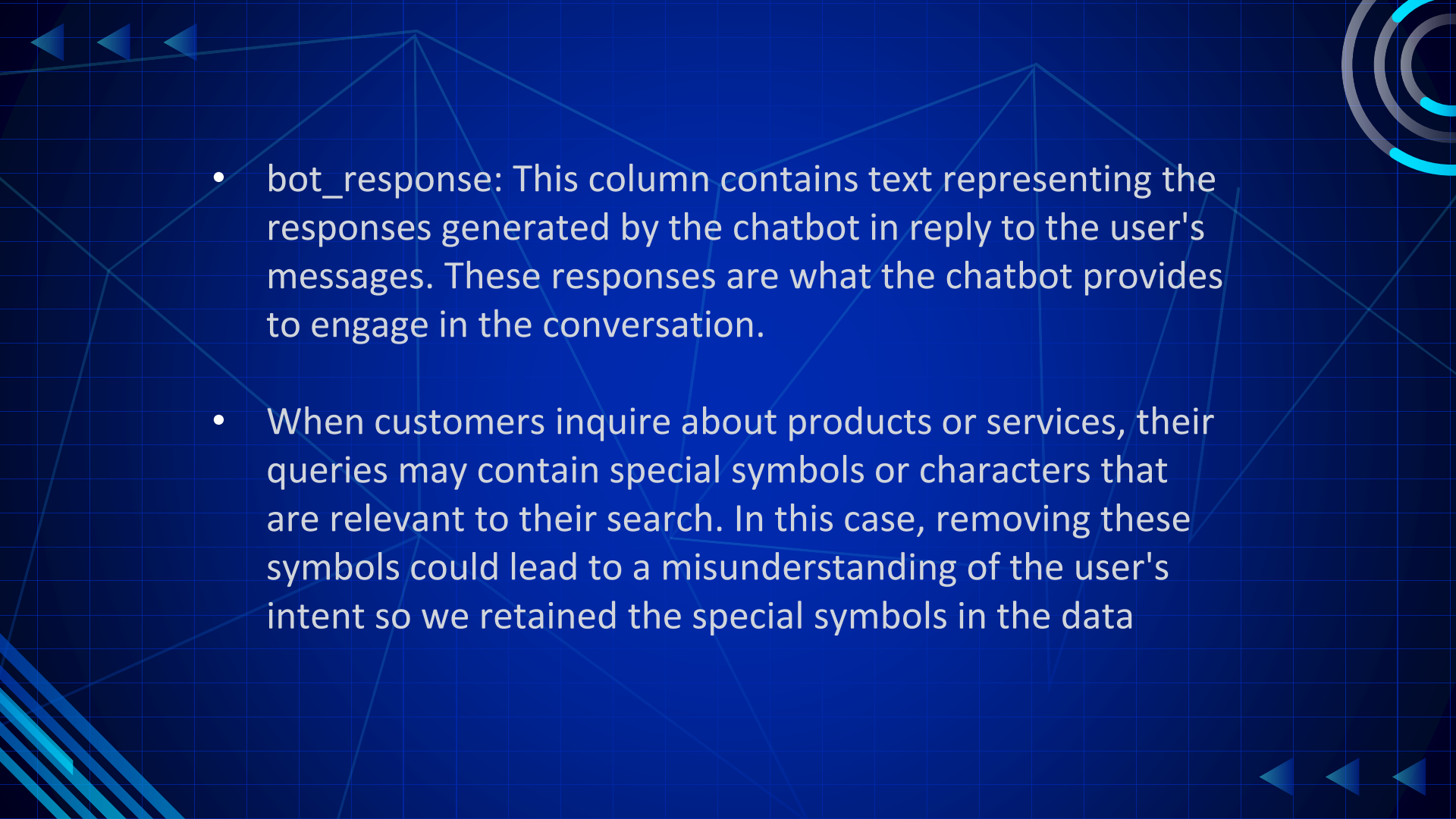
# Print the first few rows of the preprocessed data for viewing
print("Preprocessed Data:")
print(df.head())
```

We've read the dataset, removed missing values, removed leading and trailing whitespaces, converted text to lowercase for consistency, and optionally removed duplicate dialog pairs. This is a good start to ensure the data is clean and ready for further processing or for training a chatbot model.

Preprocessed Data:

	user_message \	bot_response
0	hi, how are you doing?	i'm fine. how about yourself?
1	i'm fine. how about yourself?	i'm pretty good. thanks for asking.
2	i'm pretty good. thanks for asking.	no problem. so how have you been?
3	no problem. so how have you been?	i've been great. what about you?
4	i've been great. what about you?	i've been good. i'm in school right now.

- **user_message:** This column contains text representing messages or queries from the user to the chatbot. These are the questions, statements, or requests that users make to the chatbot.

- 
- `bot_response`: This column contains text representing the responses generated by the chatbot in reply to the user's messages. These responses are what the chatbot provides to engage in the conversation.
 - When customers inquire about products or services, their queries may contain special symbols or characters that are relevant to their search. In this case, removing these symbols could lead to a misunderstanding of the user's intent so we retained the special symbols in the data

CONCLUSION

- To develop a chatbot capable of engaging in meaningful conversations, the data preprocessing phase plays a pivotal role in transforming raw dialogues into clean, structured, and consistent inputs.
- This phase has set the foundation for building a chatbot that can understand user messages and generate coherent responses.
- We'll integrate the chatbot in a web application in the final phase, when it is completed and ready to take the input from users and give the appropriate response.