

# CREATE A CHATBOT IN PYTHON

## Phase 3: Development Part 1

In this part you will begin building your project by loading and preprocessing the dataset.

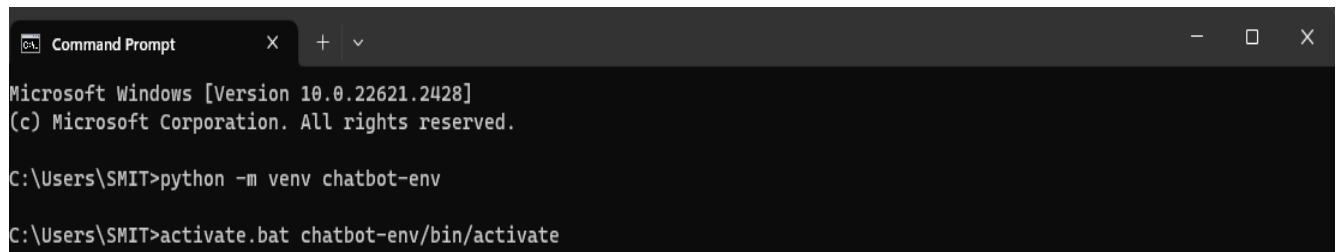
Start building the chatbot by preparing the environment and implementing basic user interactions. Install required libraries, like transformers for GPT-3 integration and flask for web app development.

**Dataset Link:** <https://www.kaggle.com/datasets/grafstor/simple-dialogs-for-chatbot>

### 1.Virtual Environment:

The very first step is to develop a chatbot, setting a virtual environment for the application.

To set an environment, create and activate the environment in Command prompt.



```
Command Prompt
Microsoft Windows [Version 10.0.22621.2428]
(c) Microsoft Corporation. All rights reserved.

C:\Users\SMIT>python -m venv chatbot-env

C:\Users\SMIT>activate.bat chatbot-env/bin/activate
```

### 2.Install Required libraries:

Install the required libraries like Pandas, transformers for GPT-3 integration and flask for web app development.

## pip install pandas

```
C:\Users\SMIT>pip install pandas
Defaulting to user installation because normal site-packages is not writeable
Collecting pandas
  Obtaining dependency information for pandas from https://files.pythonhosted.org/packages/89/c8/466196b756d74326828fe227743105fda0198a133f18916610e068540f0f/pandas-2.1.1-cp312-cp312-win_amd64.whl.metadata
    Downloading pandas-2.1.1-cp312-cp312-win_amd64.whl.metadata (18 kB)
Requirement already satisfied: numpy>=1.26.0 in c:\users\smit\appdata\roaming\python\python312\site-packages (from pandas) (1.26.0)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\smit\appdata\roaming\python\python312\site-packages (from pandas) (2.8.2)
Collecting pytz>=2020.1 (from pandas)
  Obtaining dependency information for pytz>=2020.1 from https://files.pythonhosted.org/packages/32/4d/aaf7eff5deb402fd9a24a1449a8119f00d74ae9c2efa79f8ef9994261fc2/pytz-2023.3.post1-py2.py3-none-any.whl.metadata
    Downloading pytz-2023.3.post1-py2.py3-none-any.whl.metadata (22 kB)
Collecting tzdata>=2022.1 (from pandas)
  Downloading tzdata-2023.3-py2.py3-none-any.whl (341 kB)
    341.8/341.8 kB 2.7 MB/s eta 0:00:00
Requirement already satisfied: six>=1.5 in c:\users\smit\appdata\roaming\python\python312\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Download pandas-2.1.1-cp312-cp312-win_amd64.whl (10.5 MB)
    10.5/10.5 MB 8.4 MB/s eta 0:00:00
Download pytz-2023.3.post1-py2.py3-none-any.whl (502 kB)
    502.6/502.5 kB 6.2 MB/s eta 0:00:00
Installing collected packages: pytz, tzdata, pandas
Successfully installed pandas-2.1.1 pytz-2023.3.post1 tzdata-2023.3
```

## pip install transformers

```
C:\Users\SMIT>pip install transformers
Defaulting to user installation because normal site-packages is not writeable
Collecting transformers
  Obtaining dependency information for transformers from https://files.pythonhosted.org/packages/1a/d1/3bba59606141ae808017f6fde91453882f931957f125009417b87a281067/transformers-4.34.0-py3-none-any.whl.metadata
    Downloading transformers-4.34.0-py3-none-any.whl.metadata (121 kB)
    121.5/121.5 kB 1.8 MB/s eta 0:00:00
Collecting filelock (from transformers)
  Obtaining dependency information for filelock from https://files.pythonhosted.org/packages/5e/5d/97afbafd9d584ffb45fcb354a479a3609bd97f912f8f1f6c563cb1fae21/filelock-3.12.4-py3-none-any.whl.metadata
    Downloading filelock-3.12.4-py3-none-any.whl.metadata (2.8 kB)
Collecting huggingface-hub<1.0,>=0.16.4 (from transformers)
  Obtaining dependency information for huggingface-hub<1.0,>=0.16.4 from https://files.pythonhosted.org/packages/ef/b5/b6107bd65fa4c96fdf00e4733e2fe5729bb9e5e09997f63074bb43d3ab28/huggingface-hub-0.18.0-py3-none-any.whl.metadata
    Downloading huggingface-hub-0.18.0-py3-none-any.whl.metadata (13 kB)
Requirement already satisfied: numpy>=1.17 in c:\users\smit\appdata\roaming\python\python312\site-packages (from transformers) (1.26.0)
Requirement already satisfied: packaging>=20.0 in c:\users\smit\appdata\roaming\python\python312\site-packages (from transformers) (23.2)
Requirement already satisfied: pyyaml>=5.1 in c:\users\smit\appdata\roaming\python\python312\site-packages (from transformers) (6.0.1)
Collecting regex!=2019.12.17 (from transformers)
  Obtaining dependency information for regex!=2019.12.17 from https://files.pythonhosted.org/packages/d3/10/6f2d5f8635d7714ad97ce6ade7a643358c4f3e45cde4ed12b7150734a8f3/regex-2023.10.3-cp312-cp312-win_amd64.whl.metadata
    Downloading regex-2023.10.3-cp312-cp312-win_amd64.whl.metadata (41 kB)
    42.0/42.0 kB 1.6 MB/s eta 0:00:00
Requirement already satisfied: requests in c:\users\smit\appdata\roaming\python\python312\site-packages (from transformers) (2.31.0)
Collecting tokenizers<0.15,>=0.14 (from transformers)
  Downloading tokenizers-0.14.1.tar.gz (317 kB)
    317.3/317.3 kB 3.9 MB/s eta 0:00:00

Installing build dependencies ... done
Getting requirements to build wheel ... done
```

## pip install flask

```
C:\Users\SMIT>pip install flask
Defaulting to user installation because normal site-packages is not writeable
Collecting flask
  Obtaining dependency information for flask from https://files.pythonhosted.org/packages/36/42/015c23096649b908c809c69388a805a571a3bea44362fe87e33fc3afa01f/flask-3.0.0-py3-none-any.whl.metadata
    Downloading flask-3.0.0-py3-none-any.whl.metadata (3.6 kB)
Collecting Werkzeug>=3.0.0 (from flask)
  Obtaining dependency information for Werkzeug>=3.0.0 from https://files.pythonhosted.org/packages/b6/a5/54b01f663d60d5334f6c9c87c26274e94617a4fd463d812463626423b10d/werkzeug-3.0.0-py3-none-any.whl.metadata
    Downloading werkzeug-3.0.0-py3-none-any.whl.metadata (4.1 kB)
Requirement already satisfied: Jinja2>=3.1.2 in c:\users\smit\appdata\roaming\python\python312\site-packages (from flask) (3.1.2)
Collecting itsdangerous>=2.1.2 (from flask)
  Downloading itsdangerous-2.1.2-py3-none-any.whl (15 kB)
Collecting click>=8.1.3 (from flask)
  Obtaining dependency information for click>=8.1.3 from https://files.pythonhosted.org/packages/00/2e/d53fa4befbf2cfa713304affc7ca780ce4fc1fd8710527771b58311a3229/click-8.1.7-py3-none-any.whl.metadata
    Downloading click-8.1.7-py3-none-any.whl.metadata (3.0 kB)
Collecting blinker>=1.6.2 (from flask)
  Obtaining dependency information for blinker>=1.6.2 from https://files.pythonhosted.org/packages/bf/2b/11bcdcb7dee4923253a4a21bae3be854bcc4f06295bd827756352016d97c/blinker-1.6.3-py3-none-any.whl.metadata
    Downloading blinker-1.6.3-py3-none-any.whl.metadata (1.9 kB)
Requirement already satisfied: colorama in c:\users\smit\appdata\roaming\python\python312\site-packages (from click>=8.1.3->flask) (0.4.6)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\smit\appdata\roaming\python\python312\site-packages (from Jinja2>=3.1.2->Flask) (2.1.3)
Download flask-3.0.0-py3-none-any.whl (99 kB)
    99.7/99.7 kB 1.9 MB/s eta 0:00:00
Download blinker-1.6.3-py3-none-any.whl (13 kB)
Download click-8.1.7-py3-none-any.whl (97 kB)
    97.4/97.9 kB 1.9 MB/s eta 0:00:00
Download werkzeug-3.0.0-py3-none-any.whl (226 kB)
    226.6/226.6 kB 1.2 MB/s eta 0:00:00
```

## pip install openai

```
C:\Users\SMIT>pip install openai
Defaulting to user installation because normal site-packages is not writeable
Collecting openai
  Obtaining dependency information for openai from https://files.pythonhosted.org/packages/1e/9f/385c25502f437686e4aa715969e5eaf5c2cb5e5ffa7c5cdd52f3c6ae967a/openai-0.28.1-py3-none-any.whl.metadata
  Downloading openai-0.28.1-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: requests>=2.20 in c:\users\smit\appdata\roaming\python\python312\site-packages (from openai) (2.31.0)
Collecting tqdm (from openai)
  Obtaining dependency information for tqdm from https://files.pythonhosted.org/packages/00/e5/f12a80907d0884e6dff9c16d0c0114d81b8cd07dc3ae54c5e962cc83037e/tqdm-4.66.1-py3-none-any.whl.metadata
  Downloading tqdm-4.66.1-py3-none-any.whl.metadata (57 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 57.6/57.6 kB 764.0 kB/s eta 0:00:00
Collecting aiohttp (from openai)
  Downloading aiohttp-3.8.6.tar.gz (7.4 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 7.4/7.4 MB 8.7 MB/s eta 0:00:00
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Installing backend dependencies ... done
  Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\smit\appdata\roaming\python\python312\site-packages (from requests>=2.20->openai) (3.3.0)
Requirement already satisfied: idna<4,>=2.5 in c:\users\smit\appdata\roaming\python\python312\site-packages (from requests>=2.20->openai) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\smit\appdata\roaming\python\python312\site-packages (from requests>=2.20->openai) (2.0.6)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\smit\appdata\roaming\python\python312\site-packages (from requests>=2.20->openai) (2023.7.22)
Requirement already satisfied: attrs>=17.3.0 in c:\users\smit\appdata\roaming\python\python312\site-packages (from aiohttp->openai) (23.1.0)
Collecting multidict<7.0,>=4.5 (from aiohttp->openai)
  Downloading multidict-6.0.4.tar.gz (51 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 51.3/51.3 kB 873.7 kB/s eta 0:00:00
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
```

## 3.API key for GPT-3:

### Generate API key from OpenAI to use GPT-3

#### API keys

Your secret API keys are listed below. Please note that we do not display your secret API keys again after you generate them.

Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, OpenAI may also automatically disable any API key that we've found has leaked publicly.

NAME	KEY	CREATED	LAST USED 
Secret key	sk-...JEf9	Oct 17, 2023	Never  
<a href="#">+ Create new secret key</a>			

#### Default organization

If you belong to multiple organizations, this setting controls which organization is used by default when making requests with the API keys above.

Personal 

Note: You can also specify which organization to use for each API request. See [Authentication](#) to learn more.

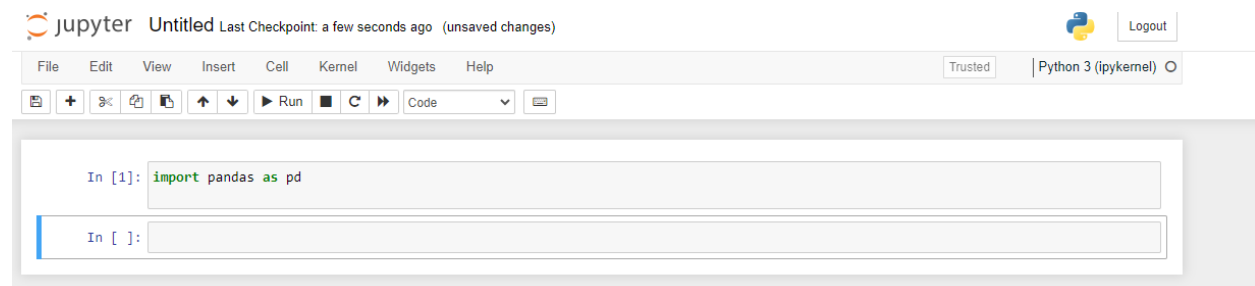
**API\_key= “ sk-qLkr4P1YkGF5Afp4BvLMT3BIbkFJlaLmbiFLtzdFDfeJAUev “**

## 4.Pre-process the Dataset:

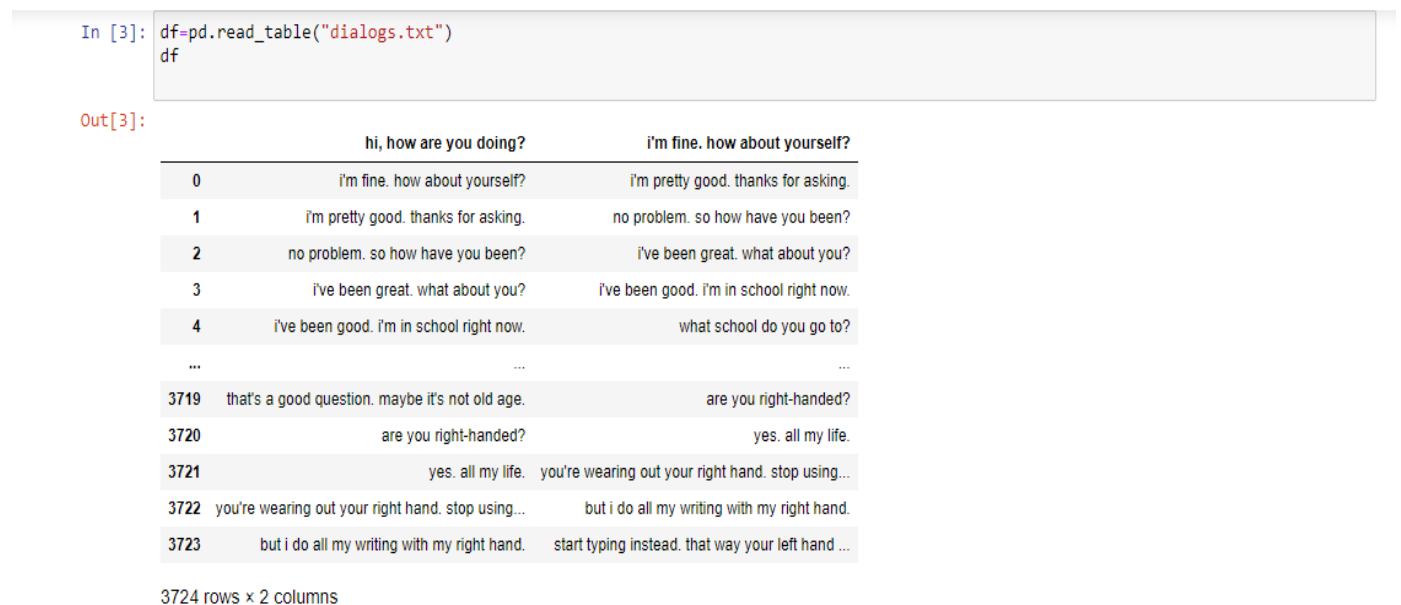
The next step is to preprocess the dataset to avoid outliers, errors and missing values while processing the application.

To preprocess the data, we use Pandas DataFrame to make it easier.

```
#import pandas
import pandas as pd
```



```
#print the dataset
df=pd.read_table("dialogs.txt")
df
```



#Check whether the dataset has null values

**df.isnull()**

3724 rows x 2 columns

In [4]: `df.isnull()`

Out[4]:

	hi, how are you doing?	i'm fine. how about yourself?
0	False	False
1	False	False
2	False	False
3	False	False
4	False	False
...	...	...
3719	False	False
3720	False	False
3721	False	False
3722	False	False
3723	False	False

3724 rows x 2 columns

#describe the dataset

**df.describe()**

In [5]: `df.describe()`

Out[5]:

	hi, how are you doing?	i'm fine. how about yourself?
count	3724	3724
unique	3509	3511
top	what do you mean?	what do you mean?
freq	22	22

In [ ]:

## 5.Pre-processed dataset:

```
3697 so, every time the plane goes up, my ear starts to hurt. there's no good
3698 that's no good. sometimes the pain goes away, and sometimes it doesn't.
3699 sometimes the pain goes away, and sometimes it doesn't. have you seen a doctor?
3700 have you seen a doctor? i've been to two doctors.
3701 i've been to two doctors. and they couldn't fix your problem?
3702 and they couldn't fix your problem? they both said i have to live with it.
3703 they both said i have to live with it. or you can stay off planes.
3704 did you see the woman with the new face? did she get a nice job?
3705 did she get a nice job? she got an "everything" job!
3706 she got an "everything" job! what do you mean?
3707 what do you mean? a team of doctors gave her a whole new face.
3708 a team of doctors gave her a whole new face. why did they do that?
3709 why did they do that? a mad dog bit most of her face off.
3710 a mad dog bit most of her face off. oh, that's terrible. what does she look like now?
3711 oh, that's terrible. what does she look like now? her face is really fat, but they say the swelling will go d
3712 her face is really fat, but they say the swelling will go down. and then will she look normal again?
3713 and then will she look normal again? i guess so.
3714 i guess so. god bless modern medicine.
3715 there's something wrong with my right hand. what's wrong with it?
3716 what's wrong with it? it aches most of the time.
3717 it aches most of the time. what do you think it is?
3718 what do you think it is? i don't know. i think it's old age.
3719 i don't know. i think it's old age. if it's old age, why don't both of your hands hurt?
3720 if it's old age, why don't both of your hands hurt? that's a good question. maybe it's not old age.
3721 that's a good question. maybe it's not old age. are you right-handed?
3722 are you right-handed? yes. all my life.
3723 yes. all my life. you're wearing out your right hand. stop using it so much.
3724 you're wearing out your right hand. stop using it so much. but i do all my writing with my right hand.
3725 but i do all my writing with my right hand. start typing instead. that way your left hand will do half the work
```

dialogs.txt

```
1 hi, how are you doing? i'm fine. how about yourself?
2 i'm fine. how about yourself? i'm pretty good. thanks for asking.
3 i'm pretty good. thanks for asking. no problem. so how have you been?
4 no problem. so how have you been? i've been great. what about you?
5 i've been great. what about you? i've been good. i'm in school right now.
6 i've been good. i'm in school right now. what school do you go to?
7 what school do you go to? i go to pcc.
8 i go to pcc. do you like it there?
9 do you like it there? it's okay. it's a really big campus.
10 it's okay. it's a really big campus. good luck with school.
11 good luck with school. thank you very much.
12 how's it going? i'm doing well. how about you?
13 i'm doing well. how about you? never better, thanks.
14 never better, thanks. so how have you been lately?
15 so how have you been lately? i've actually been pretty good. you?
16 i've actually been pretty good. you? i'm actually in school right now.
17 i'm actually in school right now. which school do you attend?
18 which school do you attend? i'm attending pcc right now.
19 i'm attending pcc right now. are you enjoying it there?
20 are you enjoying it there? it's not bad. there are a lot of people there.
21 it's not bad. there are a lot of people there. good luck with that.
22 good luck with that. thanks.
23 how are you doing today? i'm doing great. what about you?
24 i'm doing great. what about you? i'm absolutely lovely, thank you.
25 i'm absolutely lovely, thank you. everything's been good with you?
26 everything's been good with you? i haven't been better. how about yourself?
27 i haven't been better. how about yourself? i started school recently.
28 i started school recently. where are you going to school?
29 where are you going to school? i'm going to pcc.
30 i'm going to pcc. how do you like it so far?
31 how do you like it so far? i like it so far. my classes are pretty good right now.
32 i like it so far. my classes are pretty good right now. i wish you luck.
```

## 6. Create application for Flask:

Create a new directory for project and create a Python script (app.py) to build the chatbot in Visual Studio. In this script, will set up Flask application and integrate the GPT-3 model.

```
app.py > ...
1  from flask import Flask, request, jsonify
2  from transformers import GPT3Tokenizer, GPT2LMHeadModel
3  import openai # Make sure you install the 'openai' library and set your API key
4
5  app = Flask(__name__)
6
7  # Set your OpenAI GPT-3 API key
8  openai.api_key = 'sk-qLkr4P1YkGF5Afp48vLMT3B1bkFJlaLmbiFLtZdFDfeJAUev '
9
10 #import dataset
11 dataset="dialogs.txt"
12
13 # Initialize the GPT-3 model
14 tokenizer = GPT3Tokenizer.from_pretrained(dataset)
15 model = GPT2LMHeadModel.from_pretrained(dataset)
16
17 @app.route('/chat', methods=['POST'])
18 def chat():
19     user_message = request.json['user_message']
20
21     # Generate a response using GPT-3
22     input_ids = tokenizer.encode("User: " + user_message, return_tensors='pt')
23     response = model.generate(input_ids, max_length=100, num_return_sequences=1, no_repeat_ngram_size=2, top_p=0.5)
24     chat_response = tokenizer.decode(response[0], skip_special_tokens=True)
25
26     return jsonify({"response": chat_response})
27
28 if __name__ == '__main__':
29     app.run(debug=True)
30
```

## 7.Create HTML Template:

In this step, we need a HTML template to create a simple chat interface.

```
<? template.html > ...
1  <!DOCTYPE html>
2  <html>
3  <head>
4  |   <title>Chatbot</title>
5  </head>
6  <body>
7  |   <h1>Chatbot</h1>
8  |   <div id="chat">
9  |       <div id="messages"></div>
10 |       <input id="user-input" type="text" placeholder="Ask a question...">
11 |       <button onclick="sendMessage()">Send</button>
12 |   </div>
13 |   <script>
14 |       function sendMessage() {
15 |           var userMessage = document.getElementById("user-input").value;
16 |           document.getElementById("user-input").value = "";
17 |           document.getElementById("messages").innerHTML += "<p>You: " + userMessage + "</p>";
18 |
19 |           // Send the user message to the server
20 |           fetch('/chat', {
21 |               method: 'POST',
22 |               headers: {
23 |                   'Content-Type': 'application/json'
24 |               },
25 |               body: JSON.stringify({ user_message: userMessage })
26 |           })
27 |           .then(response => response.json())
28 |           .then(data => {
29 |               var botMessage = data.response;
30 |               document.getElementById("messages").innerHTML += "<p>Chatbot: " + botMessage + "</p>";
31 |           });
32 |       }
33 |   </script>
34 </body>
35 </html>
36
```