

README.md

# ML - Create a Customer Support Website with ChatGPT models

---

[Google Slides](#)

## Introduction

---

This project is to use the popular ChatGPT models to create a customer support website to answer questions from customers about their website.

## OpenAI API Models

### *text-embedding-ada-002*

OpenAI's text embeddings measure the relatedness of text strings. Embeddings are commonly used for:

- **Search** (where results are ranked by relevance to a query string)
- **Clustering** (where text strings are grouped by similarity)
- **Recommendations** (where items with related text strings are recommended)
- **Anomaly detection** (where outliers with little relatedness are identified)
- **Diversity measurement** (where similarity distributions are analyzed)
- **Classification** (where text strings are classified by their most similar label)

### *text-davinci-003*

This model builds on top of previous InstructGPT models, and improves on a number of behaviors that we've heard are important to you as developers.

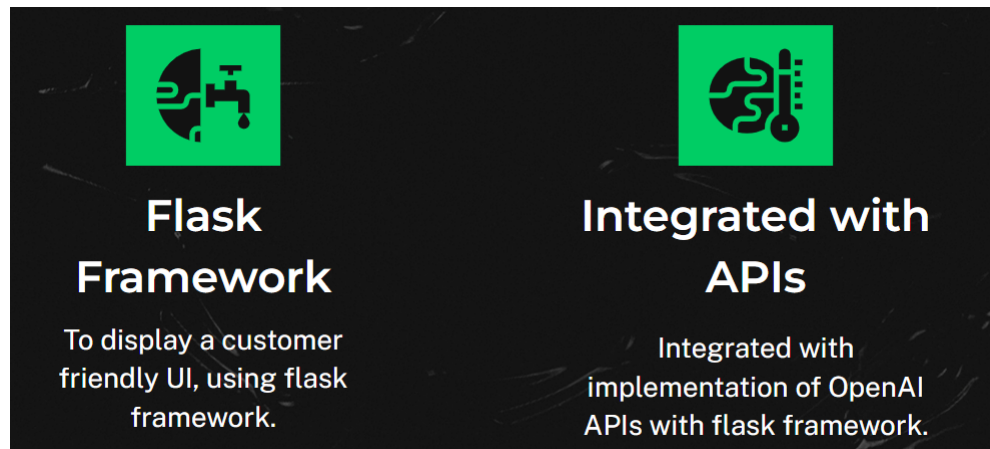
It includes the following improvements:

- It produces higher quality writing. This will help your applications deliver clearer, more engaging, and more compelling content.

- It can handle more complex instructions, meaning you can get even more creative with how you make use of its capabilities now.
- It's better at longer form content generation, allowing you to take on tasks that would have previously been too difficult to achieve.

## Design

---



### 1. Install Packages

Set up environment with installing all the packages needed in the requirements.txt file.

### 2. Crawl Data

Crawl data from the website and save the data as a .csv file for use later.

### 3. Embedded Text

Use the text-embedding-ada-002 Model to embedded the relevant data from the .csv file.

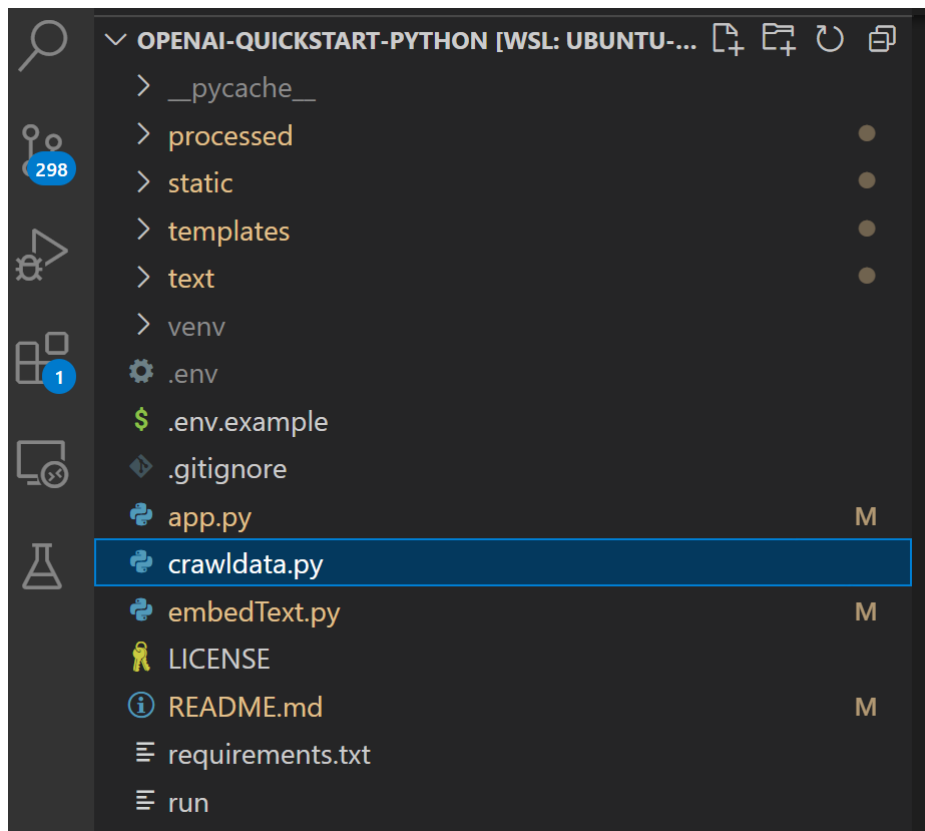
### 4. Display

Run the project with displaying the customer UI and collect questions.

## Implementation and Test

---

*This project is run on Ubuntu 22.04*



## 1. Create Python Virtual Environment and install Packages

Install the needed packages if you don't have it installed.

```
$ sudo apt install python3.10-venv
```

List all needed packages and versions in requirements.txt file.

```
$ python3 -m venv venv
$ . venv/bin/activate
$ pip install -r requirements.txt
```

You may need to update your setuptools to make the installation smooth.

```
$ pip show setuptools
$ pip3 install --upgrade pip setuptools or
$ pip install --upgrade setuptools --user python
```

## 2. Crawl Data

```
$ python3 crawldata.py
```

```
(venv) sharoncao0920@Sharon:~/openai/openai-quickstart-python$ python3 crawldata.py
https://openai.com/
https://openai.com/policies
https://openai.com/policies/api-data-usage-policies
https://openai.com/contact-sales
https://openai.com/contact-sales#content
```

## 3. Embedded Text

```
$ python3 embedText.py
```

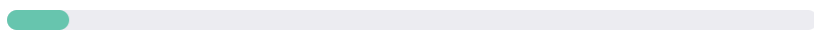
```
openai.error.RateLimitError: Rate limit reached for default-global-with-image-limits in organization org-5Wp0wasio09MFN8F2b7ki4ap on requests per min. Limit : 60 / min. Please try again in 1s. Contact support@openai.com if you continue to have issues. Please add a payment method to your account to increase your rate limit. Visit https://platform.openai.com/account/billing to add a payment method.
(venv) sharoncao0920@Sharon:~/openai/openai-quickstart
```

*No payment method linked, not working through since exceed limit of 60 requests per mins. Worked through after adding payment method.*

```
PROBLEMS OUTPUT TERMINAL DEBUG CONSOLE
bash - openai-quickstart-python
(venv) sharoncao0920@Sharon:~/openai/openai-quickstart-python$ python3 embedText.py
(venv) sharoncao0920@Sharon:~/openai/openai-quickstart-python$
```

Total cost for embedding the data.

Usage this month

 \$0.38 / \$5.00

## Tips on Saving Money for Developing

OpenAI charges based on tokens, so to minimize the cost of development for this project, I used a small size of data while developing to make the use of the grant given by the platform.

1. Run `$ python3 crawldata.py` to make sure data crawling works
2. Create your own testing data and save it as `processed/data.csv` using the same structure as of `processed/scraped.csv`.

A	B	C	D	E	F	G	H
	title	text					
0	apple	Apple is a bit sour. It has red, green, yellow. The score is 3.5.					
1	peach	Peach is very delicious and the score is 5					

```
▼ processed
  data.csv
  embeddings.csv
  scraped.csv
```

3. Modify the code in `embedText.py` to read csv file `processed/data.csv` instead of `processed/scraped.csv`.

```
df = pd.read_csv('processed/scraped.csv', index_col=0)
df.columns = ['title', 'text']
```

4. Run `$ python3 embedText.py` to embed the data.
5. Run `$ flask run` to test the project by asking questions about the testing data in browser.



## Customer Service

Ask your question about us

**You:**

Ask

**Agent:**

Apple is a fruit that is a bit sour.

6. Once the project is working as desired, link your payment method on OpenAI and embed `processed/scraped.csv` then test project with `$ flask run`.

### 4. Display

```
$ flask run
```

```
(venv) sharoncao0920@Sharon:~/openai/openai-quickstart-python$ flask run
* Serving Flask app 'app' (lazy loading)
* Environment: development
* Debug mode: on
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
* Restarting with stat
* Debugger is active!
* Debugger PIN: 361-744-753
```

You should now be able to access the app at <http://localhost:5000>

### *Sample Questions and Answers:*



## Customer Service

Ask your question about us

**You:**

What is CEO of OPENAI?

Ask

**Agent:**

Sam Altman



## Customer Service

Ask your question about us

**You:**

How can I use GPT4 as soon as I can?

Ask

**Agent:**

Join the GPT-4 API waitlist by submitting your first name, last name, email, company name, and organization ID. Priority access is given to developers who contribute exceptional model evaluations to OpenAI Evals. Researchers studying the societal impact of AI or AI alignment issues can also apply for subsidized access via the Researcher Access Program.



## Customer Service

Ask your question about us

**You:**

What is the difference between GPT3.5 and GPT4.0?

Ask

**Agent:**

GPT-4 is more reliable, creative, and able to handle much more nuanced instructions than GPT-3.5. GPT-4 scores 40% higher than GPT-3.5 on internal adversarial factuality evaluations and is better at separating fact from an adversarially-selected set of incorrect statements. GPT-4 also generally lacks knowledge of events that have occurred after the vast majority of its data cuts off and does not learn from its experience.



## Customer Service

Ask your question about us

**You:**

How much does it cost to be a premium user of chatGPT?

Ask

**Agent:**

\$20/month.

## Conclusion

---

### Accuracy



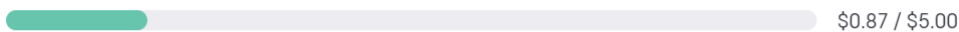
- As the questions I asked based on my own data and data crawled from the website, the answer is quite accurate.
- I think the accuracy is more determined by the data provided by the website.

## Cost

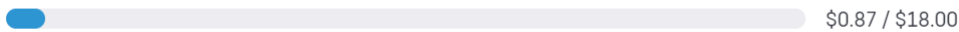
- To avoid high charges, we can use less data while developing. Use the real data only when the whole project go through as desired.
- In this project, the embedding text can be run only once to save money.

### *Final cost*

#### Usage this month



#### Credit used ⓘ



GRANT #	CREDIT GRANTED	EXPIRES (UTC)
Grant 1	\$18.00	May 1, 2023

## Enhancement

---

- Improve UI
- Implementate more APIs for customer services

## References

---

- <https://platform.openai.com/docs/guides/embeddings/what-are-embeddings>
- <https://help.openai.com/en/articles/6779149-how-do-text-davinci-002-and-text-davinci-003-differ>