

Customer Support Website - with ChatGPT

Maryam Zubair
Student Id: 19709

Table of Contents

- ❖ Introduction
- ❖ Design & Implementation
- ❖ Test
- ❖ Enhancement Ideas
- ❖ Conclusion
- ❖ References



Introduction

The purpose of the project is creating an “Answer Question System” utilizing GPT 3.5 Turbo model by OpenAI. The system aims to provide accurate and relevant answers to the questions asked related to the content on the OpenAI website.



Design & Implementation

The implementation involved the following steps:

1. **Web Scraping:** scrape relevant text content from the OpenAI website.
2. **Text Preprocessing:** Processed the collected text by tokenizing and removing unnecessary characters.
3. **Embeddings:** Utilized tokenization and embeddings to convert the text into numerical vectors
4. **GPT-3.5 Turbo for Answer Generation:** Utilized GPT-3.5 Turbo for generating context-based answers using embeddings of the question and available text.
5. **Integration:** Integrated OpenAI APIs with flask framework.



Web Crawling

The process began with scraping the “OpenAI website” to collect relevant textual content. Python libraries like **requests**, **BeautifulSoup**, and **urllib** were employed for this purpose. The requests library was used to send HTTP requests to the OpenAI website, BeautifulSoup facilitated parsing the HTML content, and urllib aided in extracting the URLs.



```
▼ text / openai.com
├─ openai.com_.txt
├─ openai.com_#.content.txt
├─ openai.com_about.txt
├─ openai.com_about#.content.txt
├─ openai.com_alignment.txt
├─ openai.com_api-data-privacy.txt
├─ openai.com_api.txt
├─ openai.com_blog_ai-and-compute_#lookingforward.txt
├─ openai.com_blog_ai-and-compute_#modern.txt
├─ openai.com_blog_ai-and-compute.txt
├─ openai.com_blog_amplifying-ai-training.txt
├─ openai.com_blog_announcing-openai-devday.txt
├─ openai.com_blog_announcing-openai-devday#.content.txt
├─ openai.com_blog_announcing-openai-devday#OpenAI.txt
├─ openai.com_blog_api-no-waitlist.txt
├─ openai.com_blog_api-no-waitlist#.content.txt
├─ openai.com_blog_api-no-waitlist#OpenAI.txt
├─ openai.com_blog_best-practices-for-deploying-language-models.txt
├─ openai.com_blog_best-practices-for-deploying-language-models#.content.txt
├─ openai.com_blog_best-practices-for-deploying-language-models#OpenAI.txt
├─ openai.com_blog_better-language-models_#update.txt
```

~/Desktop/t
openai.com
Untracked

Text Preprocessing

Upon collecting the text content from the OpenAI website, preprocessing was carried out which involved two main steps:

1. **Tokenization:** The collected text was broken down into smaller units called tokens
2. **Character Removal:** Unnecessary characters and formatting were removed to ensure a clean representation of the text.



Scraped Data:-

```
processed >  scraped.csv
1 ,fname,text
2 0,careers research scientist safety#content,"careers research scientist safety#content. Research Scientist, Safety CloseSearch Submit
3 1,research?authors=henk tillman,"research?authors=henk tillman. Research index CloseSearch Submit Skip to main contentSite Navigation
4 2,research?authors=arthur petron,"research?authors=arthur petron. Research index CloseSearch Submit Skip to main contentSite Navigati
5 3,dall e 3#content,"dall e 3#content. DALL·E 3 CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALL·E
6 4,policies service credit terms,policies service credit terms. Service credit terms
7 5,research roboschool#content,"research roboschool#content. Roboschool CloseSearch Submit Skip to main contentSite NavigationResearc
8 6,blog announcing openai devday,"blog announcing openai devday. Join us for OpenAI's first developer conference on November 6 in San Fr
9 7,blog dall e 2 update#OpenAI,"blog dall e 2 update#OpenAI. DALL·E 2 research preview update CloseSearch Submit Skip to main content
10 8,research?authors=amanda askell,"research?authors=amanda askell. Research index CloseSearch Submit Skip to main contentSite Navigati
11 9,careers graph compiler engineer#content,"careers graph compiler engineer#content. Graph Compiler Engineer CloseSearch Submit Skip to
12 10,blog image gpt,"blog image gpt. Image GPT CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALL·E
13 11,pricing#content,"pricing#content. Pricing CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALL·E
14 12,careers media relations europe lead#content,"careers media relations europe lead#content. Media Relations, Europe Lead CloseSearch
15 13,blog better language models ,"blog better language models . Better language models and their implications CloseSearch Submit Skip
16 14,blog solving rubiks cube,"blog solving rubiks cube. Solving Rubik's Cube with a robot hand CloseSearch Submit Skip to main conten
17 15,blog openai codex#OpenAI,"blog openai codex#OpenAI. OpenAI Codex CloseSearch Submit Skip to main contentSite NavigationResearchOv
18 16,research?authors=mark chen,"research?authors=mark chen. Research index CloseSearch Submit Skip to main contentSite NavigationResea
19 17,careers software engineer chatgpt data flywheel#content,"careers software engineer chatgpt data flywheel#content. Software Engineer,
20 18,careers associate general counsel emea 2#content,"careers associate general counsel emea 2#content. Associate General Counsel, EMEA
21 19,research summarizing books,"research summarizing books. Summarizing books with human feedback CloseSearch Submit Skip to main con
22 20,careers backend engineer review training platform 2,"careers backend engineer review training platform 2. Backend Engineer, Review &
23 21,blog grade school math #samples,"blog grade school math #samples. Solving math word problems CloseSearch Submit Skip to main cont
24 22,customer stories viable#content,"customer stories viable#content. Viable CloseSearch Submit Skip to main contentSite NavigationRe
25 23,blog concrete ai safety problems,"blog concrete ai safety problems. Concrete AI safety problems CloseSearch Submit Skip to main c
26 24,blog openai five finals,"blog openai five finals. OpenAI Five Finals CloseSearch Submit Skip to main contentSite NavigationResear
27 25,blog?authors=rachel lim,"blog?authors=rachel lim. Blog CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndex
28 26,customer stories morgan stanley#content,"customer stories morgan stanley#content. Morgan Stanley CloseSearch Submit Skip to main
29 27,blog our approach to alignment research,"blog our approach to alignment research. Our approach to alignment research CloseSearch
30 28,careers search#content,"careers search#content. Careers at OpenAI CloseSearch Submit Skip to main contentSite NavigationResearchOv
31 29,blog introducing superalignment#IlyaSutskever,"blog introducing superalignment#IlyaSutskever. Introducing Superalignment CloseSea
32 30,careers customer success manager#content,"careers customer success manager#content. Customer Success Manager CloseSearch Submit Ski
33 31,research?topics=dota 2,"research?topics=dota 2. Research index CloseSearch Submit Skip to main contentSite NavigationResearchOv
34 32,research?authors=rewon child,"research?authors=rewon child. Research index CloseSearch Submit Skip to main contentSite NavigationR
35 33,careers assistant controller,"careers assistant controller. Assistant Controller CloseSearch Submit Skip to main contentSite Naviga
36 34,careers software engineer hardware health#content,"careers software engineer hardware health#content. Software Engineer, Hardware He
37 35,research spam detection in the physical world#content,"research spam detection in the physical world#content. Spam detection in the
38 36,research?topics=games,"research?topics=games. Research index CloseSearch Submit Skip to main contentSite NavigationResearchOverview
39 37,product "product. Product CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALL·E 3 APIOverviewDat
```

Embeddings

Next, the preprocessed text was further transformed into numerical vectors called embeddings. Tokenization played a crucial role in this step, allowing the conversion of text into a numerical format understandable by machine learning models. These embeddings were used to represent the text in a format suitable for feeding into the GPT-3.5 Turbo model.



Embeddings Index:

```
processed > embeddings.csv
```

```
1 ,text,n_tokens,embeddings
2 0,"careers research scientist safety#content. Research Scientist, Safety CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALLÂ·E 3 APIOverviewData print
3 1,"in computer science, machine learning, or a related field, with 5+ years of related research experience.Experience in the field of AI safety
4 2,research?authors=henk tillman.,9,"[-0.006426393985748291, -0.020367804914712906, -0.006026061251759529, -0.035875432193279266, 0.0128738582
5 3,research?authors=arthur petron.,9,"[-0.004728116560727358, -0.006599515210837126, -0.02088310569524765, -0.03754138946533203, 0.01229877397
6 4,"dall e 3#content. DALL·E 3 CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALLÂ·E 3 APIOverviewData print
7 5,"As with DALLÂ·E 2, the images you create with DALLÂ·E 3 are yours to use and you don't need our permission to reprint, sell or merchandise
8 6,policies service credit terms. Service credit terms ,11,"[-0.012243076227605343, -0.011360629461705685, 0.008386715315282345, -0.0188718
9 7,"research roboschool#content. Roboschool CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALLÂ·E 3 APIOverviewData print
10 8,"Here are three of the environments we ported, with explanations of how they differ from the existingÂ environments.RoboschoolExistingComm
11 9,"blog announcing openai devday. Join us for OpenAI's first developer conference on November 6 in San Francisco CloseSearch Submit Skip
12 10,"blog dall e 2 update#OpenAI. DALL·E 2 research preview update CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALLÂ·E 3 APIOverviewData print
13 11,research?authors=amanda askell.,9,"[0.002515760250389576, -0.015122206881642342, -0.007402140647172928, -0.02916623093187809, -0.006434540
14 12,"careers graph compiler engineer#content. Graph Compiler Engineer CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALLÂ·E 3 APIOverviewData print
15 13,"blog image gpt. Image GPT CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALLÂ·E 3 APIOverviewData print
16 14,"When we train GPT-2 on images unrolled into long sequences of pixels, which we call iGPT, we find that the model appears to understand 2-
17 15," To highlight the potential of generative[^reference-17][^reference-18]Â sequence modeling[^reference-19][^reference-20][^reference-21]
18 16,"In contrast, sequences of pixels do not clearly contain labels for the images they belongÂ to.Even without this explicit supervision, the
19 17,"However, in showing that an unsupervised transformer model is competitive with the best unsupervised convolutional nets,[^reference-24]
20 18,"In the second phase, this contextualized feature is used to solve the conditional next pixel prediction task. The observed two stage per
21 19," Pre-trained on ImageNet Evaluation Model Accuracy w/o labels w/ labels CIFAR-10Linear Probe ResNet-152[^reference-50] 94.0 Â SimCLR
22 20,"Taking 15360 features from 5 layers in iGPT-XL yields 72.0% top-1 accuracy, outperforming AMDTM, MoCo, and CPC v2, but still underperform
23 21,"While unsupervised learning promises excellent features without the need for human-labeled data, significant recent progress has been made
24 22,"Our approach to semi-supervised learning is very simple since we only fit a logistic regression classifier on iGPT-L's features without a
25 23,"pricing#content. Pricing CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALLÂ·E 3 APIOverviewData print
26 24,"Once you fine-tune a model, youâ€ ll be billed only for the tokens you use in requests to that model.Learn about fine-tuningModelTraining
27 25,"careers media relations europe lead#content. Media Relations, Europe Lead CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALLÂ·E 3 APIOverviewData print
28 26,"Passion for new technology and storytelling is key.In this role, you will:Develop communication plans and manage day-to-day executionÂ For
29 27,"blog better language models . Better language models and their implications CloseSearch Submit Skip to main contentSite NavigationResearchOverviewIndexGPT-4DALLÂ·E 3 APIOverviewData print
30 28,"In addition, GPT-2 outperforms other language models trained on specific domains (like Wikipedia, news, or books) without needing to use
31 29,"PÃ³rez stated, â€ We can see, for example, that they have a common â€ language,â€ something like a dialect or dialectic.â€ Dr. PÃ³rez
32 30,"System Prompt (human-written) Miley Cyrus was caught shoplifting from Abercrombie and Fitch on Hollywood Boulevard today. Model Completion
33 31,"That is, we are developing a machine language system in the generative style with no explicit rules for producing text.We hope for future
34 32,"Frodo and Sam woke up alone in their room, and Frodo found a note on his pillow. He opened it and read:May the Power of the Ring be with
35 33,"Jefferson himself, as you well know, is an author of the Declaration of Independence who had some really crazy ideas about, well, you know
36 34,"In the months and years to come, there will be many battles in which we will have to be strong and we must give all of our energy, not to
37 35,"I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming
38 36,"When prompted with topics that are highly represented in the data (Brexit, Miley Cyrus, Lord of the Rings, and so on), it seems to be cap
39 37,"(â€ ) means a lower score isÂ better.DatasetMetricOur resultPrevious recordHumanWinograd Schema Challengeaccuracy (+)70-70%63-7%92%+1 AMB
```

GPT-3.5 Turbo for Answer Generation

After text was prepared and represented as embeddings, GPT-3.5 Turbo (language generation model) was utilized for answer generation. The embeddings of the question and the available text were fed into the GPT-3.5 Turbo model, allowing it to comprehend the context and generate context-based answers to user queries.



Testing

Testing performed to ensure the system's functionality and accuracy through two interfaces which are mentioned below:-

1. Command-Line Interface (CLI)
2. Web-Based Interface:




Command-Line Interface (CLI) Testing:

Asked variety of questions through the 'terminal' to validate the system's ability to provide relevant answers.

```
/usr/local/bin/python3 /Users/maryamz/Desktop/Generative_AI/commandLineSol.py
maryamz@Maryams-MacBook-Pro Generative_AI % /usr/local/bin/python3 /Users/maryamz/Desktop/Generative_AI/commandLineSol.py
Question: Who are the founders of OpenAI?
ChatGPT: The founders of OpenAI are Sam Altman, Greg Brockman, Trevor Blackwell, Vicki Cheung, Andrej Karpathy, Durk Kingma, John Schulman, Pamela Vagata, and Wojciech Zaremba.
Question: What are the key focus areas of OpenAI's research
ChatGPT: The key focus areas of OpenAI's research are building a household robot, developing an agent with natural language understanding, and solving a wide variety of games using a single agent.
Question: Can you tell me about GPT-3.5 Turbo and its applications?
ChatGPT: Yes, GPT-3.5 Turbo is a model that has been fine-tuned for specific use cases and can be customized by developers to perform better for their needs. It has been shown to match or even outperform base GPT-4-level capabilities on certain tasks. Over 300 applications are currently using GPT-3, and it has been utilized in various industries such as productivity, education, creativity, and games.
Question: █
```

Web-Based Interface Testing:

The process involved in creating the web-based application for question-answering is :-

1. **Flask application** was configured with defined routes to handle different endpoints.
 2. **HTML Templates** defined the structure and design of the web interface where users input their questions.
 3. **User Interaction and Form Submission:** Where users interacted with the web interface by inputting questions through a form. The input was submitted to the Flask application via HTTP requests.
 4. **Answer Presentation:** The submitted questions were processed, and context was generated using GPT-3.5 Turbo.
- 

Flask Application:

```
* Serving Flask app 'webBasedSol'
```

```
* Debug mode: on
```

```
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
```

```
* Running on http://127.0.0.1:5000
```

```
Press CTRL+C to quit
```

```
* Restarting with stat
```

```
* Debugger is active!
```

```
* Debugger PIN: 280-261-619
```

```
|
```

Web Based Interface:

OpenAI Question Answering

Ask a question:

What are the key focus areas of OpenAI's research?

Submit

Answer: The key focus areas of OpenAI's research are building a household robot, developing an agent with natural language understanding, and solving a wide variety of games using a single agent.

Sample
Question
Answer

Web Based Interface:

Sample
Question
Answer

OpenAI Question Answering

Ask a question:

Can you tell me about GPT-3.5 Turbo and its applications?

Submit

Answer: Yes, GPT-3.5 Turbo is a model that has been fine-tuned for specific use cases and can be customized by developers to perform better for their needs. It has been shown to match or even outperform base GPT-4-level capabilities on certain tasks. Over 300 applications are currently using GPT-3, and it has been utilized in various industries such as productivity, education, creativity, and games.

Web Based Interface:

OpenAI Question Answering

Ask a question:

What are OpenAI's future plans regarding AGI?

Submit

Answer: OpenAI's future plans regarding AGI include ensuring that it benefits all of humanity, empowering humanity to flourish, and navigating potential risks.

Sample
Question
Answer

Enhancement Ideas

1. **Real-time Updates:** Implementing a feature to regularly update the dataset by re-scraping the OpenAI website in order to ensure my question answer system remains up-to-date with the latest information.
2. **Multi-language Support:** I would love to enhance the system to support questions and answers in multiple languages, to make it accessible to global audience.
3. **Enhanced User Interface (UI):** I can work on creating an even more intuitive and appealing User Interface to enhance user engagement.



Conclusion

In this project, I successfully created a question-answering system using GPT-3.5 Turbo by OpenAI. I applied web scraping, data processing, embeddings, and GPT-3.5 Turbo, to craft an application that provides informative and accurate responses to user's queries. The versatility of a web-based interface and command-line interaction underscores its usability. This project signifies the power of integrating AI technologies for practical and insightful applications



References

<https://platform.openai.com/docs/tutorials/web-qa-embeddings>

<https://www.digitalocean.com/community/tutorials/how-to-create-your-first-web-application-using-flask-and-python-3>

<https://www.geeksforgeeks.org/flask-http-methods-handle-get-post-requests/>



Github Link

[https://github.com/Maryam-Zubair/MachineLearning_Assignment/tree/main/ChatGPT/Use%20ChatGPT%20to%20create%20customer%20support%20website%20\(data%20source:%20web%20pages\)](https://github.com/Maryam-Zubair/MachineLearning_Assignment/tree/main/ChatGPT/Use%20ChatGPT%20to%20create%20customer%20support%20website%20(data%20source:%20web%20pages))

