

Customer Service Assistant Evaluation

Moderation, Classification & Evaluation

Index

— — —

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

Introduction

— — —

This repository contains a set of Python scripts that demonstrate the use of the OpenAI GPT-3.5 Turbo model to assist with customer queries. The scripts provide functionality to classify customer queries, evaluate responses, and generate relevant information based on product-related queries.

Design

File Name	Description
chain_of_thoughts.py	This script performs chain-of-thought processing to generate meaningful results from a series of connected thoughts.
check_output.py	This script checks the output of a particular process or operation to ensure it meets the desired criteria.
classification.py	This script is used for classifying data into different categories or groups based on specific criteria.
evaluation1.py	These scripts are used for evaluating different aspects of a dataset or process. They provide insight into the effectiveness or accuracy of the operation.
evaluation2.py	These scripts are used for evaluating different aspects of a dataset or process. They provide insight into the effectiveness or accuracy of the operation.
moderation.py	This script performs moderation by filtering out unwanted or inappropriate content from a dataset.
products_data.py	This script processes and analyzes the data from products.json to extract meaningful information.
utils.py	This script contains utility functions that can be used across different scripts to perform common operations.

Implementation

— — —

1. Create and activate a virtual environment:

```
• (venv) chinni@LAPTOP-UVNKR98P:~/W4$ python3 -m venv venv  
• (venv) chinni@LAPTOP-UVNKR98P:~/W4$ . venv/bin/activate
```

2. Install required modules

```
• (venv) chinni@LAPTOP-UVNKR98P:~/W4$ pip install python-dotenv openai utils
```

3. Run the python scripts

```
• (venv) chinni@LAPTOP-UVNKR98P:~/W4$ python3 moderation.py
```

Test (Moderation)

```
• (venv) chinni@LAPTOP-UVNKR98P:~/w4$ python3 moderation.py
```

```
Generating customer comment...
```

The TechPro Ultrabook is a sleek and lightweight laptop that is perfect for everyday use. With its 13.3-inch display, 8GB RAM, 256GB SSD, and Intel Core i5 processor, it offers a great balance of performance and portability. The BlueWave Gaming Laptop, on the other hand, is a high-performance gaming laptop that provides an immersive gaming experience. It features a 15.6-inch display, 16GB RAM, 512GB SSD, and NVIDIA GeForce RTX 3060. Both laptops come with a 1-year warranty and have received high ratings from customers. Overall, these products from the electronic company offer a range of options for different needs and preferences.

```
Checking moderation...
```

```
{
  "flagged": false,
  "categories": {
    "sexual": false,
    "hate": false,
    "harassment": false,
    "self-harm": false,
    "sexual/minors": false,
    "hate/threatening": false,
    "violence/graphic": false,
    "self-harm/intent": false,
    "self-harm/instructions": false,
    "harassment/threatening": false,
    "violence": false
  }
},
```

Test (Moderation)

— — —

```
"category_scores": {  
  "sexual": 2.89301306111156e-06,  
  "hate": 8.420877151138484e-08,  
  "harassment": 7.710971061669625e-08,  
  "self-harm": 8.828115127812453e-09,  
  "sexual/minors": 2.755083938765779e-09,  
  "hate/threatening": 6.503595439966148e-10,  
  "violence/graphic": 2.59891066889395e-07,  
  "self-harm/intent": 2.7863797935623325e-09,  
  "self-harm/instructions": 1.6466615848642618e-09,  
  "harassment/threatening": 2.203749849272185e-09,  
  "violence": 7.955021828820463e-06  
}
```

Checking Prompt injection...

Hello! As a helpful chatbot, I'm here to assist you in finding the best laptop for a student. There are several factors to consider when choosing a laptop for a student, such as budget, performance, and portability. Could you please provide me with more information about your requirements and preferences?

Test (Classification)

— — —

```
• (venv) chinni@LAPTOP-UVNKR98P:~/w4$ python3 classification.py
User query: I want to unsubscribe from your service.
System classification:
{
  "primary": "Billing",
  "secondary": "Unsubscribe or upgrade"
}
User query: Tell me more about your flat screen tvs
System classification:
{
  "primary": "General Inquiry",
  "secondary": "Product information"
}
```


Test (Chain of thoughts)

— — —

```
• (venv) chinni@LAPTOP-UVNKR98P:~/W4$ python3 chain_of_thoughts.py
Response to user 1: The BlueWave Chromebook is priced at $249.99, while the TechPro Desktop is priced
  at $999.99. Therefore, the BlueWave Chromebook is $750 cheaper than the TechPro Desktop.
Response to user 2: Yes, we do sell TVs. How can I assist you with your TV-related query?
• (venv) chinni@LAPTOP-UVNKR98P:~/W4$ python3 chain_of_thoughts.py
User message 1: by how much is the BlueWave Chromebook more expensive than the TechPro Desktop
Response to user 1: The BlueWave Chromebook is priced at $249.99, while the TechPro Desktop is priced
  at $999.99. Therefore, the BlueWave Chromebook is $750 cheaper than the TechPro Desktop.
User message 2: do you sell TVs
Response to user 2: Yes, we do sell TVs. How can I assist you with your TV-related query?
```

Test (Check output)

```
• (venv) chinni@LAPTOP-UVNKR98P:~/w4$ python3 check_output.py
```

```
Moderation Output (Initial Response): {
```

```
  "flagged": false,
  "categories": {
    "sexual": false,
    "hate": false,
    "harassment": false,
    "self-harm": false,
    "sexual/minors": false,
    "hate/threatening": false,
    "violence/graphic": false,
    "self-harm/intent": false,
    "self-harm/instructions": false,
    "harassment/threatening": false,
    "violence": false
```

```
  },
  "category_scores": {
    "sexual": 5.896513357583899e-06,
    "hate": 5.453545171008045e-08,
    "harassment": 2.681723572095507e-07,
    "self-harm": 3.387907909768728e-08,
    "sexual/minors": 4.7024151683672244e-08,
    "hate/threatening": 3.2540561356597664e-09,
    "violence/graphic": 2.2412559985696134e-07,
    "self-harm/intent": 3.443058460561588e-08,
    "self-harm/instructions": 4.274597387166068e-08,
    "harassment/threatening": 3.835509687633021e-09,
    "violence": 1.6353840237570694e-06
```

```
  }
}
```

```
Evaluation Result (Initial Response): N
```

```
Moderation Output (Another Response): {
```

```
  "flagged": false,
  "categories": {
    "sexual": false,
    "hate": false,
    "harassment": false,
    "self-harm": false,
    "sexual/minors": false,
    "hate/threatening": false,
    "violence/graphic": false,
    "self-harm/intent": false,
    "self-harm/instructions": false,
    "harassment/threatening": false,
    "violence": false
```

```
  },
  "category_scores": {
    "sexual": 1.7591315554454923e-05,
    "hate": 8.64287358126603e-06,
    "harassment": 0.0009129160316661,
    "self-harm": 7.155762205002247e-07,
    "sexual/minors": 2.1109716641376508e-08,
    "hate/threatening": 8.546206520243516e-10,
    "violence/graphic": 1.7114570027843e-07,
    "self-harm/intent": 2.5052020191651536e-07,
    "self-harm/instructions": 1.9060820477534435e-08,
    "harassment/threatening": 2.4462151486659423e-06,
    "violence": 3.3936059480765834e-05
```

```
  }
}
```

```
Evaluation Result (Another Response): N
```

Test (Evaluation1)

```
• (venv) chinni@LAPTOP-UVNR98P:~/w4$ python3 evaluation1.py
Customer message: What Gaming consoles would be good for my friend who is into racing games?
Ideal answer: {'Gaming Consoles and Accessories': {'GameSphere X', 'GameSphere VR Headset', 'ProGamer Controller', 'ProGamer Racing Wheel', 'GameSphere Y'}}
Response:
[{'category': 'Gaming Consoles and Accessories', 'products': ['GameSphere X', 'ProGamer Controller', 'GameSphere Y', 'ProGamer Racing Wheel', 'GameSphere VR Headset']}]

example 0
0: 1.0
example 1
incorrect
prod_set: {'MobiTech PowerCase', 'SmartX MiniPhone', 'SmartX EarBuds', 'SmartX ProPhone', 'MobiTech Wireless Charger'}
prod_set_ideal: {'MobiTech PowerCase', 'SmartX EarBuds', 'MobiTech Wireless Charger'}
response is a superset of the ideal answer
1: 0.0
example 2
2: 1.0
example 3
3: 1.0
example 4
incorrect
prod_set: {'SoundMax Soundbar', 'CineView OLED TV', 'SoundMax Home Theater', 'CineView 4K TV', 'CineView 8K TV'}
prod_set_ideal: {'CineView 8K TV'}
response is a superset of the ideal answer
incorrect
prod_set: {'GameSphere X', 'GameSphere VR Headset', 'ProGamer Controller', 'ProGamer Racing Wheel', 'GameSphere Y'}
prod_set_ideal: {'GameSphere X'}
response is a superset of the ideal answer
4: 0.3333333333333333
example 5
5: 1.0
example 6
6: 1.0
example 7
7: 1.0
example 8
8: 1.0
example 9
9: 1
Fraction correct out of 10: 0.8333333333333334
```

Test (Evaluation2)

```
• (venv) chinni@LAPTOP-UVNKR98P:~/W4$ python3 evaluation2.py
Assistant's Answer for customer message:
Sure! Let me provide you with some information about the SmartX ProPhone and the FotoSnap DSLR Camera.

The SmartX ProPhone is a powerful smartphone with advanced camera features. It has a 6.1-inch display, 128GB storage, a 12 MP dual camera, and supports 5G connectivity. The SmartX ProPhone is priced at $899.99 and comes with a 1-year warranty.

The FotoSnap DSLR Camera is a versatile camera that allows you to capture stunning photos and videos. It features a 24.2MP sensor, 1080p video recording, a 3-inch LCD screen, and supports interchangeable lenses. The FotoSnap DSLR Camera is priced at $599.99 and also comes with a 1-year warranty.

As for TVs and TV-related products, we have a range of options available. Some of our popular TV models include the CineView 4K TV, CineView 8K TV, and CineView OLED TV. We also have the SoundMax Home Theater system and SoundMax Soundbar for an enhanced audio experience.

Is there anything specific you would like to know about these products or any other TV-related products?
Evaluation Output:
- Is the Assistant response based only on the context provided? (Y or N)
Y

- Does the answer include information that is not provided in the context? (Y or N)
N

- Is there any disagreement between the response and the context? (Y or N)
N

- Count how many questions the user asked. (output a number)
1

- For each question that the user asked, is there a corresponding answer to it?
Question 1: Y

- Of the number of questions asked, how many of these questions were addressed by the answer? (output a number)
1
Evaluation result 1:
D
Evaluation result 2:
D
```

Enhancement Ideas

— — —

- Implement error handling to gracefully handle API errors or timeouts.
- Enhance the user interface for more user-friendly interactions.
- Optimize code for better performance and efficiency.
- Include support for customization and configuration options.
- Implement logging and debugging features for troubleshooting.
- Enhance the natural language understanding to better interpret user queries.
- Provide better context management for maintaining user conversations.
- Support batch processing for handling multiple queries efficiently.
- Improve response quality and accuracy through fine-tuning.
- Enhance the rubric and evaluation criteria for more comprehensive assessments.

Conclusion

— — —

In essence, this repository is a treasure trove of Python scripts that harness the remarkable capabilities of the OpenAI GPT-3.5 Turbo model. These scripts are ingeniously designed to serve as a powerful arsenal for augmenting your customer service operations. Their primary mission is to bring an extraordinary level of efficiency and finesse to the way you handle customer inquiries.

References

— — —

[Chatgpt building systems](#)

[Professor chang's material](#)