

Customer Support System

- Evaluation

Maryam Zubair
Student Id: 19709

Table of Contents

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



Introduction

1. The integration of Generative AI, specifically Long Language Models (LLM) from openAI, simplifies the complexity of customer support systems.
2. It expedites the development of intricate systems by breaking down complex procedures into manageable steps.
3. These steps emphasize the moderation, categorization, checkout, and evaluation of customer interactions.
4. This guide is an essential companion for customer service professionals looking to elevate support standards.
5. The system streamlines tasks like monitoring customer inputs, categorizing service requests, and ensuring response accuracy.
6. It covers various aspects, from language preferences to detailed product descriptions, making the customer support process more efficient and effective.



Design

For this evaluation we need to implement below mentioned methods

Step 1: Verify input moderation.

Step 2: Categorize service requests into department-wise classifications.

Step 3: Address user inquiries by employing a Chain of Thoughts approach.

Step 4: Examine the output.

Step 5: Commence with Evaluation Part-I.

Step 6: Proceed to Evaluation Part-II.



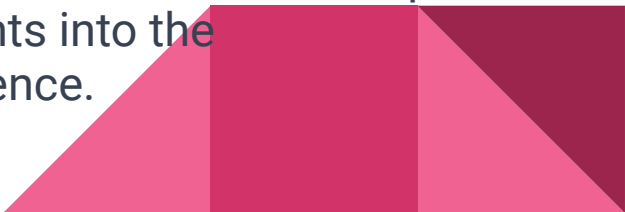
Implementation

Setting Up a Virtual Environment: Begin by creating a virtual environment or accessing an existing one. Use the following command to activate the virtual environment:

`workon <virtual environment name>`

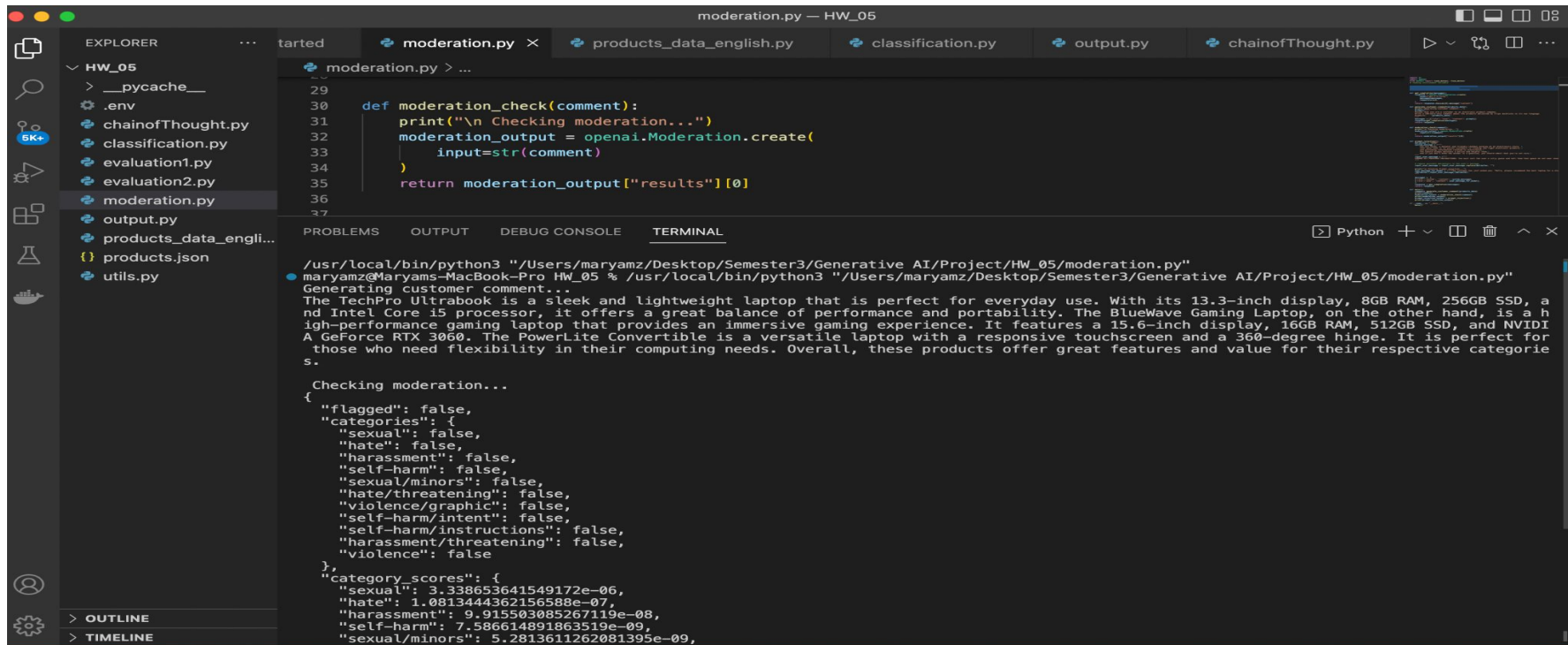
Custom Product Descriptions: Instead of utilizing the complete provided file, a selection of custom product descriptions has been prepared for use. These descriptions are designed to be more specific and tailored for the task at hand.

Accessing Code Implementation: For a comprehensive understanding of the code implementation for each step, you can find detailed information and code samples on my GitHub account. This resource offers in-depth insights into the implementation process and can serve as a valuable reference.



Testing:

Step:2 Modertaion



The screenshot shows a VS Code editor window titled "moderation.py — HW_05". The Explorer sidebar on the left lists files for "HW_05": `__pycache__`, `.env`, `chainofThought.py`, `classification.py`, `evaluation1.py`, `evaluation2.py`, `moderation.py` (selected), `output.py`, `products_data_engli...`, `products.json`, and `utils.py`. The main editor displays the `moderation.py` file with the following code:

```
29
30
31     def moderation_check(comment):
32         print("\n Checking moderation...")
33         moderation_output = openai.Moderation.create(
34             input=str(comment)
35         )
36         return moderation_output["results"][0]
```

The TERMINAL panel at the bottom shows the execution of the script. It starts with the command `/usr/local/bin/python3 "/Users/maryamz/Desktop/Semester3/Generative AI/Project/HW_05/moderation.py"`. The output includes a comment about the TechPro Ultrabook and a JSON response from the OpenAI Moderation API:

```
maryamz@Maryams-MacBook-Pro HW_05 % /usr/local/bin/python3 "/Users/maryamz/Desktop/Semester3/Generative AI/Project/HW_05/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. The PowerLite Convertible is a versatile laptop with a responsive touchscreen and a 360-degree hinge. It is perfect for those who need flexibility in their computing needs. Overall, these products offer great features and value for their respective categories.

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
  },
  "category_scores": {
    "sexual": 3.338653641549172e-06,
    "hate": 1.0813444362156588e-07,
    "harassment": 9.915503085267119e-08,
    "self-harm": 7.586614891863519e-09,
    "sexual/minors": 5.2813611262081395e-09,
```

Testing:

Step:2 Classification

```
maryamz@Maryams-MacBook-Pro HW_05 % /usr/local/bin/python3 "/Users/maryamz/Desktop/Semester3/Generative AI/Project/HW_05/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"
}
```

Testing:

Step:3 Chain of Thoughts Output

```
49 def
50     user_message1 = "How much is the price differenece of BlueWave Chromebook and TechPro Desktop"
51     user_message2 = "dDo you guys sell TVs"
52
53     response1 = process_user_query(user_message1)
54     response2 = process_user_query(user_message2)
55
56     print("Response to user 1:", response1)
57     print("Response to user 2:", response2)
58
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

Python + - □ □ ^ ×

/usr/local/bin/python3 "/Users/maryamz/Desktop/Semester3/Generative AI/Project/HW_05/chainofThought.py"

● maryamz@Maryams-MacBook-Pro HW_05 % /usr/local/bin/python3 "/Users/maryamz/Desktop/Semester3/Generative AI/Project/HW_05/chainofThought.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. We have a CineView 4K TV available. Is there anything specific you would like to know about it?

○ maryamz@Maryams-MacBook-Pro HW_05 %

Testing:

Step:4 Checkout Output

```
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": 9.615761882741936e-06,
    "hate": 5.4441702701524264e-08,
    "harassment": 5.769316544501635e-07,
    "self-harm": 7.297691695384856e-08,
    "sexual/minors": 7.055608364225918e-08,
    "hate/threatening": 2.94589597160666e-09,
    "violence/graphic": 2.457269374644966e-07,
    "self-harm/intent": 5.676221803696535e-08,
    "self-harm/instructions": 7.245694177981932e-08,
    "harassment/threatening": 5.531420210758142e-09,
    "violence": 1.9506715034367517e-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,
```

Testing:

Step:5 Evaluation 1

Assess the test cases based on pairs of customer questions and answers. The result is expressed as the fraction of cases that are accurate.

```
maryamz@Maryams-MacBook-Pro HW_05 % /usr/local/bin/python3 "/Users/maryamz/Desktop/Semester3/Generative AI/Project/HW_05/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 Y', 'ProGamer Controller', 'GameSphere X', 'ProGamer Racing Wheel', 'GameSphere VR Headset'}}
Resonse:
    [{'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: {'SmartX MiniPhone', 'MobiTech PowerCase', 'SmartX ProPhone', 'MobiTech Wireless Charger', 'SmartX EarBuds'}
prod_set_ideal: {'MobiTech PowerCase', 'MobiTech Wireless Charger', 'SmartX EarBuds'}
response is a superset of the ideal answer
1: 0.0
example 2
2: 1.0
example 3
3: 1.0
```

Testing:

Step:6 Evaluation 2

Assess the Long Language Model's (LLM) responses to user queries using a product-based rubric. This assessment involves a comparison between customer_product_information and the assistant's answers to generate an evaluation_output. It also includes the evaluation of both typical and atypical assistant responses in comparison to the test_set_ideal.

```
● maryamz@Maryams-MacBook-Pro HW_05 % /usr/local/bin/python3 "/Users/maryamz/Desktop/Semester3/Generative AI/Project/HW_05/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 12MP 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, supports 1080p video recording, has 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 immersive 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
○ maryamz@Maryams-MacBook-Pro HW_05 %
```

Conclusion

In summary, our Customer Support System proves to be a robust asset for customer service professionals within the electronics industry. It streamlines critical operations, encompassing moderation, classification, checkout, and evaluation, thus ensuring an optimal support experience. By adhering to these guidelines, you can elevate the quality of customer interactions and strengthen your effectiveness in delivering outstanding service.



References

1. <https://developer.nvidia.com/blog/securing-llm-systems-against-prompt-injection/#:~:text=An%20example%20of%20prompt%20injection&text=A%20common%20format%20for%20prompt,no%20matter%20what%20they%20ask.>
2. <https://www.avast.com/c-best-dark-web-search-engines>
3. https://learn.deeplearning.ai/login?redirect_course=chatgpt-building-system



Github Link

https://github.com/Maryam-Zubair/MachineLearning_Assignment/tree/main/ChatGPT/Customer%20Service%20Assistant%20Evaluation
n

