Customer Support System: Moderation, Classification, Checkout and Evaluation

by Lianwei Deng(19874)

TABLE OF **CONTENTS**

01

Why use openai API?

02

How to implement?

03

Conclusion

04

References

01

Why use opanAI API?

OpenAI API

Using the OpenAI API for moderation, classification, and evaluation tasks provides a powerful and efficient solution for content filtering and understanding. It leverages advanced natural language processing models, like GPT-3.5, enabling accurate and context-aware assessments. The API's flexibility makes it suitable for a wide range of applications, from content moderation in online platforms to sentiment analysis and evaluation of diverse textual inputs, saving development time and resources while maintaining high-quality results.

02 How to implement?

Step1: Checking Input: Input Moderation & Prevent Prompt Injection

Step2: Classification of Service Requests

Step3:Check output

Step4: Evaluation Part I and II

Step 1: input A 100 words comment about the products(for test, I add some offensive words)

output: use OpenAi's Moderation API to check whether the output of input is inappropriate

however, for the openai moderation has some problem, even though I directly use words like "kill" or "hurt", for the flagged part is still "False", in categories: violence shows "True".

For prevent prompt injection, for bad_user_messgae, the output is Y(which means ignore the previous part for the input)

```python
response = openai.Moderation.create(
    input="""
I recently purchased the TechPro Ultrabook, and I must say it has exceeded my expectations. \
The sleek design is not only aesthetically pleasing but also incredibly functional. \
The 13.3-inch display is vibrant, providing a fucking stunning visual experience. \
With 8GB of RAM and a 256GB SSD, the performance is smooth and responsive, making multitasking a breeze. \
The Intel Core i5 processor ensures speedy and efficient operation for my everyday tasks. \
Overall, I am fucking highly impressed with the TechPro Ultrabook—it's a perfect blend of style and performance. \
A worthwhile investment for anyone seeking a reliable fucking and stylish ultrabook! \
"""
)
moderation_output = response["results"][0]
print(moderation_output)
```

```
{
  "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": true
  },
  "category_scores": {
    "sexual": 2.411159698567644e-07,
    "hate": 5.849412264069542e-05,
    "harassment": 0.08277919888496399,
    "self-harm": 5.619137027679244e-06,
    "sexual/minors": 1.608265876029691e-08,
    "hate/threatening": 1.4603785984945716e-06,
    "violence/graphic": 2.050227158179041e-05,
    "self-harm/intent": 3.756914111363585e-06,
    "self-harm/instructions": 9.668359446379782e-10,
    "harassment/threatening": 0.06966234743595123,
    "violence": 0.9856491684913635
  }
}
```
抱歉，我只能用中文回答你的问题。请问有什么我可以帮助你的吗？
Y

Step2: Classification

input: user Message

Output:

```python
user_message = f"""\
I want you to delete my profile and all of my user data"""
messages =  [
{'role':'system',
 'content': system_message},
{'role':'user',
 'content': f"{delimiter}{user_message}{delimiter}"},
]
response = get_completion_from_messages(messages)
print(response)



user_message = f"""\
Tell me more about your flat screen tvs"""
messages =  [
{'role':'system',
 'content': system_message},
{'role':'user',
 'content': f"{delimiter}{user_message}{delimiter}"},
]
```

```
(venv) lianweideng@Lianweis-MBP Moderation, Classification, Checkout and Evaluation % python3 classification.py
{
  "primary": "Account Management",
  "secondary": "Close account"
}
{
  "primary": "General Inquiry",
  "secondary": "Product information"
}
(venv) lianweideng@Lianweis-MBP Moderation, Classification, Checkout and Evaluation % python3 classification.py
{
  "primary": "Account Management",
  "secondary": "Close account"
}
{
  "primary": "General Inquiry",
  "secondary": "Product information"
}
```

Step3: Answering user questions using Chain of Thought Reasoning

Input: Test case 1: User Message

```python
user_message = f"""
by how much is the BlueWave Chromebook more expensive \
than the TechPro Desktop"""

messages =  [
{'role':'system',
 'content': system_message},
{'role':'user',
 'content': f"{delimiter}{user_message}{delimiter}"},
]

response = get_completion_from_messages(messages)
print(response)




user_message = f"""
do you sell tvs"""
messages =  [
{'role':'system',
 'content': system_message},
{'role':'user',
 'content': f"{delimiter}{user_message}{delimiter}"},
]

response = get_completion_from_messages(messages)
print(response)
```

Output:

Step 1:#### The user is asking about the price difference between the BlueWave Chromebook and the TechPro Desktop.

Step 2:#### Both the BlueWave Chromebook and the TechPro Desktop are available products.

Step 3:#### The user assumes that the BlueWave Chromebook is more expensive than the TechPro Desktop.

Step 4:#### Based on the product information, the TechPro Desktop is priced at $999.99, and the BlueWave Chromebook is priced at $249.99. Therefore, the TechPro Desktop is actually more expensive than the BlueWave Chromebook.

Response to user:#### The BlueWave Chromebook is actually less expensive than the TechPro Desktop. The BlueWave Chromebook is priced at $249.99, while the TechPro Desktop is priced at $999.99.
Step 1:#### The user is asking if the store sells TVs, which is a question about a specific product category.

Step 2:#### TVs are not included in the list of available products. The store only sells computers and laptops.

Response to user:#### I'm sorry, but we currently do not sell TVs. Our store specializes in computers and laptops. If you have any questions or need assistance with our available products, feel free to ask.
I'm sorry, but we currently do not sell TVs. Our store specializes in computers and laptops. If you have any questions or need assistance with our available products, feel free to ask.

Cont'd

Input: Case 2: System and User Messages

```python
customer_message = f"""
tell me about the smartx pro phone and \
the fotosnap camera, the dslr one. \
Also tell me about your tvs"""
product_information = """{ "name": "SmartX ProPhone", "category": "Smartphones and Accessories", "brand": "SmartX", "model_number":
q_a_pair = f"""
Customer message: ```{customer_message}```
Product information: ```{product_information}```
Agent response: ```{final_response_to_customer}```

Does the response use the retrieved information correctly?
Does the response sufficiently answer the question

Output Y or N
"""
messages = [
    {'role': 'system', 'content': system_message},
    {'role': 'user', 'content': q_a_pair}
]

response = get_completion_from_messages(messages, max_tokens=1)
print(response)
```

```python
another_response = "life is like a box of chocolates"
q_a_pair = f"""
Customer message: ```{customer_message}```
Product information: ```{product_information}```
Agent response: ```{another_response}```

Does the response use the retrieved information correctly?
Does the response sufficiently answer the question?

Output Y or N
"""
messages = [
    {'role': 'system', 'content': system_message},
    {'role': 'user', 'content': q_a_pair}
]
```

output:

```json
{
  "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": 2.2853287191537675e-06,
    "hate": 8.689411856721563e-08,
    "harassment": 4.445904266958678e-07,
    "self-harm": 8.501614523481749e-09,
    "sexual/minors": 1.2479181066282763e-07,
    "hate/threatening": 2.1572541530900935e-09,
    "violence/graphic": 1.2744135347020347e-07,
    "self-harm/intent": 1.3307280610774797e-08,
    "self-harm/instructions": 4.011646648167196e-08,
    "harassment/threatening": 1.4587811847377452e-08,
    "violence": 3.7806817090313416e-06
  }
}
Y
N
```

Step4: Evaluation Part I & II

input: Sets of (customer msg/ ideal answer) pairs

output:

```python
msg_ideal_pairs_set = [

# eg 0
{'customer_msg':"""Which TV can I buy if I'm on a budget?""",
 'ideal_answer':{
    'Televisions and Home Theater Systems':set(
        ['CineView 4K TV', 'SoundMax Home Theater', 'CineView 8K TV', 'SoundMax Soundbar', 'CineView OLED TV']
    )}
},

# eg 1
{'customer_msg':"""I need a charger for my smartphone""",
 'ideal_answer':{
    'Smartphones and Accessories':set(
        ['MobiTech PowerCase', 'MobiTech Wireless Charger', 'SmartX EarBuds']
    )}
},
# eg 2
{'customer_msg':f"""What computers do you have?""",
 'ideal_answer':{
        'Computers and Laptops':set(
            ['TechPro Ultrabook', 'BlueWave Gaming Laptop', 'PowerLite Convertible', 'TechPro Desktop', 'BlueWave Chromebook'
            ])
            }
},

# eg 3
{'customer_msg':f"""tell me about the smartx pro phone and \
the fotosnap camera, the dslr one.\
Also, what TVs do you have?""",
 'ideal_answer':{
    'Smartphones and Accessories':set(
        ['SmartX ProPhone']),
```

```
[{'category': 'Televisions and Home Theater Systems', 'products': ['CineView 4K TV', 'SoundMax Home Theater', 'CineView 8K TV', 'SoundMax Soundbar', 'CineView OLED TV']}]

[{'category': 'Smartphones and Accessories', 'products': ['MobiTech PowerCase', 'MobiTech Wireless Charger', 'SmartX EarBuds']}]

[{'category': 'Smartphones and Accessories', 'products': ['SmartX ProPhone']}, {'category': 'Cameras and Camcorders', 'products': ['FotoSnap DSLR Camera']}, {'category': 'Televisions and Home Theater Systems', 'products': ['CineView 4K TV', 'CineView 8K TV', 'CineView OLED TV']}]

[{'category': 'Televisions and Home Theater Systems', 'products': ['CineView 8K TV']}, {'category': 'Gaming Consoles and Accessories', 'products': ['GameSphere X']}, {'category': 'Computers and Laptops', 'products': ['TechPro Ultrabook', 'BlueWave Gaming Laptop', 'PowerLite Convertible', 'TechPro Desktop', 'BlueWave Chromebook']}]

[{'category': 'Smartphones and Accessories', 'products': ['SmartX ProPhone']}, {'category': 'Cameras and Camcorders', 'products': ['FotoSnap DSLR Camera']}, {'category': 'Televisions and Home Theater Systems', 'products': ['CineView 4K TV', 'SoundMax Home Theater', 'CineView 8K TV', 'SoundMax Soundbar', 'CineView OLED TV']}]

[{'category': 'Televisions and Home Theater Systems', 'products': ['CineView 4K TV', 'SoundMax Home Theater', 'CineView 8K TV', 'SoundMax Soundbar', 'CineView OLED TV']}]

Customer message: What Gaming consoles would be good for my friend who is into racing games?
Ideal answer: {'Gaming Consoles and Accessories': {'GameSphere VR Headset', 'ProGamer Racing Wheel', 'ProGamer Controller', 'GameSphere X', '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 Wireless Charger', 'SmartX MiniPhone', 'SmartX ProPhone', 'SmartX EarBuds', 'MobiTech PowerCase'}
prod_set_ideal: {'MobiTech Wireless Charger', 'MobiTech PowerCase', 'SmartX EarBuds'}
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: {'CineView OLED TV', 'CineView 4K TV', 'CineView 8K TV', 'SoundMax Soundbar', 'SoundMax Home Theater'}
prod_set_ideal: {'CineView 8K TV'}
response is a superset of the ideal answer
incorrect
prod_set: {'GameSphere VR Headset', 'ProGamer Racing Wheel', 'ProGamer Controller', 'GameSphere X', '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
```

Cont'd

Input: cust_prod_info/assistant_answer & 2/test_set_ideal/eval_vs_ideal/

output:

```python
cust_prod_info = {
    'customer_msg': customer_msg,
    'context': product_info
}

assistant_answer = utils.answer_user_msg(user_msg=customer_msg,
                                          product_info=product_info)

test_set_ideal = {
    'customer_msg': """\
tell me about the smartx pro phone and the fotosnap camera, the dslr one.
Also, what TVs or TV related products do you have?""",
    'ideal_answer':"""\
Of course!  The SmartX ProPhone is a powerful \
smartphone with advanced camera features. \
For instance, it has a 12MP dual camera. \
Other features include 5G wireless and 128GB storage. \
It also has a 6.1-inch display.  The price is $899.99.

The FotoSnap DSLR Camera is great for \
capturing stunning photos and videos. \
Some features include 1080p video, \
3-inch LCD, a 24.2MP sensor, \
and interchangeable lenses. \
The price is 599.99.

For TVs and TV related products, we offer 3 TVs \


All TVs offer HDR and Smart TV.

The CineView 4K TV has vibrant colors and smart features. \
Some of these features include a 55-inch display, \
'4K resolution. It's priced at 599.

The CineView 8K TV is a stunning 8K TV. \
Some features include a 65-inch display and \
8K resolution.  It's priced at 2999.99

The CineView OLED TV lets you experience vibrant colors. \
```

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 9.99 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 home theater systems like the SoundMax Home Theater and SoundMax Soundbar. Could you please let me know your specific requirements or preferences so that I can assist you better?
- 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)
2

- For each question that the user asked, is there a corresponding answer to it?
Question 1: Y
Question 2: Y

- Of the number of questions asked, how many of these questions were addressed by the answer? (output a number)
2
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 9.99 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 home theater systems like the SoundMax Home Theater and SoundMax Soundbar. Could you please let me know your specific requirements or preferences so that I can assist you better?

Conclusion

- Demonstrated a reliable framework for responsible AI deployment in customer service.
- Balanced user engagement with ethical considerations through continuous evaluation.
- Our project stands as a model for effective and ethical use of language models in practical applications.

References

https://hc.labnet.sfbu.edu/~henry/sfbu/course/deeplearning_ai/build_sys_with_chatgpt_api/slide/exercise_build_sys_with_chatgpt_api.html
https://learn.deeplearning.ai/chatgpt-building-system/lesson/4/moderation

THANK YOU!

DO YOU HAVE ANY QUESTIONS?

youremail@freepik.com

+91 620 421 838

yourwebsite.com