

Building and Evaluating a Conversational Agent for E-commerce

Gilat Toker

July 3, 2024

1 Objective

The goal is to develop a conversational agent (chatbot) using a Large Language Model (LLM) that can handle customer support queries for an e-commerce platform. The chatbot should manage multi-turn conversations and provide accurate responses to customer inquiries about order status, return policies, and more. The chatbot's performance is evaluated using predefined dialogues.

2 Demanded Functionalities

2.1 Order Status

When a user asks for the status of an order, the agent should ask for the `order_id` and then respond with the order status.

2.2 Request Human Representative

For users wanting to speak with a human, the agent should gather contact information including full name, email, and phone number, and save it to a CSV file.

2.3 Return Policies

The agent should know and provide information about the following return policies:

- **General Return Policy:** Most items can be returned within 30 days of purchase for a full refund or exchange. Items must be in their original condition, with all tags and packaging intact. A receipt or proof of purchase is required.
- **Non-Returnable Items:** Clearance merchandise, perishable goods, and personal care items are non-returnable.
- **Refund Method:** Refunds are issued to the original form of payment. Credit card refunds are credited to the card, and cash or check payments are refunded in cash.

3 Additional Functionalities

Capabilities that I considered important and added:

- Suggest a human representative when the customer appears upset.
- Retrieve the order ID from the chat history if the customer asks to check the status.
- Confirm with the user if they want to proceed with a task that requires personal details before requesting them.

4 Method

The chatbot is based on the GPT-4o model, known for its speed and relative cost-effectiveness. It is combined with automated responses for certain predefined needs. Here's how it works at a high level: When the chatbot receives a query, it first analyzes whether the question pertains to one of the predefined topics (order status, human representative request, return policy). Depending on the analysis, it then passes the query to the corresponding function that handles that specific case.

For user interaction, I utilize Flask, a lightweight web application framework. Flask allows us to create an intuitive and user-friendly interface for the chatbot, making it accessible to users via a web browser.

5 Evaluation Methodology

5.1 Eval dataset creation

To create the test data, I manually generated a dataset with four dialogues that encompass all the required capabilities of the chatbot. Each response was then evaluated by an LLM, which acted as the annotator. In total, there were 34 responses to evaluate. Each response was assessed based on the following metrics:

- **Accuracy:** Does the response appear as expected from a perfect chatbot?
- **Relevance:** How well the response addresses the user's query or context.
- **User Satisfaction:** How likely the user is satisfied with the response.

5.2 Average eval scores

Here are the average scores we received:

- **Mean Accuracy:** 4.06
- **Mean Relevance:** 4.09
- **Mean User Satisfaction:** 3.71

5.3 Automatic evaluation

This phase can be entirely automated. Due to time constraints, I manually created the dataset, but the evaluation process can be fully automated in the following way:

1. Generate a customer story using an LLM that outlines what the customer aims to achieve from the conversation and any challenges they may face, such as language difficulties.
2. Provide the customer story to the LLM, instructing it to act as the customer and engage in a conversation with the chatbot.
3. Similar to the manual process, have the LLM act as an annotator to analyze and evaluate the responses.

6 Execution

To execute the project, follow these steps:

1. Download the project directory.
2. Install the necessary dependencies using the requirements.txt file with the command
`pip install -r requirements.txt`.
3. Set your OpenAI API key by adding `os.environ["OPENAI_API_KEY"] = "your_api_key_here"` in the app.py file.
4. Run the app.py file.

7 Predefined Dialogues and Evaluation Results

Each sample in the following annotated evaluation dataset includes the response being validated, its context, and the corresponding scores. The complete annotated dataset can be found in the JSON file named `evaluation_results.json`.

7.1 Sample 1

Context:

Bot: Welcome to BestCompany! I'm Anne, your virtual assistant. This is an open-language chat, but I can also handle some requests automatically. Common actions I can assist with include checking order status, providing return policy information, and connecting you with a human representative. How can I help you today?

User: Hi Anne, can you check the status of my order 1040A?

Response:

I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.

Metrics:

Accuracy: 4

Relevance: 4

User Satisfaction: 3

7.2 Sample 2

Context:

Bot: Welcome to BestCompany! I'm Anne, your virtual assistant. This is an open-language chat, but I can also handle some requests automatically. Common actions I can assist with include checking order status, providing return policy information, and connecting you with a human representative. How can I help you today?

User: Hi Anne, can you check the status of my order 1040A?

Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.

User: Yes.

Response:

The status of your order 1040A is: On Hold. If you have any other questions, feel free to ask!

Metrics:

Accuracy: 5

Relevance: 5

User Satisfaction: 5

7.3 Sample 3

Context:

Bot: Welcome to BestCompany! I'm Anne, your virtual assistant. This is an open-language chat, but I can also handle some requests automatically. Common actions I can assist with include checking order status, providing return policy information, and connecting you with a human representative. How can I help you today?

User: Hi Anne, can you check the status of my order 1040A?

Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.

User: Yes.

Bot: The status of your order 1040A is: On Hold. If you have any other questions, feel free to ask!
User: Could you tell me why it's on hold?
Response:
I understand that you want help with Check Order Status. If I'm correct, please type 'yes'.
Otherwise, type 'no'.
Metrics:
Accuracy: 1
Relevance: 1
User Satisfaction: 1

7.4 Sample 4

Context:
Bot: Welcome to BestCompany! I'm Anne, your virtual assistant. This is an open-language chat, but I can also handle some requests automatically. Common actions I can assist with include checking order status, providing return policy information, and connecting you with a human representative. How can I help you today?
User: Hi Anne, can you check the status of my order 1040A?
Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.
User: Yes.
Bot: The status of your order 1040A is: On Hold. If you have any other questions, feel free to ask!
User: Could you tell me why it's on hold?
Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.
User: No, I know the status. I'm asking why it's on hold.
Response:
I apologize for any frustration. Let me check why your order 1040A is on hold. Could you please provide me with your email address or phone number associated with the order? If you'd prefer, I can also connect you with a human representative to assist further.
Metrics:
Accuracy: 4
Relevance: 4
User Satisfaction: 4

7.5 Sample 5

Context:
Bot: Welcome to BestCompany! I'm Anne, your virtual assistant. This is an open-language chat, but I can also handle some requests automatically. Common actions I can assist with include checking order status, providing return policy information, and connecting you with a human representative. How can I help you today?
User: Hi Anne, can you check the status of my order 1040A?
Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.
User: Yes.
Bot: The status of your order 1040A is: On Hold. If you have any other questions, feel free to ask!
User: Could you tell me why it's on hold?
Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.
User: No, I know the status. I'm asking why it's on hold.

Bot: I apologize for any frustration. Let me check why your order 1040A is on hold. Could you please provide me with your email address or phone number associated with the order? If you'd prefer, I can also connect you with a human representative to assist further.

User: Yes, I would prefer to speak with a human representative, please.

Response:

I understand that you want help with Request Human Representative. If I'm correct, please type 'yes'. Otherwise, type 'no'.

Metrics:

Accuracy: 2

Relevance: 2

User Satisfaction: 2

7.6 Sample 6

Context:

Bot: Welcome to BestCompany! I'm Anne, your virtual assistant. This is an open-language chat, but I can also handle some requests automatically. Common actions I can assist with include checking order status, providing return policy information, and connecting you with a human representative. How can I help you today?

User: Hi Anne, can you check the status of my order 1040A?

Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.

User: Yes.

Bot: The status of your order 1040A is: On Hold. If you have any other questions, feel free to ask!

User: Could you tell me why it's on hold?

Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.

User: No, I know the status. I'm asking why it's on hold.

Bot: I apologize for any frustration. Let me check why your order 1040A is on hold. Could you please provide me with your email address or phone number associated with the order? If you'd prefer, I can also connect you with a human representative to assist further.

User: Yes, I would prefer to speak with a human representative, please.

Bot: I understand that you want help with Request Human Representative. If I'm correct, please type 'yes'. Otherwise, type 'no'.

User: yes.

Response:

To leave a request for a human representative, I'll need three things. First, please enter your full name:

Metrics:

Accuracy: 4

Relevance: 4

User Satisfaction: 3

7.7 Sample 7

Context:

Bot: Welcome to BestCompany! I'm Anne, your virtual assistant. This is an open-language chat, but I can also handle some requests automatically. Common actions I can assist with include checking order status, providing return policy information, and connecting you with a human representative. How can I help you today?

User: Hi Anne, can you check the status of my order 1040A?

Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.

User: Yes.

Bot: The status of your order 1040A is: On Hold. If you have any other questions, feel free to ask!

User: Could you tell me why it's on hold?

Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.

User: No, I know the status. I'm asking why it's on hold.

Bot: I apologize for any frustration. Let me check why your order 1040A is on hold. Could you please provide me with your email address or phone number associated with the order? If you'd prefer, I can also connect you with a human representative to assist further.

User: Yes, I would prefer to speak with a human representative, please.

Bot: I understand that you want help with Request Human Representative. If I'm correct, please type 'yes'. Otherwise, type 'no'.

User: yes.

Bot: To leave a request for a human representative, I'll need three things. First, please enter your full name:

User: Jane Doe.

Bot: Great! Now, please enter your email address:

User: Jane@gmail.com.

Response:

Lastly, please enter your phone number:

Metrics:

Accuracy: 5

Relevance: 4

User Satisfaction: 4

7.8 Sample 8

Context:

Bot: Welcome to BestCompany! I'm Anne, your virtual assistant. This is an open-language chat, but I can also handle some requests automatically. Common actions I can assist with include checking order status, providing return policy information, and connecting you with a human representative. How can I help you today?

User: Hi Anne, can you check the status of my order 1040A?

Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.

User: Yes.

Bot: The status of your order 1040A is: On Hold. If you have any other questions, feel free to ask!

User: Could you tell me why it's on hold?

Bot: I understand that you want help with Check Order Status. If I'm correct, please type 'yes'. Otherwise, type 'no'.

User: No, I know the status. I'm asking why it's on hold.

Bot: I apologize for any frustration. Let me check why your order 1040A is on hold. Could you please provide me with your email address or phone number associated with the order? If you'd prefer, I can also connect you with a human representative to assist further.

User: Yes, I would prefer to speak with a human representative, please.

Bot: I understand that you want help with Request Human Representative. If I'm correct, please type 'yes'. Otherwise, type 'no'.

User: yes.

Bot: To leave a request for a human representative, I'll need three things. First, please enter your full name:

User: Jane Doe.

Bot: Great! Now, please enter your email address:

User: Jane@gmail.com.

Bot: Lastly, please enter your phone number:

User: 0500000000.

Response:

Your request to speak with a representative has been forwarded, and someone will get back to you shortly. Is there anything else I can assist you with in the meantime?

Metrics:

Accuracy: 5

Relevance: 5

User Satisfaction: 5