

## Data Science Chatbot Challenge: E-Commerce Expert Assistant

Welcome to the challenge! Your task is to create a chatbot that serves as a smart e-commerce assistant, capable of answering questions about both product details and order information. This challenge will test your ability to think on your feet, solve complex problems independently, and implement advanced chatbot functionalities while balancing performance with cost-efficiency. Feel very free, and encouraged to think out loud. Ask questions. Commit as often as you usually would, working on any standard project.

---

### Challenge Details

#### Datasets

1. **Product Information Dataset:**

- Contains 5000 rows of data about various musical instruments and other products, including:
  - Product title, description, features, ratings, price, and categories.
  - Example: *"Ernie Ball Mondo Slinky Nickelwound Electric Guitar Strings"* (rating: 4.8, price: \$6.99).
- **Use Cases:**
  - Answer user questions related to product specifics
  - Suggest products based on user preferences or needs.

2. **Order Data Dataset:**

- Includes customer order history with details such as:
    - Order date, customer ID, product category, sales, shipping cost, and payment method.
    - Example: *Order ID 37077: "Car Media Players" purchased for \$140 on 2018-01-02.*
  - **Use Cases:**
    - Fetch order details for a specific customer
- 

#### Mock API Integration

The provided mock API functions allow interaction with the **Order Data Dataset**. You must use these functions to:

- Retrieve order details by customer ID.
- Process and return the queried data in a user-friendly format.

---

## Retrieval-Augmented Generation (RAG)

You are expected to use **RAG** to improve the chatbot's ability to provide accurate, dynamic responses. The chatbot should:

- Retrieve relevant information from the **Product Dataset** based on user queries.
  - Integrate the retrieved information with an LLM to generate responses.
- 

## Deliverables

### 1. Time Estimates

- The first task is essential, and often overlooked as an important and difficult quality of advanced engineering. Spend some preliminary drawing board time to break this project into smaller pieces, and estimate how long each will take you. Update the attached CSV with your estimated information as your first commit. Time yourself and update this file as time goes on, with a new column showing actual logged time so it can be compared to the initial guess. Good engineers are bad at this sometimes, because of the notorious unpredictability of unknowns in technical projects. We are looking at your process, and promise we will not be phased by poor estimates.

### 2. Core Functionality:

- Build a chatbot capable of:
  - Answering queries from both datasets.
  - Using the mock API for order data retrieval.
  - Employing RAG to improve accuracy and contextual relevance.
- Demonstrate cost-efficiency by documenting trade-offs between model quality, inference time, and hosting costs.

### 3. Sample Questions and Responses:

- Submit at least **10 sample questions and their chatbot-generated responses** for each dataset:
  - Product Dataset queries.
  - Order Data queries via the mock API.

### 4. Implementation:

- Develop the chatbot in a shared Git repository.
- Ensure it runs in your provided Jupyter Notebook for easy testing.
- Include a readme with detailed instructions for use and a detail description of your process

### 5. Bonus Features:

- **Hosting:** Host the chatbot on a cloud-based interface for interactive testing.
- **Interface:** Create a basic UI for basic testing (local or hosted)

- **Model Comparison:** Experiment with both closed-source models (e.g., GPT-4o mini, Claude 3 Haiku, etc), and open-source (e.g., Mistral, LLaMA) to assess trade-offs.
  - **Fine-Tuning:** Experiment with fine-tuning using a dataset like the ECI dataset for better performance.
  - **Advanced Features:** Include unique or creative functionalities
- 

## Guidelines

- **Cost Efficiency:** Avoid relying solely on expensive, high-quality models. Instead, strive for a balance between performance and cost, demonstrating resourceful use of open-source models or lightweight architectures where appropriate.
  - **Experimentation:** Explore multiple approaches and clearly document the trade-offs in terms of accuracy, latency, and cost.
- 

## Evaluation Criteria

1. **Cost Efficiency (Main Criterion):**
  - Are you demonstrating an ability to balance quality and cost effectively?
  - Are the trade-offs clearly documented and justified?
2. **Accuracy (Main Criterion) :**
  - Are responses correct and relevant to user queries?
3. **Integration:**
  - Does the chatbot effectively utilize datasets and API calls?
  - Are RAG techniques used effectively?
4. **Innovation:**
  - Are advanced techniques like fine-tuning, or custom features implemented?
5. **Submission Quality:**
  - Are all deliverables provided in a clear and professional manner?
  - Is the testing and the sample questions/responses realistic, varied, and aligned with the use cases?
6. **Presentation:**
  - Is the codebase well-documented and easy to understand?
  - Are bonus features implemented to enhance functionality?

## Sample Questions and Expected Answers

### Product Dataset Queries

**User:** "What are the top 5 highly-rated guitar products?"

**Chatbot:** "Here are some of the top-rated guitar products you might love:

- The Ernie Ball Mondo Slinky Nickelwound Electric Guitar Strings is a popular choice with a 4.8-star rating. At just \$6.99, it's a great pick for electric guitar players.
- If you need a reliable stand, the Amazon Basics Adjustable Guitar Folding A-Frame Stand also has a 4.8-star rating and is priced at \$17.75.
- For acoustic players, the D'Addario Guitar Strings - Phosphor Bronze is highly rated at 4.7 stars and costs \$10.99.

Let me know if you'd like more details on any of these or if you'd like suggestions tailored to your specific needs!"

---

**User:** "What's a good product for thin guitar strings?"

**Chatbot:** "The *D'Addario Guitar Strings - Phosphor Bronze Acoustic Guitar Strings* for \$10.99 might be just what you're looking for. They're specifically designed to be compatible with thin strings, offering a warm, balanced tone that's perfect for acoustic guitars. They have a solid rating of 4.7 stars with over 60,000 reviews. Let me know if you'd like more details or if you're interested in other options!"

---

**User:** "Is the BOYA BYM1 Microphone good for a cello?"

**Chatbot:** "The *Boya BY-M1* is an omnidirectional lavalier microphone primarily designed for capturing speech in video recordings. Its omnidirectional pickup pattern captures sound from all directions, making it suitable for interviews, presentations, and general voice recording. While it can record musical instruments, its design and frequency response are optimized for vocals rather than the dynamic range and nuances of musical instruments. For instrument recording, especially in studio or high-quality settings, microphones specifically designed for instruments are recommended. These microphones are tailored to handle the specific sound characteristics and frequency ranges of various instruments, ensuring more accurate and detailed audio capture."

---

## Order Dataset Queries

**User:** "What are the details of my last order?"

**Chatbot:** "Sure, I can help with that! Could you please provide your Customer ID?"

**User:** "37077"

**Chatbot:** "Here's what I found for your most recent order:

You placed it on January 2, 2018, for a *Car Media Player*. The total price was \$140, and the shipping cost was \$4.60.

Is there anything else you'd like to know about this order?"\*

---

**User:** "Fetch 5 most recent high-priority orders."

**Chatbot:** "Here are the 5 most recent high-priority orders I found for you:

1. On December 30, 2018, *Car Pillow & Neck Rest* was ordered for \$231 with a shipping cost of \$14.20. (Customer ID: 36086).
2. Also on December 30, 2018, *Car Seat Covers* were ordered for \$114 with a shipping cost of \$1.10. (Customer ID: 35081).
3. On the same day, *Tyre* was ordered for \$250 with a shipping cost of \$13.30. (Customer ID: 26306).
4. Another order for *Car Pillow & Neck Rest* was placed for \$231 with a shipping cost of \$10.50. (Customer ID: 50454).
5. Finally, on December 30, 2018, *Car & Bike Care* was ordered for \$118 with a shipping cost of \$2.90. (Customer ID: 41577).

Let me know if you'd like more details about any of these orders!"

---

**User:** "What are the details of my most recent order?"

**Chatbot:** "Please provide your Customer ID to retrieve your order details."

**User:** "My Customer ID is 37077."

**Chatbot:** "Your most recent order was placed on 2018-01-02 for 'Car Media Players.' The total sales amount was \$140, with a shipping cost of \$4.6. The order priority is marked as 'Medium.' Would you like further details or help with another order?"

---

**User:** "What is the status of my car body covers?"

**Chatbot:** *"Please provide your Customer ID to check the status of your car body cover order."*

**User:** *"My Customer ID is 41066."*

**Chatbot:** *""You placed an order for 5 'Car Body Covers' on 2018-11-08 with a 'Critical' priority. If you'd like more information on shipping or delivery, I recommend checking your order confirmation or contacting support."*

.....  
**User:** *"What is the status of my cell-phone order?"*

**Chatbot:** *"Please provide your Customer ID to retrieve your cell-phone order."*

**User:** *"My Customer ID is 53639."*

**Chatbot:** *"You have two orders for samsung mobile phones: one placed on 2018-06-14 with another placed on 2018-08-08. Which order would you like to inquire about?"*

**User:** *The more recent one"*

**Chatbot:** *"Your order for one Samsung mobile phone placed on 2018-06-14 has a Medium priority and cost you \$220 plus \$13.8 shipping. Let me know if you'd like more details or need assistance with this order."*

---