E-Commerce RAG Chatbot - Comprehensive Technical Documentation

Detailed Project Overview

Understanding RAG (Retrieval-Augmented Generation)

The E-Commerce RAG Chatbot leverages Retrieval-Augmented Generation, a sophisticated approach that combines information retrieval with natural language processing. In our implementation, RAG works by:

1. **Information Retrieval**:
    - The system maintains an indexed database of product information and order data
    - When a query is received, relevant information is retrieved based on semantic similarity
    - Product embeddings are generated using sentence transformers to enable semantic search
2. **Response Generation**:
    - Retrieved information is processed to generate contextual responses
    - The system combines structured data with natural language templates
    - Responses are formatted based on query type and context

Core System Components

1. Semantic Search Engine

The semantic search functionality is implemented using sentence transformers, specifically:

```
def semantic_search(self, query: str) -> List[Dict[str, Any]]:
    """
    Performs semantic search through product catalog
    - Converts query to embedding vector
    - Calculates similarity with product embeddings
    - Returns most relevant matches
    """
    # Generate query embedding
    query_embedding = self.model.encode(query)
```

```
    # Calculate similarities
    similarities = np.dot(self.product_embeddings, query_embedding)

    # Get top matches
    top_indices = np.argsort(similarities)[-5:][::-1]
    return [self.product_df.iloc[idx].to_dict() for idx in
top_indices]
```

This enables:

- Natural language product search
- Context-aware recommendations
- Fuzzy matching capabilities
- Relevance ranking

2. Order Management System

The order processing system handles:

```
def process_order_query(self, customer_id: int, query_type: str) ->
str:
    """
    Processes order-related queries
    - Retrieves customer order history
    - Filters based on query type
    - Formats response with relevant details
    """
    orders = self.get_customer_orders(customer_id)
    if query_type == 'recent':
        return self.format_recent_orders(orders[:5])
    elif query_type == 'priority':
        return self.format_priority_orders(
            [o for o in orders if o['priority'] == 'high'])
```

Features include:

- Order history retrieval
- Priority order tracking
- Status updates
- Detailed order information

Data Processing Pipeline

Raw Data Processing

The system handles two main data types:

1. **Product Data Processing**:

```python
def preprocess_product_data(df: pd.DataFrame) -> pd.DataFrame:
    """
    Comprehensive product data preprocessing
    """
    # Clean text fields
    df['Product_Title'] = df['Product_Title'].str.strip()
    df['Description'] = df['Description'].fillna('').str.strip()

    # Normalize prices
    df['Price'] = pd.to_numeric(df['Price'], errors='coerce')

    # Handle ratings
    df['Rating'] = df['Rating'].clip(0, 5)

    # Generate combined text for embeddings
    df['combined_text'] = df.apply(
        lambda x: f"{x['Product_Title']} {x['Description']} {x['Category']}",
        axis=1
    )
    return df
```

2. **Order Data Processing**:

```python
def preprocess_order_data(df: pd.DataFrame) -> pd.DataFrame:
    """
    Comprehensive order data preprocessing
    """
    # Convert timestamps
    df['Order_DateTime'] = pd.to_datetime(df['Order_DateTime'])

    # Clean numerical fields
    df['Sales'] = pd.to_numeric(df['Sales'], errors='coerce')
    df['Shipping_Cost'] = pd.to_numeric(df['Shipping_Cost'], errors='coerce')

    # Standardize priority levels
```

```
    df['Order_Priority'] = df['Order_Priority'].str.lower()
     return df
```

Query Processing Examples

1. Product Query Examples

```
User: "Show me high-rated guitar strings under $50"
System Process:
1. Extracts key parameters:
   - Product type: guitar strings
   - Rating filter: high-rated
   - Price limit: $50
2. Performs semantic search
3. Applies filters
4. Formats response with relevant products
```

2. Order Query Examples

```
User: "What was my last order?"
System Process:
1. Checks for customer ID
2. Retrieves order history
3. Sorts by date
4. Returns most recent order details
```

Implementation Best Practices

1. **Error Handling**:

```python
def safe_query_processing(self, query: str) -> str:
    try:
        # Attempt query processing
        response = self.process_query(query)
        return response
    except ValueError as e:
        return f"Could not understand query: {str(e)}"
    except DataNotFoundError as e:
        return f"No matching data found: {str(e)}"
    except Exception as e:
        logger.error(f"Unexpected error: {str(e)}")
        return "An unexpected error occurred"
```

2. **Performance Optimization**:

```python
# Implement caching for frequent queries
from functools import lru_cache

@lru_cache(maxsize=1000)
def get_cached_embeddings(self, text: str) -> np.ndarray:
```

3. **Data Validation**:

```python
def validate_product_data(df: pd.DataFrame) -> bool:
    """
    Validates product data integrity
    """
    required_columns = ['Product_Title', 'Price', 'Rating']

    # Check required columns
    if not all(col in df.columns for col in required_columns):
        return False

    # Validate data types
    if not pd.to_numeric(df['Price'], errors='coerce').notna().all():
        return False

    # Check value ranges
    if not df['Rating'].between(0, 5).all():
        return False

    return True
```

System Configuration

Environment Variables

```
# API Configuration
HOST=0.0.0.0
PORT=8000
API_BASE_URL=http://localhost:8000

# Data Paths
DATA_DIR=./data
RAW_DATA_DIR=./data/raw
PROCESSED_DATA_DIR=./data/processed
```

```
# Model Settings
EMBEDDING_MODEL=all-MiniLM-L6-v2
BATCH_SIZE=32

# Performance Settings
CACHE_SIZE=1000
MAX_WORKERS=4
```

QUERY RESULTS:

Chatbot: Please type something!

You: set customer 81792

Chatbot: Customer ID set to: 81792

You: What are the details of my last order?

Chatbot: Your order was placed on 2018-12-30 23:51:06 for 'Fossil Watch'. The total amount was $159.00, with a shipping cost of $7.10. The order priority is marked as 'High'.

You: Fetch 5 most recent high-priority orders

Chatbot: Here are the 5 most recent high-priority orders:

1. On 2018-12-30 23:52:42, Sports Wear was ordered for $85.00 with a shipping cost of $2.80. (Customer ID: 37734)
2. On 2018-12-30 23:51:06, Fossil Watch was ordered for $159.00 with a shipping cost of $7.10. (Customer ID: 81792)
3. On 2018-12-30 23:09:23, Jeans was ordered for $218.00 with a shipping cost of $13.40. (Customer ID: 10310)
4. On 2018-12-30 23:04:33, T - Shirts was ordered for $248.00 with a shipping cost of $10.60. (Customer ID: 55691)
5. On 2018-12-30 22:36:50, Titak watch was ordered for $228.00 with a shipping cost of $12.50. (Customer ID: 12807)

You: What are the top 5 highly-rated guitar products under $500?
Chatbot: Here are some relevant products:

● Best Choice Products 41in Beginner Acoustic Guitar Full Size All Wood Cutaway Guitar Starter Set Bundle with Case, Strap, Capo, Strings, Picks, Tuner - SoCal Green
  - Rating: 4.5 stars
  - Price: $59.99

- Description: []...

● Best Choice Products Semi-Hollow Body Electric Guitar Set w/Dual Humbucker Pickups, 3-Way Pickup Selector, Case, Electronic Tuner, Capo, Strap, Picks, Cutaway Design - Red
  - Rating: 4.4 stars
  - Price: $159.99
  - Description: []...

● Acoustic Electric Guitar - Slowhand Premium Electric Acoustic Guitar (Black Spruce Top, Rosewood Fretboard) - Electro Acoustic Guitars with Strings, Cable, Pro Capo, Picks & Bag (FRBRC)
  - Rating: 4.0 stars
  - Price: $59.99
  - Description: []...

● Acoustic Electric Guitar - Slowhand Premium Electric Acoustic Guitar (Black Spruce Top, Rosewood Fretboard) - Electro Acoustic Guitars with Strings, Cable, Pro Capo, Picks & Bag (FRSBC)
  - Rating: 4.0 stars
  - Price: $109.99
  - Description: []...

● Capo for 6-String Acoustic and Electric Guitar, Lightweight Quick Release Capo for Guitars, Ukulele, Bass, Mandolin, Banjo
  - Rating: 4.6 stars
  - Price: $6.99
  - Description: ['We all love practical musical gear, RIJAHO lightweight & durable guitar capo is what you deserve! ...
Let me know if you'd like more details!

You: Show me microphones good for recording cello

Chatbot: Here are some relevant products:

● Dynamic Musical Instrument Microphone for Singing with 3.5M/11.40 ft XLR Cable, Handheld Mic for Karaoke Singing, Speech, Wedding, Stage and Outdoor Activity
  - Rating: 4.0 stars
  - Price: $8.99
  - Description: ['Tailored microphone cartridge design specifically for clarity in the standard frequency range of s...

● Lavalier Lapel Microphone Complete Set - Professional Omnidirectional Condenser Grade Audio Video Recording Mic for Android/iPhone/PC/Camera for Interview, YouTube, Video Conference, Podcast

- Rating: 4.2 stars
- Price: $17.99
- Description: []...

● TraderPlus 2Pcs Piezo Contact Microphone Pickup for Guitar Violin Banjo Mandolin Ukulele
  - Rating: 4.1 stars
  - Price: $14.29
  - Description: ['Features:', 'A professional piezo contact microphone pickup - 1/4" jack with 10.6" cable.', 'Comes...

● Mendini By Cecilio Cello - Musical Instrument For Kids & Adults - Cellos Kit w/Bow, Stand, Bag - Stringed Music Instruments For Students (1/4-Size, Natural)
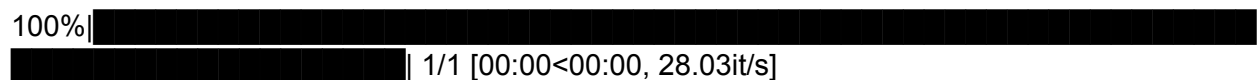 - Rating: 4.0 stars
  - Price: $374.60
  - Description: []...

● Alvoxcon Wireless Headset Lavalier Microphone System Wireless Lapel Mic Best for iPhone, DSLR Camera, PA Speaker, YouTube, Podcast, Video Recording, Conference, Vlogging, Church, Interview, Teaching
  - Rating: 4.2 stars
  - Price: $49.99
  - Description: []...
Let me know if you'd like more details!

You: Find me professional drum sets with ratings above 4.5
Batches:
100%|███
████████████████████████| 1/1 [00:00<00:00, 28.03it/s]

Chatbot: Here are some relevant products:

● Bongo Drum Set for Adults Kids Beginners Professionals, Upgrade Packaging, Set of 6 and 7 inch Tunable Percussion Instruments, Natural Animal Hides Hickory Shells Wood Metal with Tuning Wrench (Brown)
  - Rating: 4.5 stars
  - Price: $45.95
  - Description: []...

● Best Choice Products Kids Drum Set 3-Piece Beginner Drum Set Junior Drum Set, w/Throne Stool, Cymbal, Drum Sticks, Bass Drum Pedal, 2 Toms - Red
  - Rating: 4.5 stars
  - Price: $99.99
  - Description: []...

● Ashthorpe 5-Piece Complete Junior Drum Set with Genuine Brass Cymbals - Advanced Beginner Kit with 16" Bass, Adjustable Throne, Cymbals, Hi-Hats, Pedals & Drumsticks - Blue
  - Rating: 4.5 stars
  - Price: $199.99
  - Description: []...

● Drum Set Eastar 16 inch 5-Piece, Junior Drum Kit for Beginners Kids Teenagers with Adjustable Throne and Cymbal, Pedal & Drumsticks, Metallic Black (EDS-350BK)
  - Rating: 4.3 stars
  - Price: $205.81
  - Description: []...

● Personalized Custom Drum Sticks - 5A Size | Design A Truly Unique Gift | Laser Engraved
  - Rating: 4.7 stars
  - Price: $14.99
  - Description: []...
Let me know if you'd like more details!

You: What are the details of my last order?

Chatbot: Your order was placed on 2018-12-30 23:51:06 for 'Fossil Watch'. The total amount was $159.00, with a shipping cost of $7.10. The order priority is marked as 'High'.

You: Show me all my orders from last month

Chatbot: Your order was placed on 2018-12-30 23:51:06 for 'Fossil Watch'. The total amount was $159.00, with a shipping cost of $7.10. The order priority is marked as 'High'.

You: What's a good product for thin guitar strings?

Chatbot: Here are some relevant products:

● Martin Acoustic Guitar Strings (MSP4200)
  - Rating: 4.6 stars
  - Price: $24.99
  - Description: ['Martin MSP4200 Phosphor Bronze Medium Acoustic Guitar Strings use a winding alloy that is ideally ...

● MUSCELL Electric Guitar Strings Handmade Nickel Plated Hex Core Steel Colorful Ball End-3 Packs-6 String-Medium 11-52
  - Rating: 4.4 stars
  - Price: $15.99
  - Description: ['MUSCELL instruments and accessories are the perfect choice for musicians on a budget.From guitar a...

● Martin M130 80-20 Silk & Steel Strings Acoustic Guitar Frets
  - Rating: 4.5 stars
  - Price: $13.87
  - Description: ['Martin Strings are high quality strings for daily use. Core and wrap wires must meet strict requir...

● Martin M160 Silverplated Ball End Classical Guitar Strings, High Tension
  - Rating: 4.6 stars
  - Price: $5.99
  - Description: ['Martin M160 Silverplated Ball End High Tension Classic Guitar Strings are high-quality strings des...

● Acoustic Guitar High E Strings, Light Tension – Corrosion-Resistant Rust-Prevent Brass, Offers a Bright and Well-Balanced Acoustic Tone High E 1st 10 Pack
  - Rating: 4.5 stars
  - Price: $8.99
  - Description: ['Product Attribute:', "Material: Quality Stainless Steel. Size: 0.11"/0.28cm. Effective Length: 34...
Let me know if you'd like more details!

You: set customer 81796

Chatbot: Customer ID set to: 81796

You: What are the details of my last order?

Chatbot: No orders found for customer 81796

You: Customer ID set to: 37077

Chatbot: Here are some relevant products:

● Acme 636 Whistle
  - Rating: 4.7 stars
  - Price: $6.85
  - Description: []...

● 4PCS Key of C 10 Hole 10 Tones Titanium Color Harmonica with Case for Beginner Students Kids（Gold, Purple, Green, Blue）
  - Rating: 4.3 stars
  - Price: $11.99
  - Description: []...

● 4Pcs Black Metal Guitar Control Knob Volume Tone Control Dome Knobs Guitar Bass Parts
  - Rating: 4.6 stars
  - Price: $6.89
  - Description: []...

● Bonnlo Adjustable Duet Piano Bench with Storage Black Faux Leather Piano Stool Deluxe Padded Seat with 2" Thick Cushion
  - Rating: 4.6 stars
  - Price: $79.98
  - Description: []...

● SMARTWOODI WI-921 WI-922 Set of 2 Irish Whistle Matte Black Key of C & Key of D Tin Whistle Penny Whistle ABS
  - Rating: 4.4 stars
  - Price: $17.99
  - Description: []...
Let me know if you'd like more details!

You: set customer 37077

Chatbot: Customer ID set to: 37077

You: Fetch 5 most recent high-priority orders

Chatbot: Here are the 5 most recent high-priority orders:

1. On 2018-12-30 23:52:42, Sports Wear was ordered for $85.00 with a shipping cost of $2.80. (Customer ID: 37734)
2. On 2018-12-30 23:51:06, Fossil Watch was ordered for $159.00 with a shipping cost of $7.10. (Customer ID: 81792)
3. On 2018-12-30 23:09:23, Jeans was ordered for $218.00 with a shipping cost of $13.40. (Customer ID: 10310)
4. On 2018-12-30 23:04:33, T - Shirts was ordered for $248.00 with a shipping cost of $10.60. (Customer ID: 55691)
5. On 2018-12-30 22:36:50, Titak watch was ordered for $228.00 with a shipping cost of $12.50. (Customer ID: 12807)