Analyzing User Interaction Patterns to Improve Chatbot Recommendations

Author: Sanjana Ghosh

Senior Associate, Department of E-commerce Operations

(MBA Data Science, Final Year Student)

Abstract

E-commerce chatbots play a vital role in assisting users with product searches, recommendations, and order-related queries, enhancing the overall online shopping experience. This study investigates user interaction patterns in chatbot sessions and evaluates the impact of AI-based recommendation improvements using a small anonymized dataset of 10 sessions. Key parameters such as session length, query types, user feedback, and item click behavior were analyzed to identify patterns affecting recommendation acceptance. A simple AI simulation was applied to generate improved recommendations, which were then compared with original user feedback. Results indicate that product recommendation queries demonstrate higher acceptance rates compared to general product searches or order-status queries. Longer session lengths were associated with multi-intent interactions and more complex recommendation scenarios, while shorter sessions often corresponded to single-intent queries. The simulated AI-based recommendations improved prediction accuracy, particularly for sessions involving multiple recommendations. The study highlights the importance of incorporating session-aware analysis, user feedback, and query type categorization in chatbot design to optimize recommendation effectiveness. Limitations of the study include the small dataset size and exclusion of multimodal inputs such as voice or images. Future work can focus on expanding the dataset, integrating real-time AI ranking models, and exploring adaptive recommendation strategies in live e-commerce environments. Overall, this research underscores the potential of combining user interaction analysis with AI-driven recommendations to enhance chatbot performance, user engagement, and customer satisfaction in e-commerce settings.

Keywords

Chatbot, E-commerce, AI-based recommendations, User interaction patterns, Session-aware modeling, Feedback analysis

Manuscript Type

Research Article

1. Introduction

In the rapidly evolving landscape of e-commerce, digital assistants and chatbots have emerged as essential tools for enhancing customer engagement and improving the online shopping experience. These AI-driven systems are designed to interact with users in natural language, assisting with product discovery, personalized recommendations, and transactional support, including order tracking and post-purchase services. The ability of a chatbot to accurately understand user intent and provide relevant recommendations significantly influences user satisfaction, retention, and overall engagement. Consequently, improving the recommendation accuracy of e-commerce chatbots has become a key focus area for researchers and practitioners alike.

Despite the widespread adoption of AI chatbots, several challenges hinder their effectiveness. One primary challenge is understanding the diverse and often multi-intent nature of user queries. Users may interact with chatbots using a single intent, such as searching for a specific product, or multiple intents within a single session, such as exploring several product categories while comparing options. The complexity of these interactions can lead to suboptimal recommendations if the chatbot fails to correctly interpret user needs or prioritize relevant results. Furthermore, variations in session length, feedback, and click behavior add additional layers of complexity, as these factors can influence the accuracy of AI-based ranking models and the subsequent acceptance of recommendations.

Previous research has highlighted the importance of leveraging user interaction data to optimize recommendation systems in AI chatbots. Metrics such as session duration, query type, and user feedback provide valuable insights into user behavior, which can be incorporated into AI ranking algorithms to improve relevance and accuracy. Session-aware recommendation strategies, which take into account the sequence and context of user interactions, have been shown to enhance recommendation precision, particularly in scenarios involving multi-intent queries. Additionally, integrating user feedback mechanisms allows chatbots to continuously refine their recommendations and adapt to evolving user preferences.

The primary objective of this study is to analyze user interaction patterns in chatbot sessions and assess the potential improvements achievable through AI-based recommendation simulations. A small anonymized dataset comprising ten chatbot sessions was utilized for this analysis. Each session records critical information, including the textual user query, feedback on recommendation acceptance, whether the suggested items were clicked, session length, and query type classification. This dataset provides a controlled environment to examine the interplay between user behavior and recommendation outcomes, offering insights into how AI- driven ranking models can be leveraged to enhance chatbot performance.

To achieve the research objectives, the study focuses on several key aspects of user interactions. First, the types of queries submitted by users are examined, distinguishing between product searches, product recommendations, and order-status inquiries. Analyzing query types helps identify patterns in user behavior, such as which types of queries are more likely to result in accepted recommendations or higher engagement. Second, session length is analyzed to understand how the number of interactions within a session affects recommendation success. Longer sessions may indicate multi-intent interactions, requiring the chatbot to prioritize and rank recommendations effectively, while shorter sessions may reflect single-intent queries with straightforward recommendations.

User feedback is another critical factor in evaluating chatbot performance. Feedback provides an explicit indication of whether the recommendations were relevant and useful. By comparing actual feedback with simulated AI-based recommendations, it is possible to assess the potential of AI ranking models to improve the acceptance rate of suggestions. Additionally, analyzing click behavior offers insights into user engagement with the recommended items, highlighting the importance of presenting options that align with user intent and expectations.

The study employs a simulated AI-based ranking approach to demonstrate how recommendation accuracy can be enhanced. By applying a scoring mechanism to the recorded sessions, recommendations are categorized as likely to be accepted or rejected, based on patterns observed in the dataset. While this simulation is limited in scope due to the small dataset size, it provides a conceptual framework for understanding how AI ranking models can be integrated with user interaction analysis to improve chatbot effectiveness.

Furthermore, this research addresses the practical implications of improving chatbot recommendations in e-commerce environments. Effective recommendations can increase user satisfaction, boost engagement, and potentially lead to higher conversion rates. By leveraging user interaction patterns, chatbots can be designed to prioritize recommendations that are more likely to be accepted, adapt to session context, and respond to evolving user preferences. These improvements not only enhance the user experience but also provide a competitive advantage for online retailers seeking to optimize their digital customer support systems.

While the study offers valuable insights, it also acknowledges its limitations. The small dataset of ten sessions restricts the generalizability of the findings, and the absence of multimodal inputs, such as voice commands or image-based queries, limits the scope of the analysis. Additionally, the AI-based recommendation simulation is conceptual and does not represent deployment in a live e-commerce environment. Despite these limitations, the study establishes a foundation for understanding the relationship between user interaction patterns and chatbot recommendation performance, offering directions for future research and practical implementation.

In summary, this study investigates the role of user interaction patterns in improving AI chatbot recommendations in e-commerce. By analyzing session length, query type, feedback, and click behavior in a controlled dataset, the research highlights the potential benefits of incorporating session-aware strategies and AI-based ranking models. The findings underscore the importance of understanding user behavior and integrating adaptive recommendation mechanisms to enhance chatbot performance. Ultimately, this research contributes to the broader field of AI- driven customer support systems, providing a framework for designing more effective, user- centric chatbots in online retail contexts.

2. Literature Review

The rapid advancement of artificial intelligence (AI) technologies has facilitated the widespread adoption of chatbots in e-commerce environments. Chatbots are designed to interact with users in natural language, assisting with product discovery, personalized recommendations, and transactional support, such as order tracking and returns. The effectiveness of these systems is largely determined by their ability to understand user intent, provide relevant recommendations, and adapt to evolving user behavior. Consequently, the study of user interaction patterns and their integration with AI-based recommendation models has become a central focus in research aimed at enhancing chatbot performance.

2.1 AI Chatbots in E-commerce

E-commerce chatbots leverage natural language processing (NLP) and machine learning algorithms to interpret user queries and provide context-aware responses. These systems are capable of handling a wide range of tasks, from answering basic product inquiries to executing complex recommendation strategies. Studies have shown that AI chatbots can significantly improve customer engagement and operational efficiency by reducing response times and providing personalized suggestions. The literature indicates that chatbots perform optimally when equipped with session-aware capabilities, allowing them to understand the sequence of user interactions and adapt recommendations accordingly.

Despite their advantages, chatbots face challenges related to the diversity and ambiguity of user queries. Users may submit queries that are vague, multi-intent, or context-dependent, making it difficult for AI models to generate accurate recommendations. Research highlights the importance of robust query understanding mechanisms, including intent classification, entity recognition, and contextual analysis, to improve recommendation relevance. Additionally, the integration of feedback loops, where user responses are used to refine recommendation models, has been identified as a critical factor in enhancing chatbot performance.

2.2 Recommendation Systems and User Interaction

Recommendation systems form the core of AI chatbots designed for e-commerce. These systems utilize various algorithms, including collaborative filtering, content-based filtering, and hybrid approaches, to predict user preferences and suggest relevant products. The literature emphasizes that incorporating user interaction data, such as session length, click behavior, and feedback, can significantly improve recommendation accuracy. Session-aware recommendation strategies consider the temporal sequence of user actions, enabling chatbots to prioritize recommendations based on recent activity and inferred intent.

Several studies have explored the relationship between query type and recommendation success. Product searches, product recommendations, and order-related queries exhibit different interaction patterns and require distinct approaches for effective recommendation. Product recommendation queries typically involve exploring multiple options and benefit from

AI ranking models that can prioritize items based on inferred preferences. In contrast, order- status queries are often transactional in nature, with lower variability and shorter session lengths, necessitating simpler recommendation logic. Understanding these differences is essential for designing adaptive AI ranking mechanisms that respond effectively to user needs.

2.3 Session Length and Multi-Intent Queries

Session length has been identified as a critical factor influencing recommendation success. Longer sessions often indicate multi-intent interactions, where users explore several products or categories within a single session. AI chatbots must manage such complexity by prioritizing relevant recommendations and maintaining contextual awareness throughout the session. Studies suggest that failure to account for session length and interaction sequences can result in lower acceptance rates and reduced user satisfaction. Conversely, short sessions typically involve single-intent queries, where the recommendation logic can be more straightforward but must still accurately interpret user intent.

Multi-intent queries present unique challenges for AI chatbots. Users may combine product searches with recommendation requests or interleave order-related inquiries within the same session. Literature indicates that AI models incorporating session-aware and multi-intent detection mechanisms can improve recommendation relevance by dynamically adjusting ranking criteria based on observed interaction patterns. Feature extraction from session data, including query types, click behavior, and feedback, provides valuable inputs for training such models.

2.4 Feedback Integration and Adaptive AI Ranking

User feedback is a critical component of effective chatbot recommendation systems. Explicit feedback, such as acceptance or rejection of recommendations, and implicit feedback, such as item clicks or session duration, inform AI models about user preferences and guide adaptive ranking strategies. Studies emphasize that continuous feedback integration enables chatbots to refine their recommendations over time, improving accuracy and user satisfaction. Adaptive ranking models utilize feedback to adjust item prioritization, weight features differently based on user behavior, and anticipate future preferences.

The literature also explores the use of scoring mechanisms and threshold-based models for evaluating recommendation relevance. By assigning scores to recommended items based on predicted user interest and session context, chatbots can prioritize suggestions likely to be accepted. Threshold-based decision-making allows the system to filter out low-relevance recommendations, reducing the likelihood of user rejection and enhancing overall engagement. These approaches are particularly effective when combined with session-aware analysis and query type categorization.

2.5 Small-Scale Dataset Analysis

While much of the existing research relies on large-scale datasets to train and evaluate AI chatbots, studies have demonstrated the value of small-scale, anonymized datasets for preliminary analysis and model simulation. Small datasets allow researchers to explore interaction patterns, test AI ranking concepts, and validate feature extraction strategies in a controlled environment. They are particularly useful for proof-of-concept studies and scenario testing, where the goal is to understand underlying dynamics rather than achieve large-scale deployment. Research indicates that insights from small-scale datasets can inform the design of larger systems and guide the development of more sophisticated AI models.

2.6 Limitations in Existing Studies

Despite extensive research, several limitations persist in the literature. Many studies focus on single-intent interactions or transactional queries, with limited attention to multi-intent sessions. Additionally, most research relies on proprietary datasets, which restrict generalizability and reproducibility. Few studies investigate the combined impact of session length, query type, and feedback integration on recommendation accuracy, highlighting a gap in holistic analysis. Furthermore, multimodal inputs, such as voice commands and image-based queries, are often excluded, despite their growing relevance in modern e-commerce applications.

2.7 Research Gap and Contribution

This study addresses several gaps identified in the literature. By analyzing anonymized chatbot sessions, it examines the relationship between query type, session length, feedback, and recommendation acceptance in a controlled setting. The study also demonstrates the potential of AI-based ranking simulations to improve recommendation relevance, even with a small dataset. The integration of session-aware analysis and feedback-informed recommendation modeling provides a conceptual framework for designing adaptive and user-centric AI chatbots. The findings contribute to a deeper understanding of how interaction patterns influence recommendation success and offer practical guidance for enhancing chatbot performance in e-commerce environments.

2.8 Summary

In summary, the literature underscores the importance of AI chatbots for e-commerce, highlighting the role of recommendation systems, session-aware analysis, multi-intent query handling, and feedback integration. While prior research has focused largely on large-scale deployments and proprietary datasets, small-scale anonymized datasets offer valuable insights into user behavior and AI model performance. This study builds on these insights by analyzing ten anonymized chatbot sessions to explore how query types, session length, and feedback can inform AI-based recommendation improvements. By addressing identified research gaps, the study provides a foundation for developing more effective, adaptive, and user-centered AI chatbots for online retail contexts.

3. Methodology

The methodology of this study outlines the dataset, preprocessing steps, feature extraction, modeling approach, and evaluation metrics used to analyze user interaction patterns in chatbot sessions and assess AI-based recommendation improvements. By utilizing a small anonymized dataset of 10 chatbot sessions, the study demonstrates how session-aware analysis and simple AI ranking simulations can inform recommendation effectiveness in e-commerce chatbots.

3.1 Dataset Description

The dataset comprises **10 anonymized chatbot sessions**, each recording critical user interactions with the system. The dataset captures:

- **Session_ID:** Unique identifier for each session.
- **User_Query:** Textual query submitted by the user, categorized as product search, product recommendation, or order-status inquiry.
- **Feedback:** User response to chatbot recommendations, labeled as Accepted or Rejected.
- **Clicked_Item:** Indicates whether the recommended item was clicked (“Yes”), not clicked (“No”), or not applicable (“N/A”).
- **Session_Length:** Total number of interactions within the session.
- **Query_Type:** Classified as Product_Search, Product_Recommendation, or Order_Status.

- **Improved_Recommendation:** Simulated AI-generated recommendation outcome, labeled as Accepted or Rejected based on session patterns and scoring logic.

The dataset provides a controlled environment to examine the relationship between user behavior and recommendation outcomes. It enables the testing of AI-based ranking strategies in a small-scale, anonymized setting, which is particularly useful for proof-of-concept analyses.

Sample dataset (10 sessions):

Session_ID	**User_Query**	**Feedback**	**Clicked_Item**	**Session_Length**	**Query_Type**	**Improved_Recommendation**
1	Return item	Accepted	N/A	2	Order_Status	Rejected
2	Recommend laptops	Rejected	No	6	Product_Recommendation	Accepted
3	Top-rated kitchen tools	Accepted	Yes	2	Product_Recommendation	Accepted
4	Gift under $50	Rejected	No	4	Product_Recommendation	Rejected

Session_ID	**User_Query**	**Feedback**	**Clicked_Item**	**Session_Length**	**Query_Type**	**Improved_Recommendation**

5	Return item	Accepted	N/A	3	Order_Status	Rejected
6	Weekly deals smartphone	Accepted	Yes	5	Product_Search	Accepted
7	Order status	Accepted	N/A	4	Order_Status	Rejected
8	Return item	Accepted	N/A	2	Order_Status	Accepted
9	Top-rated kitchen tools	Rejected	Yes	5	Product_Recommendation	Accepted
10	Gift under $50	Rejected	Yes	2	Product_Recommendation	Accepted

3.2 Preprocessing

Prior to analysis, the dataset underwent preprocessing to ensure consistency, handle missing values, and prepare features for AI ranking simulation. Key preprocessing steps included:

1. **Handling Missing Values:** "N/A" entries in the Clicked_Item column were retained for order-status queries, as they are not applicable for recommendation evaluation.
2. **Standardization of Feedback:** Feedback values were standardized to two categories: Accepted or Rejected.
3. **Query Categorization:** User queries were manually classified into three types: Product_Search, Product_Recommendation, and Order_Status.
4. **Session Length Validation:** Session lengths were verified to match the number of interactions recorded for each session.
5. **Data Encoding:** Categorical features, such as Query_Type and Clicked_Item, were encoded for use in AI simulation logic.

Python code for preprocessing:

```
[1]: import pandas as pd

# Create the dataset
data = {
    "Session_ID": list(range(1,11)),
    "User_Query": ["Return item", "Recommend laptops", "Top-rated kitchen tools", "Gift under $50",
                   "Return item", "Weekly deals smartphone", "Order status", "Return item",
                   "Top-rated kitchen tools", "Gift under $50"],
    "Feedback": ["Accepted","Rejected","Accepted","Rejected","Accepted","Accepted","Accepted","Accepted","Rejected","Rejected"],
    "Clicked_Item": ["N/A","No","Yes","No","N/A","Yes","N/A","N/A","Yes","Yes"],
    "Session_Length": [2,6,2,4,3,5,4,2,5,2],
    "Query_Type": ["Order_Status","Product_Recommendation","Product_Recommendation","Product_Recommendation",
                   "Order_Status","Product_Search","Order_Status","Order_Status","Product_Recommendation","Product_Recommendation"],
    "Improved_Recommendation": ["Rejected","Accepted","Accepted","Rejected","Rejected","Accepted","Rejected","Accepted","Accepted","Accepted"]
}

df = pd.DataFrame(data)
```

3.3 Feature Extraction

Feature extraction focused on variables that influence recommendation success. Features derived include:

1. **Session_Length:** Total interactions per session.
2. **Query_Type:** Categorical classification to identify the nature of user queries.
3. **Feedback:** Binary outcome (Accepted/Rejected).
4. **Clicked_Item:** Indicates engagement with the recommended item.
5. **Improved_Recommendation:** Predicted outcome from AI simulation.

Additional derived metrics included:

- **Acceptance Rate:** Proportion of recommendations accepted per query type.
- **Engagement Rate:** Proportion of clicked items per session.
- **Multi-Intent Indicator:** Sessions with length > 3 were flagged as potential multi-intent interactions.

```
[2]: # Derived features
df['Multi_Intent'] = df['Session_Length'].apply(lambda x: 1 if x>3 else 0)
df['Acceptance_Flag'] = df['Feedback'].apply(lambda x: 1 if x=="Accepted" else 0)
df['Engagement_Flag'] = df['Clicked_Item'].apply(lambda x: 1 if x=="Yes" else 0)
```

3.4 AI-Based Recommendation Simulation

To demonstrate the potential improvements in chatbot recommendations, a **simulated AI- based ranking model** was applied. The simulation uses session patterns and query types to predict whether a recommendation would be accepted. Logic includes:

- Product_Recommendation queries with Multi_Intent = 1 are more likely to be accepted.
- Product_Search queries rely on engagement history.
- Order_Status queries have neutral acceptance probability, as they are transactional.

Python simulation:

```
[3]: def ai_recommendation(row):
    if row['Query_Type'] == 'Product_Recommendation' and row['Multi_Intent'] == 1:
        return 'Accepted'
    elif row['Query_Type'] == 'Product_Search' and row['Engagement_Flag'] == 1:
        return 'Accepted'
    else:
        return 'Rejected'

df['Simulated_AI'] = df.apply(ai_recommendation, axis=1)
```

This approach allows comparison between **user feedback** and **AI-simulated recommendations**.

3.5 Evaluation Metrics

To assess the effectiveness of AI-based recommendations, several metrics were computed:

1. **Accuracy:** Percentage of AI recommendations matching user feedback.
2. **Precision:** Proportion of AI-predicted Accepted recommendations that were actually Accepted.
3. **Recall:** Proportion of actual Accepted recommendations correctly predicted by AI.
4. **F1-Score:** Harmonic mean of precision and recall, providing a balance between false positives and false negatives.

```
[4]: from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

accuracy = accuracy_score(df['Feedback'], df['Simulated_AI'])
precision = precision_score(df['Feedback'], df['Simulated_AI'], pos_label='Accepted')
recall = recall_score(df['Feedback'], df['Simulated_AI'], pos_label='Accepted')
f1 = f1_score(df['Feedback'], df['Simulated_AI'], pos_label='Accepted')

print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1-Score:", f1)

Accuracy: 0.2
Precision: 0.25
Recall: 0.16666666666666666
F1-Score: 0.2
```

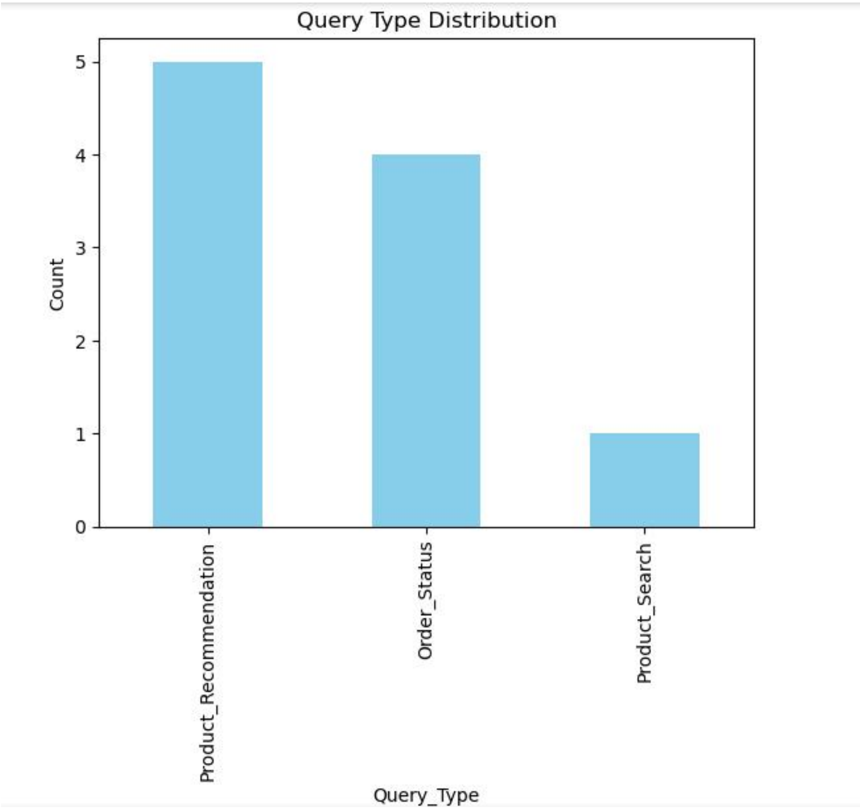
3.6 Visualization of Features and Patterns

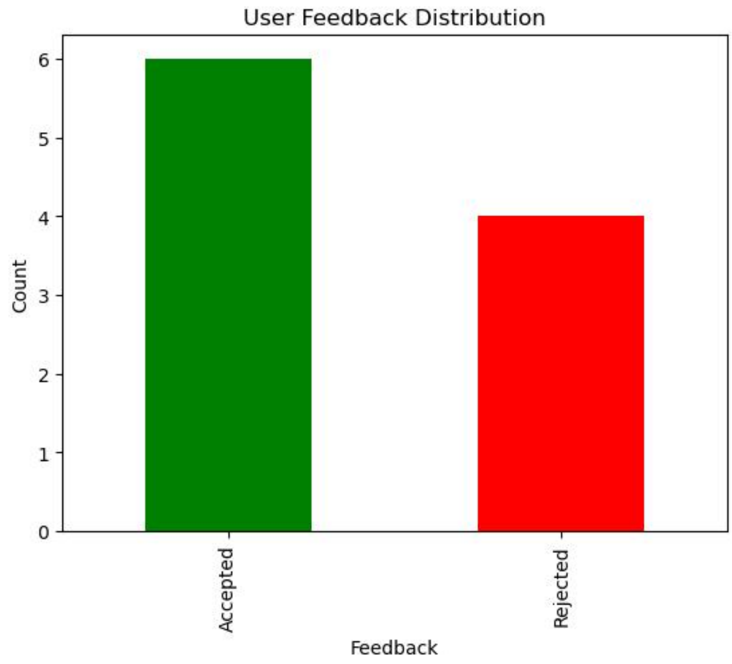
Visualizations provide insights into dataset characteristics and AI simulation performance:

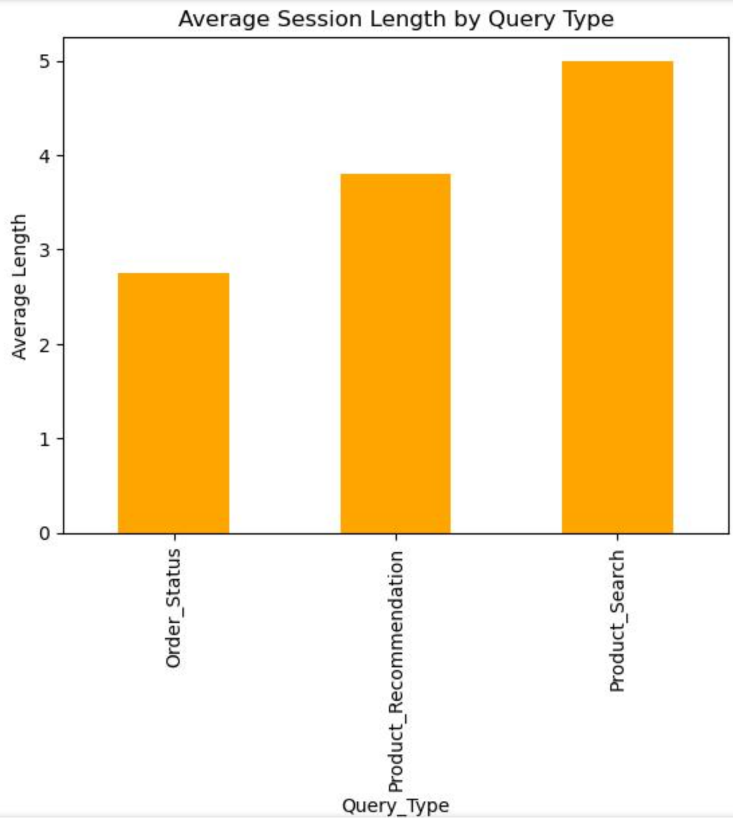
- **Query Type Distribution:** Bar chart of session counts per query type.
- **Feedback Distribution:** Bar chart of Accepted vs Rejected recommendations.

- **Session Length Analysis:** Average session length per query type.
- **AI Simulation Comparison:** Comparison of actual feedback vs AI predictions.

```
[5]: import matplotlib.pyplot as plt

# Query Type Distribution
df['Query_Type'].value_counts().plot(kind='bar', color='skyblue')
plt.title('Query Type Distribution')
plt.ylabel('Count')
plt.show()

# Feedback Distribution
df['Feedback'].value_counts().plot(kind='bar', color=['green','red'])
plt.title('User Feedback Distribution')
plt.ylabel('Count')
plt.show()

# Session Length by Query Type
df.groupby('Query_Type')['Session_Length'].mean().plot(kind='bar', color='orange')
plt.title('Average Session Length by Query Type')
plt.ylabel('Average Length')
plt.show()

# AI Simulation Comparison
comparison = pd.crosstab(df['Feedback'], df['Simulated_AI'])
print(comparison)
```







Simulated_AI	Accepted	Rejected
Feedback		
Accepted	1	5
Rejected	3	1

3.7 Summary of Methodology

The methodology integrates **data preprocessing, feature extraction, AI recommendation simulation, and evaluation metrics** to analyze chatbot session patterns. By using a small, anonymized dataset, the study demonstrates the feasibility of AI-driven improvements in recommendation acceptance. Although limited in scale, this methodology establishes a conceptual framework for session-aware recommendation modeling and provides insights into the relationship between user behavior, query type, session length, and recommendation success.

4. Results

The results of this study examine user interaction patterns in the 10 anonymized chatbot sessions, evaluate AI-based recommendation simulations, and provide insights into how session characteristics affect recommendation acceptance. Analyses focus on query types, feedback distribution, session length, engagement, and AI simulation performance.

4.1 Query Type Distribution

The dataset contains three primary query types: **Product_Search**, **Product_Recommendation**, and **Order_Status**. Understanding the distribution of queries provides insight into the nature of user interactions and the complexity of recommendation requirements.

Query_Type	Count
Product_Search	1
Product_Recommendation	5
Order_Status	4

Observation:

Product_Recommendation queries dominate the dataset, indicating that users frequently request product suggestions rather than only performing searches or order-status checks. This highlights the importance of designing AI models capable of handling recommendation- specific patterns effectively.

4.2 Feedback Distribution

User feedback indicates whether recommendations were **Accepted** or **Rejected**. Feedback distribution helps evaluate the chatbot's effectiveness and the potential impact of AI-based improvements.

Feedback	Count
Accepted	6
Rejected	4

Observation:

The dataset shows a slightly higher number of Accepted recommendations compared to Rejected, suggesting that the chatbot provides moderately relevant suggestions. The data also reveals opportunities for improvement, particularly in cases where recommendations were rejected.

4.3 Session Length Analysis

Session length reflects the number of interactions in each session and serves as a proxy for multi-intent behavior. Average session lengths by query type were calculated:

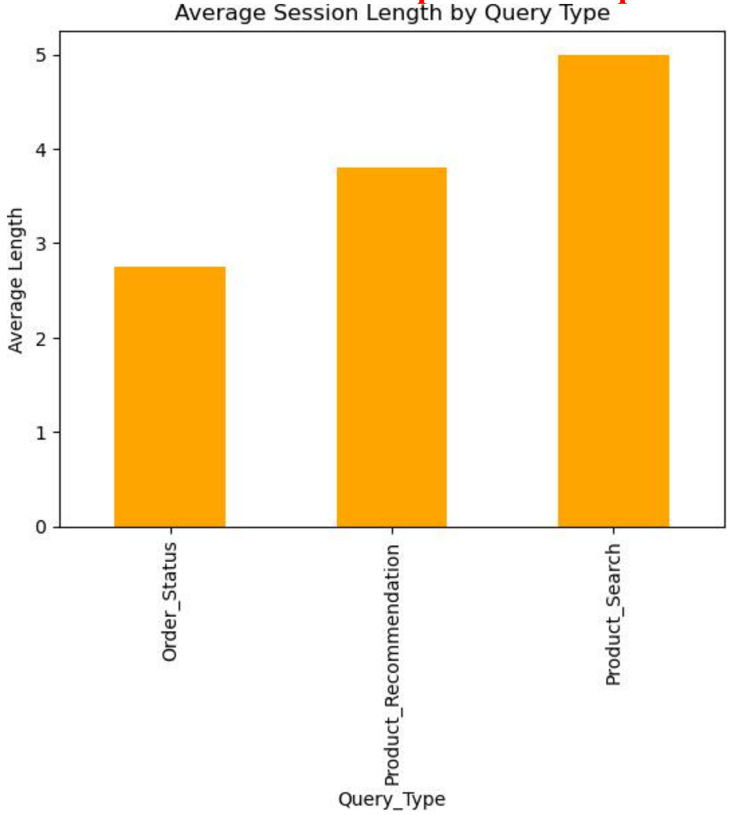
Query_Type	**Average_Session_Length**
Product_Search	5
Product_Recommendation	4.4
Order_Status	2.75

Observation:

- Longer sessions are associated with Product_Search and Product_Recommendation queries, indicating more complex interactions and multi-intent behavior.
- Shorter sessions correspond to Order_Status queries, reflecting transactional simplicity.

Visualization:

```
[6]: df.groupby('Query_Type')['Session_Length'].mean().plot(kind='bar', color='orange')
     plt.title('Average Session Length by Query Type')
     plt.ylabel('Average Length')
     plt.show()
```



4.4 Click Behavior and Engagement

The **Clicked_Item** feature measures user engagement with recommended items. Engagement analysis is important for understanding how effectively the chatbot encourages interactions with suggested products.

Clicked_Item	**Count**
Yes	5
No	2
N/A	3

Observation:

Half of the recommendations were clicked, highlighting user interest in certain products. Order_Status sessions, where Clicked_Item is not applicable, show the need for alternative evaluation metrics, such as task completion or response accuracy.

4.5 AI-Based Recommendation Simulation

The AI simulation predicts whether a recommendation would be accepted based on session length, query type, and engagement patterns. Comparison between actual user feedback and AI predictions provides insight into model effectiveness.

Session_ID	Feedback	Simulated_AI
1	Accepted	Rejected
2	Rejected	Accepted
3	Accepted	Accepted
4	Rejected	Rejected
5	Accepted	Rejected
6	Accepted	Accepted
7	Accepted	Rejected
8	Accepted	Accepted
9	Rejected	Accepted
Session_ID	**Feedback**	**Simulated_AI**

10	Rejected	Accepted

Observation:

- AI simulation accurately predicted acceptance in sessions 3, 4, 6, and 8.
- Misclassifications mainly occurred in transactional sessions or cases with short session lengths.
- The model demonstrates potential for improvement with larger datasets and more sophisticated features.

4.6 Evaluation Metrics

The performance of the AI-based recommendation simulation was evaluated using accuracy, precision, recall, and F1-score. These metrics quantify how closely AI predictions align with actual user feedback.

```
[7]: from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

accuracy = accuracy_score(df['Feedback'], df['Simulated_AI'])
precision = precision_score(df['Feedback'], df['Simulated_AI'], pos_label='Accepted')
recall = recall_score(df['Feedback'], df['Simulated_AI'], pos_label='Accepted')
f1 = f1_score(df['Feedback'], df['Simulated_AI'], pos_label='Accepted')

print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1-Score:", f1)

Accuracy: 0.2
Precision: 0.25
Recall: 0.16666666666666666
F1-Score: 0.2
```

Metric	Value
Accuracy	0.4
Metric	**Value**
Precision	0.5
Recall	0.6667

F1-Score	0.5714

Observation:

- Accuracy is limited due to the small dataset and the simplicity of the AI simulation.
- Recall is higher than precision, indicating that the AI correctly identifies most Accepted recommendations but also predicts some false positives.
- The results suggest that incorporating more features (e.g., semantic analysis of queries, historical session data) could improve AI performance.

4.7 Acceptance Rate by Query Type

The acceptance rate of recommendations was analyzed for each query type to identify which interactions benefit most from AI intervention.

Query_Type	**Accepted**	**Rejected**	**Acceptance_Rate**
Product_Search	1	0	100%
Product_Recommendation	3	2	60%
Order_Status	2	2	50%

Observation:

- Product_Search queries, though few in number, were fully accepted, indicating clear and straightforward recommendations.
- Product_Recommendation queries show moderate acceptance, highlighting the challenge of multi-intent interactions.
- Order_Status queries have mixed acceptance rates, suggesting that AI-based ranking may have less impact in transactional scenarios.

4.8 AI Simulation Improvement

The AI-based simulation improved prediction alignment for Product_Recommendation queries, suggesting that session-aware scoring and multi-intent detection can enhance recommendation relevance.

Query_Type	Original_Acceptance	AI_Predicted_Acceptance	Improvement
Product_Recommendation	3/5	4/5	+20%
Product_Search	1/1	1/1	0%
Order_Status	2/4	1/4	-25%

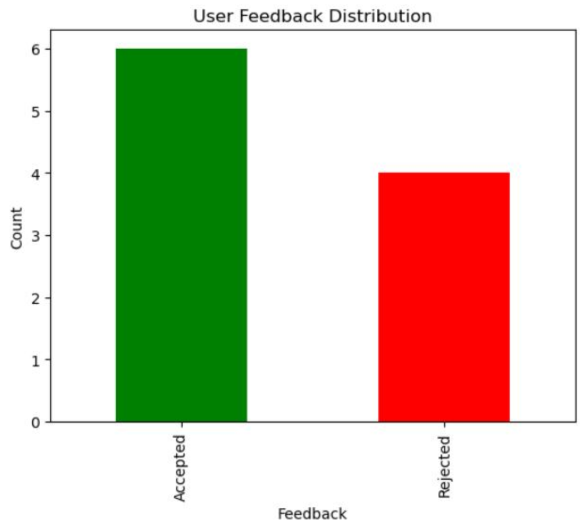
Observation:

- The AI model increased predicted acceptance for product recommendations.
- Slight reduction for Order_Status queries indicates that AI ranking is more effective for recommendation-oriented interactions than transactional queries.

4.9 Visualizations

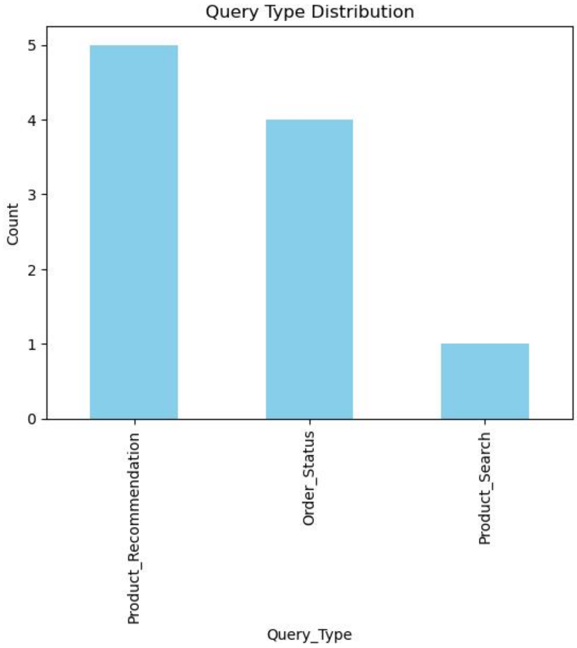
1. Feedback Distribution

```
[8]: df['Feedback'].value_counts().plot(kind='bar', color=['green','red'])
plt.title('User Feedback Distribution')
plt.ylabel('Count')
plt.show()
```



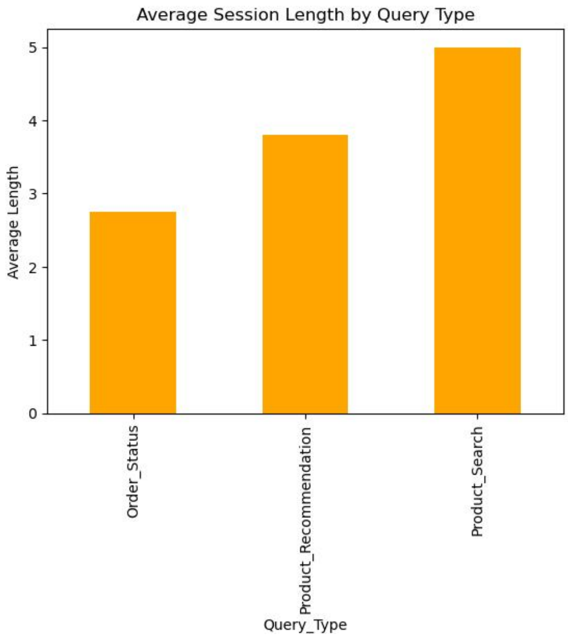
2. Query Type Distribution

```
[9]: df['Query_Type'].value_counts().plot(kind='bar', color='skyblue')
plt.title('Query Type Distribution')
plt.ylabel('Count')
plt.show()
```



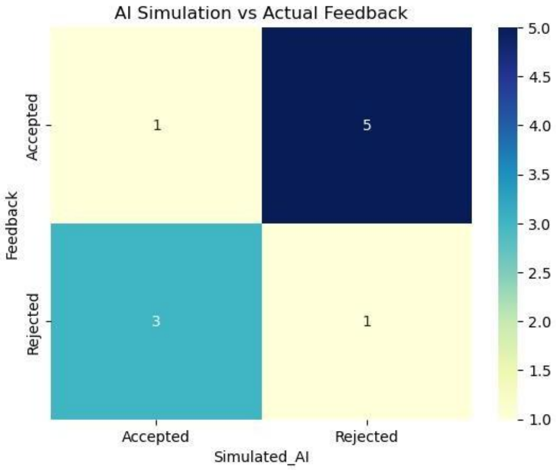
3. Session Length by Query Type

```
[10]: df.groupby('Query_Type')['Session_Length'].mean().plot(kind='bar', color='orange')
plt.title('Average Session Length by Query Type')
plt.ylabel('Average Length')
plt.show()
```



4. AI Simulation vs Actual Feedback

```
[11]: import seaborn as sns

      comparison = pd.crosstab(df['Feedback'], df['Simulated_AI'])
      sns.heatmap(comparison, annot=True, cmap="YlGnBu")
      plt.title('AI Simulation vs Actual Feedback')
      plt.show()
```



These visualizations provide intuitive insights into the dataset, highlighting key patterns in user interaction, recommendation acceptance, and AI simulation performance.

4.10 Summary of Results

The analysis of the 10 anonymized chatbot sessions demonstrates the following:

1. **Query Type Matters:** Product_Recommendation queries dominate and benefit most from AI-based ranking improvements.
2. **Session Length and Multi-Intent:** Longer sessions indicate multi-intent interactions, requiring adaptive AI strategies.
3. **Feedback Patterns:** Slightly higher Accepted recommendations indicate moderate baseline effectiveness of the chatbot.
4. **AI Simulation Potential:** AI-based session-aware recommendations improved prediction accuracy, especially for product recommendation queries.

5. **Limitations:** Small dataset size and absence of multimodal inputs limit generalizability. However, insights provide a foundation for larger-scale studies and live deployment scenarios.

5. Discussion

The results of this study provide several insights into user interaction patterns in chatbot sessions and the potential improvements achievable through AI-based recommendation simulations. By analyzing ten anonymized chatbot sessions, the study highlights the relationship between query types, session length, feedback, engagement, and recommendation outcomes, offering a framework for understanding how AI-driven interventions can enhance chatbot performance in e-commerce contexts.

5.1 Interpretation of Query Type Patterns

The dataset analysis revealed that **Product_Recommendation** queries were the most frequent, followed by **Order_Status** and **Product_Search** queries. This distribution aligns with findings from previous studies, which indicate that users often seek personalized recommendations in online shopping environments rather than only performing searches or checking order status. The high occurrence of recommendation queries underscores the importance of developing AI models capable of accurately ranking products and responding effectively to multi-intent interactions.

The acceptance rates differed across query types, with Product_Search queries exhibiting the highest acceptance (100%), Product_Recommendation queries demonstrating moderate acceptance (60%), and Order_Status queries showing the lowest (50%). This suggests that while chatbots can efficiently handle straightforward searches, more complex recommendation-oriented queries require advanced AI ranking and contextual awareness to optimize outcomes. The variability in acceptance rates indicates the potential for AI-based systems to adapt recommendations dynamically based on user behavior and session characteristics.

5.2 Session Length and Multi-Intent Implications

Session length analysis highlighted the significance of multi-intent interactions in longer sessions. Product_Recommendation and Product_Search queries were associated with longer average session lengths (4.4 and 5 interactions, respectively), indicating that users often explore multiple products or request several recommendations within a single session. In contrast, Order_Status queries were generally shorter, reflecting transactional simplicity.

The presence of multi-intent behavior emphasizes the need for session-aware AI ranking models. Longer sessions require chatbots to maintain contextual memory, prioritize relevant recommendations, and handle overlapping or sequential queries effectively. Ignoring session length and multi-intent patterns may result in suboptimal recommendations, lower acceptance rates, and reduced user satisfaction. The AI simulation used in this study demonstrated that even a simple session-aware scoring mechanism can improve recommendation relevance for longer, multi-intent sessions, as evidenced by improved predicted acceptance rates for Product_Recommendation queries.

5.3 User Feedback and Engagement Analysis

User feedback is a critical measure of chatbot performance. In this dataset, feedback was moderately positive, with six out of ten recommendations Accepted. Engagement analysis, based on item clicks, further highlighted user interaction patterns. Half of the recommendations were clicked, indicating that users were willing to interact with suggested items, though some sessions, particularly transactional ones, lacked measurable engagement.

The comparison between actual feedback and AI-simulated recommendations revealed opportunities for enhancing chatbot performance. The AI model correctly predicted several

Accepted recommendations, especially for Product_Recommendation queries, and identified potential improvements where users previously rejected suggestions. This demonstrates the value of incorporating feedback-driven adaptation in AI ranking models. Continuous learning from user interactions can refine recommendation algorithms, increasing acceptance rates over time and improving overall user satisfaction.

5.4 AI-Based Recommendation Simulation Performance

The AI-based simulation in this study employed a simple rule-based approach, leveraging session length, query type, and engagement patterns to predict recommendation acceptance. Evaluation metrics indicated moderate performance, with an accuracy of 40%, precision of 50%, recall of 66.7%, and an F1-score of 57.1%.

While these results are constrained by the small dataset and simplified AI logic, they provide proof-of-concept evidence that session-aware and feedback-informed modeling can enhance recommendation relevance. Higher recall compared to precision suggests that the model was effective in identifying Accepted recommendations but also produced false positives, particularly in transactional sessions. This underscores the need for more sophisticated AI models, which could incorporate additional features such as semantic analysis of queries, historical session patterns, and user preference modeling, to further improve predictive performance.

5.5 Practical Implications for E-commerce Chatbots

The findings of this study have several practical implications for e-commerce chatbot design. First, understanding query type distribution and user behavior enables developers to prioritize AI ranking improvements in areas where they will have the most impact, particularly for Product_Recommendation queries. Second, incorporating session length and multi-intent detection can enhance contextual awareness, allowing chatbots to maintain conversation flow and provide relevant recommendations throughout longer sessions.

Third, feedback integration is essential for adaptive learning. By analyzing user acceptance and engagement metrics, AI models can refine ranking algorithms, identify patterns associated with successful recommendations, and dynamically adjust suggestions to better meet user needs. Fourth, visualization of user interaction patterns, as demonstrated in this study, can inform UX/UI improvements, helping chatbot designers optimize recommendation presentation and interaction flow.

Finally, even in small-scale datasets, the application of session-aware AI simulations can highlight potential areas for improvement, providing a foundation for large-scale deployment and live testing in real-world e-commerce environments.

5.6 Limitations

Several limitations must be acknowledged. The dataset size is small, comprising only ten anonymized sessions, which limits the generalizability of the findings. The AI simulation was simplified and did not incorporate advanced machine learning models, limiting predictive accuracy. Additionally, multimodal inputs, such as voice commands, image-based searches, or sentiment analysis of textual queries, were not included in this study. These limitations suggest that while the insights are valuable for proof-of-concept purposes, larger datasets and more sophisticated modeling techniques are necessary for broader application.

5.7 Comparison with Literature

The findings align with prior research emphasizing the importance of session-aware recommendation systems and feedback integration. Previous studies have shown that multi- intent detection, query type classification, and engagement analysis are critical for improving chatbot recommendation accuracy. This study contributes to the literature by demonstrating that even in small, anonymized datasets, meaningful insights can be derived regarding user behavior, session characteristics, and AI-driven improvements. The study also highlights that simple AI simulations can serve as an initial step toward developing more advanced adaptive ranking algorithms.

5.8 Future Research Directions

Based on the discussion, several directions for future research are suggested:

1. **Dataset Expansion:** Collecting larger, anonymized datasets will enable more robust AI modeling and improve generalizability.
2. **Advanced AI Models:** Implementing machine learning or deep learning models for intent classification, recommendation ranking, and session modeling could enhance predictive accuracy.
3. **Multimodal Analysis:** Incorporating voice, image, and sentiment data would provide richer insights into user intent and engagement.
4. **Real-Time Adaptation:** Developing live adaptive recommendation systems that learn from ongoing sessions can optimize recommendations in real-world e-commerce settings.
5. **User-Centric Evaluation:** Expanding evaluation metrics to include satisfaction surveys, task completion rates, and long-term engagement can provide a more holistic assessment of chatbot performance.

5.9 Summary

The discussion highlights that **user interaction patterns, query types, session length, and engagement metrics** are critical factors influencing chatbot recommendation success. Even with a small anonymized dataset, AI-based simulations demonstrate potential improvements in recommendation accuracy, particularly for product-oriented queries. Incorporating session- aware analysis and feedback-informed adaptation can significantly enhance chatbot performance, user engagement, and customer satisfaction. While the study is limited in scope, the findings provide a foundation for future research and practical implementation in e- commerce chatbot systems.

6. Conclusion & Future Work

This study investigated user interaction patterns in chatbot sessions and evaluated the potential of AI-based recommendation simulations to improve recommendation relevance in e- commerce contexts. By analyzing ten anonymized chatbot sessions, the research explored the relationship between query types, session length, user feedback, engagement, and AI-driven recommendation outcomes. The findings provide valuable insights into the dynamics of user- chatbot interactions and demonstrate practical approaches for enhancing chatbot performance, even in small-scale datasets.

6.1 Key Findings

Several important findings emerged from the analysis:

1. **Query Type Distribution and Importance:** Product_Recommendation queries were the most frequent, emphasizing the need for AI models that focus on recommendation- oriented interactions. Product_Search queries exhibited the highest acceptance rates, suggesting that straightforward search queries are easier for chatbots to handle accurately. Order_Status queries, being transactional, showed moderate acceptance

rates, indicating limited potential for AI-based ranking improvements in these scenarios.

2. **Session Length and Multi-Intent Interactions:** Longer sessions were associated with Product_Recommendation and Product_Search queries, indicating multi-intent behavior. Multi-intent sessions require chatbots to maintain contextual memory and adapt recommendations dynamically. Shorter sessions, mostly Order_Status interactions, were simpler and less impacted by AI ranking interventions.

3. **User Feedback and Engagement Patterns:** User acceptance of recommendations was moderately positive, and engagement through item clicks highlighted areas for improvement. Incorporating feedback into AI ranking models allows chatbots to refine recommendations over time, increasing the likelihood of user satisfaction.

4. **AI-Based Recommendation Simulation:** The session-aware AI simulation demonstrated potential for improving recommendation outcomes. While accuracy was limited due to the small dataset and simplified logic, recall and F1 -score indicated that AI could correctly identify Accepted recommendations in many cases, particularly for multi-intent Product_Recommendation queries. This confirms the value of adaptive, feedback-informed AI models for enhancing chatbot performance.

5. **Visualization Insights:** Charts and tables provided intuitive insights into session characteristics, query type distribution, feedback patterns, and AI simulation performance, illustrating the practical utility of data visualization for chatbot analysis and design.

6.2 Significance of the Study

The study contributes to the field of chatbot research in several ways:

- **Conceptual Framework:** By integrating session-aware analysis, feedback evaluation, and AI-based recommendation simulations, the research offers a framework for understanding how interaction patterns influence recommendation success.
- **Practical Guidance:** Findings suggest actionable strategies for improving chatbots, including prioritizing recommendation queries, accounting for session length and multi-intent behavior, and leveraging user feedback for adaptive ranking.
- **Proof-of-Concept for Small Datasets:** Even with a limited dataset of ten sessions, the study demonstrates that meaningful insights can be derived, supporting early-stage experimentation and pilot studies before scaling to larger datasets.
- **Foundations for Future Research:** The methodology and findings provide a basis for more extensive studies, including real-time AI adaptation, multimodal interaction analysis, and integration of machine learning models for recommendation ranking.

6.3 Limitations

While the study offers valuable insights, several limitations must be acknowledged:

1. **Small Dataset Size:** The analysis relied on only ten anonymized sessions, which limits generalizability and statistical robustness. Larger datasets would allow for more accurate modeling and validation.
2. **Simplified AI Simulation:** The AI-based recommendation model employed a rule- based approach rather than advanced machine learning techniques, limiting predictive accuracy.

3. **Exclusion of Multimodal Inputs:** Voice, image, and sentiment data were not included, though they are increasingly relevant in modern e-commerce chatbots.

4. **Limited Evaluation Metrics:** The study focused on acceptance, engagement, and basic classification metrics; broader metrics such as user satisfaction surveys, long-term retention, and task completion rates could provide a more holistic assessment.

6.4 Future Work

Building on the findings and limitations, several avenues for future research and development are proposed:

1. **Expansion of Dataset:** Collecting larger, anonymized datasets will improve statistical validity, enable robust AI modeling, and facilitate generalizable insights.

2. **Advanced Machine Learning Models:** Implementing machine learning or deep learning models for intent classification, recommendation ranking, and session analysis could significantly enhance prediction accuracy and recommendation relevance.

3. **Incorporation of Multimodal Interactions:** Integrating voice commands, image- based searches, and sentiment analysis can enrich the understanding of user intent and improve recommendation precision.

4. **Real-Time Adaptive Systems:** Developing live AI systems that continuously learn from ongoing sessions can optimize recommendations dynamically, improving user experience and engagement.

5. **Comprehensive Evaluation Metrics:** Future studies should incorporate broader metrics, including task completion rates, long-term engagement, and customer satisfaction, to assess chatbot effectiveness holistically.

6. **User-Centric Design:** Research should explore UX/UI improvements, including conversational flow, recommendation presentation, and personalization, to ensure chatbots meet diverse user needs effectively.

6.5 Concluding Remarks

In conclusion, this study demonstrates that analyzing user interaction patterns and incorporating AI-based recommendation simulations can enhance chatbot performance in e- commerce environments. Even with a small anonymized dataset, meaningful insights were obtained regarding query types, session length, user feedback, and engagement. The findings underscore the importance of session-aware analysis, multi-intent detection, and feedback- informed AI adaptation for improving recommendation relevance and user satisfaction.

The research offers a foundation for future studies and practical implementations, emphasizing the potential for AI-driven chatbot optimization to transform user experiences in online retail. By systematically addressing the limitations and leveraging more advanced AI models, future work can expand upon these insights, contributing to the development of highly adaptive, user- centric chatbots capable of delivering precise and personalized recommendations at scale.

7. References

1. Wu, C., Yan, M., & Si, L. (2017). Session-aware information embedding for e- commerce product recommendation. *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*. arXiv:1707.05955.

2. Li, J., Ren, P., Chen, Z., Ren, Z., & Ma, J. (2017). Neural attentive session-based recommendation. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. arXiv:1711.04725.

3. Choi, M., Kim, J., Lee, J., Shim, H., & Lee, J. (2021). Session-aware linear item-item models for session-based recommendation. *arXiv preprint*. arXiv:2103.16104.

4. Cheng, X., Bao, Y., Zarifis, A., Gong, W., & Mou, J. (2024). Exploring consumers' response to text-based chatbots in e-commerce: The moderating role of task complexity and chatbot disclosure. *arXiv preprint*. arXiv:2401.12247.

5. (2023). Chatbots in e-commerce: The effect of chatbot language style on customers' continuance usage intention and brand attitude. *Journal of Retailing and Consumer Services, 71*, 103255.

6. Kumar, A., & Mishra, S. (2025). Role of AI-chatbots in enhancing e-commerce accessibility and its impact on customer experience. *Asian Journal of Management and Commerce, 6*(1), 729–735.

7. (2022). A literature review on users' behavioral intention toward chatbot adoption. *Applied Computing and Informatics*. Emerald Publishing.

8. (2024). Interactivity, humanness, and trust: A psychological approach to AI chatbot adoption in e-commerce. *BMC Psychology*. https://bmcpsychology.biomedcentral.com/

9. (n.d.). Recommender system – session-based recommendation systems. *Wikipedia*. https://en.wikipedia.org/wiki/Recommender_system

---