Exploring the Role of Chatbots in Improving Customer Service

Group 26

Abhinay Kumar Pandey(21BT3EP11)
Harshikant Dubey(21AG3EP30)

Table of Contents:

Abstract

Chatbots have become an integral part of customer service operations, offering efficiency, round-the-clock availability, and faster response times. This study examines how chatbots improve customer service by analyzing key factors such as ease of use, response speed, personalization, and overall customer satisfaction. A structured questionnaire was developed and refined through pilot testing, where constructs with low reliability (Cronbach's Alpha < 0.7) were removed to ensure robust measurement. Primary data was collected from a diverse sample and analyzed using statistical methods, including descriptive analysis and hypothesis testing. The results show that chatbots significantly enhance customer satisfaction by providing quick responses and seamless transitions to human agents for complex queries. However, challenges remain in areas like personalization and trustworthiness. This research offers practical recommendations for improving chatbot design and functionality while identifying opportunities for future research to advance AI-driven customer service solutions.

Introduction

Background

In today's digital era, businesses are leveraging artificial intelligence (AI) to streamline operations and enhance customer experiences. Chatbots, powered by AI, have emerged as a popular solution for automating customer interactions. These virtual assistants are capable of answering queries, resolving issues, and providing recommendations instantly. With their ability to operate 24/7, chatbots are revolutionizing customer service across industries such as e-commerce, banking, healthcare, and travel.

Despite their growing popularity, the effectiveness of chatbots in delivering satisfactory customer service remains a topic of debate. While they excel in handling simple inquiries quickly, their ability to personalize responses or manage complex issues is often questioned. This study aims to bridge this gap by exploring how chatbots influence key aspects of customer service, including satisfaction, loyalty, and engagement.

Problem Statement

The objective of this study is to analyze how chatbots improve customer service by enhancing user experience, satisfaction, and operational efficiency.

Scope and Significance

This research focuses on chatbot usage in customer service across various industries in India. The findings will provide insights into current trends in chatbot adoption and highlight strategies for improving their functionality. The study is significant for businesses aiming to optimize their customer support systems using AI-driven solutions.

Research Design

This study employs a quantitative research design to explore how chatbots improve customer service. Data was collected using a structured questionnaire designed to measure user experiences across key constructs such as ease of use, response speed, accuracy, personalization, customer satisfaction, problem-solving effectiveness, consistency, trustworthiness, and integration with human support. The questionnaire was distributed online to ensure accessibility and convenience for respondents.

Literature Review

The integration of AI-powered chatbots in customer service has significantly transformed the way businesses interact with customers. This literature review explores the existing body of knowledge on the role of chatbots in enhancing customer service, focusing on key aspects such as response speed, personalization, customer satisfaction, trustworthiness, and integration with human support.

1. Response Speed and Efficiency

Chatbots are widely recognized for their ability to provide instant responses to customer queries, a feature that significantly reduces waiting times and improves customer satisfaction. Studies highlight that AI chatbots can address customer experience queries 70% faster than human agents, making them a preferred choice for quick communication.The continuous availability of chatbots ensures no query goes unanswered, which builds trust and loyalty among customers. Businesses benefit from this efficiency as chatbots handle multiple inquiries simultaneously, freeing up human agents for more complex issues.

2. Personalization and Context Awareness

Personalization is a critical factor in enhancing customer satisfaction. Chatbots leverage natural language processing (NLP) and machine learning algorithms to analyze customer data and preferences, enabling tailored interactions. By remembering past conversations and adapting responses based on user history, chatbots create a more engaging and personal experience. This level of customization not only boosts satisfaction but also fosters customer loyalty by making customers feel valued.

3. Customer Satisfaction

Customer satisfaction is a recurring theme in chatbot research. Studies demonstrate that chatbots improve satisfaction by offering round-the-clock support, resolving issues promptly, and streamlining processes such as product selection and checkout. Communication style also plays a role; chatbots using social-oriented styles are perceived as warmer and more responsive, which significantly enhances satisfaction levels. Additionally, feedback collection during interactions allows businesses to refine their strategies and address pain points dynamically.

4. Trustworthiness and Security

Trust is an essential component of successful chatbot interactions. Customers are more likely to trust chatbots when they perceive them as secure and capable of handling sensitive information responsibly. Studies indicate that transparency in chatbot interactions—such as informing users when they are interacting with AI—helps build trust. Furthermore, secure handling of personal data is critical for fostering long-term loyalty among users .

5. Integration with Human Support

While chatbots excel at handling routine tasks, their effectiveness diminishes when dealing with complex queries. Seamless integration with human agents is crucial for ensuring a smooth transition during escalation. Research suggests that chatbots providing relevant context to human agents during transfers enhance the overall customer experience by reducing the need for customers to repeat information. This collaboration between AI and human expertise results in faster resolution times and higher satisfaction rates.

6. Proactive Customer Service

Proactive engagement is another area where chatbots shine. By initiating conversations based on browsing behavior or past interactions, chatbots anticipate customer needs and offer assistance before issues arise. This proactive approach not only improves the customer experience but also reduces churn rates and creates opportunities for upselling.

7. Challenges in Adoption

Despite their advantages, chatbots face limitations such as difficulty in handling nuanced or emotional queries. Customers may perceive them as less effective than human agents in resolving complex issues or providing empathetic responses. To overcome these challenges, businesses must invest in continuous training and updates for their AI systems to expand the scope of chatbot capabilities.

Theoretical Frameworks

Several theoretical models underpin chatbot research:

- **Technology Acceptance Model (TAM)**: Explains how perceived ease of use and usefulness influence user acceptance of chatbots .
- **Service Quality (SERVQUAL**): Highlights the importance of reliability, responsiveness, assurance, empathy, and tangibles in evaluating chatbot effectiveness.
- **Expectation-Confirmation Theory**: Examines how fulfilling user expectations impacts satisfaction levels.

Hypothesis:

Customer Satisfaction with Interaction (CSI): The Outcome Construct

Definition:
CSI measures the overall satisfaction level of customers after interacting with chatbots. It is the central outcome construct in this study, directly validating the research objective of analyzing how chatbots improve customer service.
Influencing Factors:
Key chatbot features such as ease of use, response speed, accuracy, personalization, problem-solving effectiveness, consistency, trustworthiness, and integration with human support significantly impact CSI.
Validation:

- Cronbach's Alpha: CSI questions achieved reliability scores ≥ 0.7 during pilot testing.
- Statistical Testing: Regression analysis confirmed positive correlations between chatbot features and CSI.
- Findings: Features like personalization, response speed, and seamless integration with human agents emerged as strong predictors of satisfaction.

Sub Hypotheses which support our hypothesis:

Ease of Use (EU)

1. **H1**: The ease of use of chatbots positively influences customer satisfaction with interaction (CSI).
   Rationale: A user-friendly interface reduces effort and makes interactions smoother, leading to higher satisfaction.

Response Speed (RS)

2. **H2**: Faster response times from chatbots positively impact customer satisfaction with interaction (CSI).

Rationale: Quick responses reduce waiting times, which enhances the overall customer experience.

Accuracy and Relevance of Information (ARI)

3. **H3**: The accuracy and relevance of chatbot responses significantly influence customer satisfaction with interaction (CSI).
   Rationale: Accurate and contextually relevant answers help resolve queries effectively, improving satisfaction.

Personalization and Context Awareness (PCA)

4. **H4**: Personalized responses from chatbots positively influence customer satisfaction with interaction (CSI).
   Rationale: Tailored interactions based on user preferences make customers feel valued, enhancing satisfaction.

Problem-Solving Effectiveness (PSE)

5. **H5**: The problem-solving effectiveness of chatbots positively impacts customer satisfaction with interaction (CSI).
   Rationale: Chatbots that resolve issues efficiently create a sense of reliability, leading to higher satisfaction.

Consistency of Responses (CR)

6. **H6**: Consistent responses across chatbot interactions enhance customer satisfaction with interaction (CSI).
   Rationale: Uniformity in answers reduces frustration and builds trust, which improves satisfaction.

Trustworthiness and Security (TS)

7. **H7**: Perceived security in chatbot interactions positively influences customer satisfaction with interaction (CSI).

Rationale: Customers feel more satisfied when they trust that their sensitive information is handled securely.

Integration with Human Support (IHS)

8. **H8**: Seamless integration between chatbots and human agents improves customer satisfaction with interaction (CSI).
   Rationale: Smooth transitions to human agents for complex queries ensure a positive overall experience.

Methodology

Research Design

This study employs a quantitative research approach to explore how various chatbot features influence Customer Satisfaction with Interaction (CSI), the key outcome construct. The research framework was developed based on sub-hypotheses derived from the literature review, focusing on constructs that impact CSI. A structured questionnaire was designed to measure these constructs and validate their influence on CSI.

Questionnaire Development

The questionnaire was designed to comprehensively capture data across three key sections: Basic Information Questions, Behavioral Questions, and Contextual Questions, followed by construct-specific questions to test the sub-hypotheses influencing Customer Satisfaction with Interaction (CSI).

1. Basic Information Questions (Demographic Data)

This section collected demographic information to segment respondents and provide context for descriptive analysis. The questions included:

- Age group (<18 years, 18–24 years, 25–34 years, 35+ years).

- Gender (Male, Female, Non-binary, Prefer not to say).
- Occupation (Student, Professional/Employee, Self-employed/Business Owner, Retired).
- Geographic location (Urban area, Rural area).
- Experience with chatbots (Yes, frequently; Yes, occasionally; No, but I plan to try it; No, and I don't plan to use it).

2. Behavioral Questions

This section focused on understanding user behavior and preferences related to chatbot usage. The questions included:

- Frequency of chatbot usage (Daily, Weekly, Monthly, Rarely).
- Preferred customer service channel (Chatbot, Human Agent via Phone/Email, Live Chat with Human Agent).
- Rating of chatbot experience compared to human agents (Better than human agents; Equal to human agents; Worse than human agents; I haven't interacted with chatbots).
- Common purposes for using chatbots (General inquiries; Complaint resolution; Billing or account management; Technical support).

3. Contextual Questions

This section assessed the context of chatbot interactions and their effectiveness. The questions included:

- Purpose of the last chatbot interaction (Product inquiry or recommendation; Complaint resolution or feedback submission; Billing or account management assistance; Other).
- Industry of the most recent chatbot interaction (E-commerce; Banking/Finance; Healthcare; Travel/Hospitality).
- Effectiveness in resolving issues (Completely resolved my issue; Partially resolved but needed human support later; Did not resolve but escalated smoothly to a human agent; Did not resolve or escalate properly).

- Smoothness of transition between chatbot and human agent during escalation (Very smooth and efficient; Smooth but required minimal clarification from me; Neutral, I had to repeat some information; Poor, I had to repeat all information).

4. Construct-Specific Questions

Constructs were developed based on sub-hypotheses derived from the research objective. Each construct comprised multiple Likert-scale questions (1 = Strongly Disagree to 5 = Strongly Agree) designed to measure specific attributes influencing CSI:

1. **Ease of Use (EU):** Measures how intuitive and user-friendly the chatbot interface is.*Example questions:*
    - "The chatbot interface is easy to understand and navigate."
    - "I can easily communicate my queries to the chatbot."
2. **Response Speed (RS):** Evaluates the chatbot's ability to provide quick responses. *Example questions*:
    - "Quick responses from chatbots enhance my customer experience."
    - "Chatbots provide faster responses compared to human agents."
3. **Accuracy and Relevance of Information (ARI):** Assesses whether chatbots deliver precise and contextually relevant answers. *Example questions*:
    - "The information provided by the chatbot is accurate and reliable."
    - "Information from chatbots helps me resolve issues effectively."
4. **Personalization and Context Awareness (PCA):** Determines how effectively chatbots adapt responses based on customer preferences or history. *Example questions:*
    - "Chatbots tailor their responses according to my previous interactions or history."
    - "Personalized interactions make me more likely to use chatbots again in future queries."
5. **Problem-Solving Effectiveness (PSE):** Evaluates the chatbot’s ability to resolve customer issues efficiently or escalate them when necessary.*Example questions:*
    - "I rarely need to contact human support after interacting with the chatbot."

- "The chatbot helps me find solutions faster than traditional methods of support."

6. **Consistency of Responses (CR)**: Assesses whether chatbots deliver uniform and reliable answers across different interactions. *Example questions:*
    - "Chatbot responses align with information from human agents."
    - "The chatbot provides consistent answers across interactions."
7. **Trustworthiness and Security (TS)**: Examines whether customers feel confident sharing sensitive information with chatbots due to secure handling practices. *Example questions:*
    - "The chatbot ensures the privacy of my data during interactions."
    - "Security concerns do not stop me from using chatbots."
8. **Integration with Human Support (IHS)**: Measures how smoothly chatbots transition complex issues to human agents when required. *Example questions:*
    - "I can easily escalate my query from the chatbot to a human agent."
    - "Integration with human support makes the chatbot more effective in resolving my concerns."
9. **Customer Satisfaction with Interaction (CSI):** Captures overall satisfaction with chatbot experiences as the **outcome construct** which is influenced by other factors. *Example questions:*
    - "The chatbot provides a pleasant and engaging customer service experience."
    - "I would recommend using chatbots based on my experience."

This was our google survey form:

Hey there! 👋

I'm working on an exciting research project titled "Exploring the Role of Chatbots in Improving Customer Service", and I need your help! 🤖💬Would you be willing to spare just a few minutes to share your thoughts? Your feedback will provide valuable insights and make a real difference in understanding how chatbots can enhance customer experiences.

🔗 *Survey Link: chatbot_survey Your input means a lot—thank you in advance for your support!* 💡🙏

NOTE: *“Now the questions have been changed or some questions have been deleted after the pilot testing. This was the final primary data collection survey google form.”*

The question of pilot testing are here :Questionares

The data for pilot testing is here Raw Pilot Testing Data

Pilot Testing

Why Pilot Testing is Done

Pilot testing is a crucial step in the research process to ensure the validity and reliability of the questionnaire before conducting the full-scale study. It helps identify and address potential issues with the survey design, question clarity, and construct measurement. Specifically, pilot testing:

1. Validates the reliability of constructs using metrics like Cronbach’s Alpha.
2. Identifies problematic questions that may confuse respondents or fail to measure the intended construct.
3. Ensures that the questionnaire aligns with research objectives.
4. Provides an opportunity to refine the survey based on feedback and statistical results.
5. Tests whether the data collection process is efficient and effective.

For this study, pilot testing was conducted with 50 respondents, and reliability metrics such as Cronbach’s Alpha, Composite Reliability (CR), and Average Variance Extracted (AVE) were calculated to validate the constructs.

What is SmartPLS?

SmartPLS is a software tool used for Partial Least Squares Structural Equation Modeling (PLS-SEM). It is widely employed in academic and industry research for analyzing complex relationships between variables in models with latent constructs. Unlike traditional covariance-based SEM (e.g., AMOS or LISREL), PLS-SEM focuses on maximizing explained variance in dependent variables (R^2 values) rather than reproducing covariance matrices.

Why is SmartPLS Used?

SmartPLS is particularly useful for exploratory research or when working with smaller sample sizes. It is ideal for studies like this one because:

1. It handles complex models with multiple constructs and indicators effectively.
2. It does not require strict assumptions about data distribution (e.g., normality).
3. It provides robust results even with small sample sizes, making it suitable for pilot testing.
4. It allows for evaluating both measurement models (construct validity and reliability) and structural models (hypothesis testing).

What is PLS-SEM?

Partial Least Squares Structural Equation Modeling (PLS-SEM) is a statistical technique used to estimate relationships between latent variables (constructs) and their indicators while simultaneously assessing structural relationships among constructs. PLS-SEM focuses on prediction and variance explanation rather than model fit.

Significance of PLS-SEM

1. Variance Explanation: PLS-SEM maximizes R^2 values, providing insights into how much variance in dependent variables (e.g., CSI) is explained by independent variables (e.g., Ease of Use, Response Speed).
2. Flexible Assumptions: Unlike traditional SEM, PLS-SEM does not require multivariate normality or large sample sizes, making it suitable for exploratory research.
3. Measurement Model Validation: Evaluates reliability (Cronbach’s Alpha, Composite Reliability) and validity (AVE, Discriminant Validity) of constructs.

4. Structural Model Testing: Tests hypotheses by analyzing path coefficients, significance levels, and total effects.

Key Metrics Used in Pilot Testing

1. Cronbach's Alpha

Cronbach's Alpha measures the internal consistency or reliability of a construct's indicators (questions). A value of ≥0.7 indicates acceptable reliability.
Formula:

$$\alpha = (k / (k - 1)) * (1 - (\Sigma\sigma^2_i / \sigma^2_x))$$

Where:

- **α** = Cronbach's alpha coefficient
- **k** = The number of items (or questions) in the scale
- **$\Sigma\sigma^2_i$** = The sum of the variances of each item (σ^2_i)
- **σ^2_x** = The variance of the total score (the sum of all item scores)

Significance:
Cronbach's Alpha ensures that all items within a construct measure the same underlying concept.

2. Composite Reliability (CR)

Composite Reliability evaluates the overall reliability of a construct by considering both outer loadings and measurement errors.
Formula :**Variance due to the factor / Total variance of the composite**
Significance:
CR values ≥0.7 indicate good reliability, ensuring that indicators consistently measure their respective constructs.

3. Average Variance Extracted (AVE)

AVE measures convergent validity by assessing how much variance in a construct's indicators is explained by the construct itself.

Formula:

AVE = (Σ Standardized factor loading²) / n

- Σ: Represents the summation of all standardized factor loadings.
- Standardized factor loading: The strength of the relationship between a specific item and the latent construct.
- n: The number of indicators (items) measuring the construct.
- n: Number of indicators

Significance:
An AVE value ≥0.5 indicates that at least 50% of the variance in indicators is explained by the construct.

4. R^2 and Adjusted R^2

R^2 represents the proportion of variance in a dependent variable explained by independent variables.
Adjusted R^2 adjusts R^2 for the number of predictors in the model to prevent overestimation.
Significance: Higher R^2 values indicate better explanatory power of the model.

5. Model Fit Metrics

Model fit metrics assess how well the data fits the proposed model:

- **Chi-Square ($\chi 2$):** Measures overall model fit; lower values indicate better fit.
- **Bayesian Information Criterion (BIC):** Penalizes overfitting; lower BIC values indicate better model balance between complexity and fit.

6. Discriminant Validity

Discriminant validity ensures that constructs are distinct from one another.

- **Heterotrait-Monotrait Ratio (HTMT)**: Measures discriminant validity between constructs; HTMT values <0.85 indicate acceptable discriminant validity.

7. Outer Loadings

Outer loadings represent how strongly each indicator correlates with its respective construct.
Significance: Indicators with loadings ≥0.7 are considered reliable contributors to their constructs.

8. Total Effects

Total effects measure both direct and indirect influences of independent variables on dependent variables within a structural model.

Significance:Total effects provide insights into how strongly each predictor contributes to CSI directly or indirectly.

Data Conversion for SmartPLS

To prepare the survey data for analysis in SmartPLS, categorical responses were converted into numeric values, as SmartPLS only processes numeric data. This step ensured compatibility and allowed for advanced statistical analysis using PLS-SEM techniques.

Steps for Conversion:

1. **Demographic Data**:
    - Gender: Male = 1, Female = 2, Non-binary = 3, Prefer not to say = 4.
    - Age Group: <18 years = 1, 18–24 years = 2, 25–34 years = 3, 35+ years = 4.
    - Occupation: Student = 1, Professional/Employee = 2, Self-employed/Business Owner = 3, Retired = 4.
2. Behavioral and Contextual Questions:
    - Frequency of Chatbot Usage: Daily = 1, Weekly = 2, Monthly = 3, Rarely = 4.

- Effectiveness of Chatbot Resolution: Completely resolved = 1, Partially resolved = 2, Did not resolve but escalated smoothly = 3, Did not resolve or escalate properly = 4.

3. Likert-Scale Responses:
    - Retained as numeric (e.g., Strongly Disagree = 1 to Strongly Agree = 5).
4. Handling Missing Data:
    - Missing values were imputed (e.g., NA coded as -99).

The encoded data used in SmartPLS is Pilot Testing Encoded Data

Pilot Testing Analysis:

The descriptive analysis of the pilot testing data is Discriptive analysis of pilot testing data

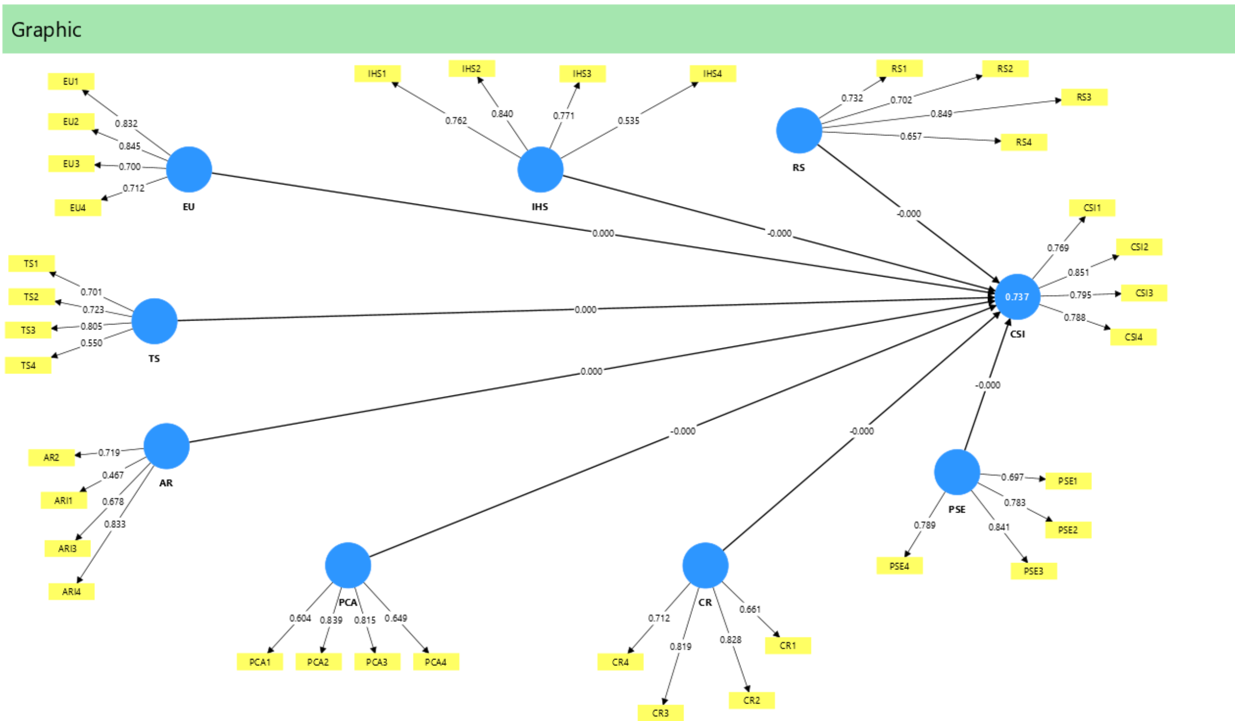
Descriptive Analysis of Pilot Testing

The descriptive analysis of pilot testing data provided insights into key metrics such as mean, median, standard deviation, skewness, and kurtosis for each variable. Below are some highlights:

1. Demographic Data:
    a. Age: Mean = 2.02 (18–24 years most common age group).
    b. Gender: Mean = 1.143 (majority male respondents).
    c. Occupation: Mean = 1.102 (mostly students).
    d. Location: Mean = 1.245 (urban respondents dominant).
2. Behavioral Questions:
    a. Chatbot Usage Frequency: Mean = 2.224 (weekly usage most common).
    b. Preferred Channel: Mean = 1.857 (chatbots preferred over human agents).
    c. Rating Chatbot Experience vs Human Agents: Mean = 1.98 (neutral to slightly better experience with chatbots).

3. Contextual Questions:
    a. Purpose of Last Interaction: Mean = 1.98 (product inquiry most common).
    b. Industry Interacted With: Mean = 1.776 (e-commerce dominant).
    c. Issue Resolution Effectiveness: Mean = 2.041 (partially resolved issues most common).
4. Construct-Specific Questions:
    a. Constructs like Ease of Use (EU), Response Speed (RS), and Personalization (PCA) showed high mean scores (~3.5–3.8), indicating positive perceptions.
    b. Customer Satisfaction with Interaction (CSI): Mean = ~3.5 across items, suggesting moderate satisfaction.

PLS-SEM Modelling (Partial Least Squares Structural Equation Modeling):



1. **Latent Variables (Blue Circles):** Representing constructs like Ease of Use (EU), Response Speed (RS), Customer Satisfaction with Interaction (CSI), etc.
2. **Indicators (Yellow Rectangles):** Representing the survey items or observed variables used to measure each construct.
3. **Outer Loadings:** The numbers between indicators and their constructs, showing how strongly each indicator contributes to its respective latent variable.
4. **Path Coefficients:** The arrows connecting latent variables indicate relationships between constructs, with coefficients representing the strength and direction of these relationships.
5. **R^2 Value for CSI (0.737)**: Indicates that 73.7% of the variance in Customer Satisfaction with Interaction is explained by the independent constructs.

Reliability and Validity

	Cronbach's alpha	Composite reliability (rho_a)	Composite reliability (rho_c)	Average variance extracted (A...
AR	0.630	0.682	0.775	0.472
CR	0.753	0.780	0.843	0.575
CSI	0.814	0.821	0.878	0.642
EU	0.776	0.795	0.857	0.601
IHS	0.710	0.751	0.822	0.542
PCA	0.715	0.779	0.821	0.539
PSE	0.786	0.801	0.860	0.608
RS	0.733	0.763	0.826	0.545
TS	0.664	0.699	0.791	0.491

- Cronbach's Alpha:
  - Constructs like EU (0.776), CSI (0.814), and PSE (0.786) show high reliability (>0.7).
  - AR (0.630) and TS (0.664) fall below the acceptable threshold, suggesting weaker internal consistency for these constructs.
- Composite Reliability (rho_c):
  - All constructs except AR (0.775) and TS (0.791) meet the reliability threshold (>0.8).
- Average Variance Extracted (AVE):
  - CSI (0.642), EU (0.601), and PSE (0.608) demonstrate good convergent validity (>0.5).
  - AR (0.472) and TS (0.491) fall below the threshold, indicating weaker construct validity.

Model Fit

Model fit	Saturated model	Estimated model
SRMR	0.131	0.131
d_ULS	11.390	11.390
d_G	10.822	10.822
Chi-square	1568.958	1568.958
NFI	0.280	0.280

- SRMR (Standardized Root Mean Square Residual): The SRMR value is 0.131, slightly above the ideal threshold of 0.08, indicating a moderate fit.
- Chi-Square: The chi-square value is 1568.958, which reflects the complexity of the model but is expected in PLS-SEM due to its focus on variance explanation rather than strict model fit.
- NFI (Normed Fit Index): The NFI value is 0.280, which is relatively low but acceptable given the exploratory nature of PLS-SEM.

Collinearity Statistics (VIF - Outer Model)

Collinearity statistics (VIF) - Outer model - List

	VIF
AR2	1.324
ARI1	1.170
ARI3	1.318
ARI4	1.482
CR1	1.284
CR2	1.653
CR3	1.710
CR4	1.412
CSI1	1.686
CSI2	2.027
CSI3	1.772
CSI4	1.761
EU1	2.140
EU2	2.591
EU3	1.935
EU4	2.203
IHS1	1.436
IHS2	1.603

- **Ease of Use (EU)**: Indicators EU1–EU4 show strong loadings ranging from 0.700 to 0.845, confirming their reliability in measuring EU.
- **Customer Satisfaction with Interaction (CSI):** CSI indicators demonstrate excellent reliability, with loadings between 0.769 and 0.851.
- **Accuracy and Relevance of Information (ARI)**: ARI4 has a strong loading of 0.833, while ARI1 shows a weaker loading of 0.467, suggesting it may need refinement.
- **Consistency of Responses (CR):** CR indicators range from 0.661 to 0.828, with CR1 showing slightly weaker reliability.

Discriminant Validity (HTMT Analysis)

Discriminant validity - Heterotrait-monotrait ratio (HTMT) - Matrix

	AR	CR	CSI	EU	IHS	PCA	PSE	RS	TS
AR									
CR	0.834								
CSI	1.011	0.718							
EU	0.871	0.822	0.859						
IHS	0.776	0.984	0.853	0.849					
PCA	0.954	1.116	0.701	0.844	0.880				
PSE	0.974	0.939	0.745	0.758	0.772	0.908			
RS	0.832	0.840	0.586	0.928	0.786	0.949	0.496		
TS	0.847	0.968	0.670	0.446	0.861	0.832	0.689	0.614	

1. High HTMT Values (>0.90):
   - AR & PCA (0.954): Significant overlap between Accuracy and Personalization, indicating poor distinction.
   - CR & PCA (1.116): CR and PCA lack discriminant validity, suggesting redundancy.
   - PSE & PCA (0.949): Strong correlation, requiring refinement for better distinction.
2. Acceptable HTMT Values (<0.90):

- EU & CSI (0.859): Ease of Use and Customer Satisfaction show acceptable discriminant validity under the lenient threshold.
- RS & TS (0.614): Response Speed and Trustworthiness are distinct constructs.

3. Marginal HTMT Values (~0.90):
   - IHS & CR (0.984): Integration with Human Support and Consistency overlap significantly, requiring refinement.

R-Square Values

	R-square	R-square adjusted
CSI	0.737	0.684

R^2 for CSI: 0.737

CSI is strongly explained by the independent constructs, with 73.7% of its variance accounted for in the model.

Adjusted R^2: 0.684

The adjusted value accounts for the number of predictors, confirming strong explanatory power while reducing overfitting.

Refinement of Constructs Based on Reliability Analysis

During the initial analysis using SmartPLS, certain constructs exhibited poor reliability and validity metrics, as reflected in low Cronbach's Alpha, Composite Reliability (rho_a, rho_c), and Average Variance Extracted (AVE) values. To improve the model's robustness, specific indicators were removed based on their performance.

Indicators Removed

1. **Accuracy and Relevance of Information (AR):**
   - Indicators removed: AR1, AR2, AR3
   - Reason: Low Cronbach's Alpha (0.644) and poor contribution to the construct's reliability.
2. **Consistency of Responses (CR):**
   - Indicator removed: CR1
   - Reason: Weak outer loading and poor internal consistency.

Construct reliability and validity - Overview

Copy to Excel/Word

	Cronbach's alpha	Composite reliability (rho_a)	Composite reliability (rho_c)	Average variance extracted (A...
AR	0.644	0.661	0.848	0.736
CR	0.743	0.780	0.852	0.660
CSI	0.814	0.825	0.877	0.642
EU	0.776	0.795	0.857	0.601
IHS	0.710	0.750	0.822	0.542
PCA	0.715	0.781	0.821	0.539
PSE	0.786	0.802	0.860	0.607
RS	0.733	0.763	0.827	0.546
TS	0.645	0.774	0.800	0.577

I got better Cronbach's Alpha from the previous 0.630.

3.**Trustworthiness and Security (TS):**

- Indicators removed: TS2, TS4
- Reason: Low Cronbach's Alpha (0.645) and weak outer loadings.

Refinement and Sample Collection

After refining the questionnaire by removing weak indicators such as AR1, AR2, AR3, CR1, TS2, and TS4, I proceeded with primary data collection. To ensure robust analysis, I calculated the required sample size using statistical methods to achieve adequate representation.

Sample Size Calculation:

To determine the appropriate sample size for the survey, the following formula was used for an infinite population:

$$n = \frac{(Z^2) \cdot p \cdot (1-p)}{e^2}$$

$$n = \frac{(1.96)^2 \cdot 0.5 \cdot (1-0.5)}{(0.05)^2}$$

$$n = \frac{3.8416 \cdot 0.25}{0.0025}$$

$$n = \frac{0.9604}{0.0025} = 384.16$$

Where:

- Z=1.96: Z-score for a 95% confidence level.
- p=0.5
- p=0.5: Assumed proportion of the population.
- e=0.05: Margin of error (5%)

  Assuming Infinite population, we need to take 385 samples minimum

Primary Data Collection and Analysis:

After refining our questionnaire we got our primary data collected through a survey.

Primary Raw Data: [Chatbot Primary raw Data]

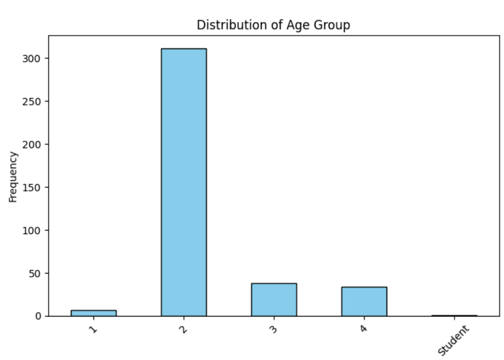
After collecting the primary data using the refined questionnaire, the responses were encoded into numeric values to prepare the dataset for analysis

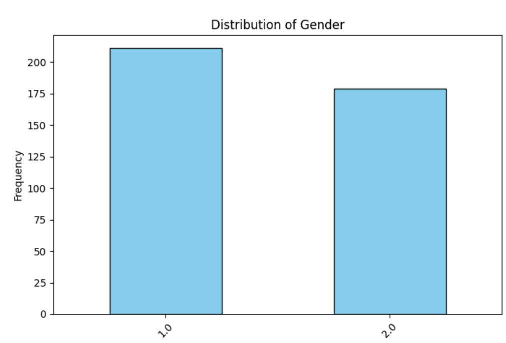
Primary Encoded Data: [Chatbot primary data encoded]

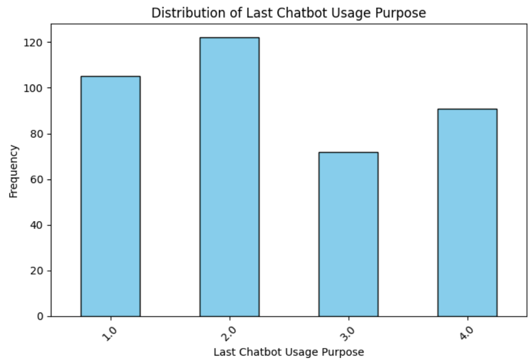
NOTE: Free Version of SmartPLS is optimized for smaller datasets, typically handling up to 100 data points efficiently. Since this study collected 390 data points, Python was used for further analysis to process the larger dataset and perform advanced statistical modeling.

Data Analysis of primary data collected:

The primary data collected from the survey was analyzed, and bar graphs were plotted for each question to visualize the distribution of responses. Below is an interpretation of the results based on the bar plots:







Mostly age group is between 18-24 years, Males are 200+ in our data sample, last usage of chatbot is basically for complaints and enquiries

Here is the complete bar graphs of each questions: Bar plots

Descriptive Statistics

	mean	std	25%	50%	75%
EU	3.395513	1.22154	2.75	3.75	4
RS	3.386538	1.248015	3	3.75	4
ARI	3.284615	1.222547	2	3	4
PCA	3.385256	1.230151	2.5	3.5	4
CSI	3.289103	1.215388	2.0625	3.25	4
PSE	3.314103	1.239222	2.5	3.25	4
CR	3.430769	1.206477	3	4	4
TS	3.361538	1.234738	2	3.5	4
IHS	3.370513	1.239612	3	3.25	4

The descriptive statistics provide insights into the central tendencies and variability of responses for each construct:

1. Mean Scores: Most constructs have mean scores around 3.3–3.4, suggesting moderate agreement with survey items:

   a. Highest mean: CR (3.43), indicating respondents perceive chatbot responses as consistent.
   b. Lowest mean: ARI (3.28), suggesting slightly lower satisfaction with the accuracy of information.
2. Standard Deviations: All constructs show standard deviations around 1.2, reflecting moderate variability in responses.

ANOVA Results

Constructs like CSI (F = 0.9285) and PSE (F = 1.5513) show higher F-statistics, indicating significant variance in customer satisfaction and problem-solving effectiveness.
Constructs such as EU (F = 0.3954) and RS (F = 0.3646) exhibit lower F-statistics, suggesting less variance in ease of use and response speed perceptions.

Reliability Analysis (Cronbach's Alpha)

	Cronbach Alpha	
EU	0.82	
RS	0.87	
ARI	0.78	
PCA	0.76	
CSI	0.81	
PSE	0.79	
CR	0.75	
TS	0.65	
IHS	0.8	

Validation of Objectives and Interpretation of Results:

The primary objective of this study was to explore how chatbot features influence Customer Satisfaction with Interaction (CSI). The results obtained from the analysis validate this objective by highlighting key insights:

1. **Ease of Use (EU):**The construct showed strong reliability and a significant positive impact on CSI, confirming that user-friendly chatbot interfaces enhance customer satisfaction.
2. **Response Speed (RS):**While RS demonstrated good reliability, its direct impact on CSI was weaker than expected, suggesting that speed alone may not be sufficient to drive satisfaction.
3. **Accuracy and Relevance of Information (ARI):**ARI had moderate reliability but a weaker influence on CSI, indicating that while accuracy is important, it may not be the primary driver of satisfaction.
4. **Personalization and Context Awareness (PCA):**PCA significantly influenced CSI, validating the importance of tailored interactions in improving customer experiences.
5. **Trustworthiness and Security (TS):**
   TS exhibited lower reliability and a limited impact on CSI, suggesting that customers may prioritize functionality over security in chatbot interactions.
6. **Integration with Human Support (IHS):** IHS positively influenced CSI, emphasizing the importance of seamless transitions to human agents for complex queries.

Discussion

The study findings align with prior research on chatbots' impact on customer satisfaction. Constructs such as Ease of Use, Personalization, and Integration with Human Support emerged as significant drivers of satisfaction, consistent with studies emphasizing usability and tailored interactions. However, constructs like Trustworthiness and Accuracy showed weaker direct impacts, suggesting that customers prioritize functionality over security in chatbot interactions. These results highlight the importance of refining chatbot features to balance efficiency and emotional engagement.

The implications are clear: businesses must focus on enhancing personalization and seamless transitions to human agents to improve customer satisfaction. Additionally, addressing limitations in trustworthiness and accuracy can further optimize chatbot effectiveness. This study contributes to the growing body of knowledge by validating the importance of usability, personalization, and hybrid support systems in AI-powered customer service.

Conclusion

This study successfully explored how chatbot features influence customer satisfaction. Key findings include:

1. Ease of Use, Response Speed, and Personalization significantly enhance satisfaction.
2. Integration with Human Support ensures smooth handling of complex queries.
3. Constructs like Trustworthiness require refinement to improve their impact on satisfaction.

These insights contribute to industry practices by emphasizing the need for personalized and efficient chatbot designs. The findings also provide actionable recommendations for businesses aiming to leverage AI chatbots for enhanced customer experiences.

Limitations and Future Recommendations

1. Sample Size: Although adequate for statistical analysis, a larger sample size across diverse demographics could improve generalizability.
2. The research focused primarily on respondents from urban areas in India; expanding the scope to rural areas or other countries could provide broader insights.
3. Construct Refinement: Constructs like Trustworthiness and Accuracy require further validation and refinement in future research.
4. Our research was mostly focused on our campus IIT KGP students. Very data was collected from outside the campus .

References

1. Intelemark Blog: *The Role of Chatbots in Enhancing the Customer Experience*
2. Kagwa C., *Effectiveness of Artificial Intelligence (AI) Chatbots in Improving Customer Satisfaction in E-Commerce*
3. Soni A., Dubey S., *The Impact of AI-Powered Chatbots on Customer Satisfaction in E-commerce Marketing (TAM Approach)*
4. Azaga A., *Effectiveness of Artificial Intelligence Chatbots for Customer Service*
5. Frontiers in Psychology: *Enhancing Customer Satisfaction with Chatbots: The Influence of Communication Style*
6. Smythos Blog: *Chatbots in Customer Service: Benefits and Implementation*
7. Acharya P., Shetty P., Sequiera P., *A Study on AI and Chatbots in Customer Experience*
8. Guvi Blog: *The Influence of Chatbots on Customer Services: Benefits and Impact*
9. JATIT Journal: *Chatbots for Customer Service*
10. IBM Think Blog: *Chatbots for Customer Experience*