Gurugram University Business Review (GUBR)
Volume 6 • Issue 1 • January–June 2026
ISSN: 2582-9718 (Online)

Assessing Users' Perception on Chatbots Effectiveness

A. Srikanth[1] & T. Mani Sri Anmitha[2]

ABSTRACT

This study looks at how K.L. Business School MBA students in Andhra Pradesh feel about voice-activated chatbots like Siri, Google Assistant, Amazon Alexa, and Bixby. It addresses the paucity of research on user assessments of chatbots performance by examining perceived efficacy, responsiveness, and overall value. Assessing general efficacy perceptions, comparing efficacy across chatbots kinds, determining user preferences, and examining the impact of demographics (gender, place of residence, and location) on perceptions and usage are the goals. A cross-sectional survey of 321 students chosen by proportionate stratified random sampling was utilized to gather data for the exploratory design. JASP, Excel, and R were utilized to analyse responses obtained via Google Forms.

The results show that consumers' opinions about chatbots efficacy are constant across chatbot types and demographic groups. Nonetheless, use patterns indicate brand domination, with Siri, Alexa, and Google Assistant becoming more popular. This implies that user choice may be influenced more by marketing, ecosystem integration, and brand exposure than by perceived performance.

Keywords: Chatbot, Alexa, Google Assistant, Siri, Bixby

INTRODUCTION

Chatbots, also known as conversational agents, are programs created to communicate with users using natural language via text, voice, or multimodal interfaces (Jalawadai, 2024). Processing user input and producing pertinent answers to voice or text enquiries is their main duty (Ghazala Bilquise, 2022). As artificial intelligence (AI) develops quickly, chatbots are using machine learning (ML) and natural language processing (NLP) more and more to mimic human interaction and improve communication accuracy. Large language models like GPT and Gemini have further changed chatbots from rule-based programs to context-aware, adaptive systems that can have dynamic, personalised discussions (Reddy, 2025). Voice-activated assistants, like as Google Assistant, Amazon Alexa, Siri, and Bixby, are now incorporated into smart phones, smart homes, and larger digital ecosystems, greatly impacting customer behaviour and service provision (Martin Adam, 2020). Depending on earlier studies, conversational AI systems with human-like characteristics increase user responsiveness and engagement (Ying Xu, 2022). Nevertheless, user opinions of chatbot efficacy continue to differ despite advancements in technology. Perceived performance and sustained engagement are still influenced by elements like trust, dependability, personalisation, and anthropomorphic design (Najaf & Ding, 2024).

NEED OF THE STUDY

Assessing how consumers see chatbots' efficacy is crucial as businesses use them more and more on digital channels. While technology developments have improved chatbot capabilities, sustained adoption and value are determined by user assessments of responsiveness, usefulness, dependability, and overall performance. In order to strategically improve chatbot design and interaction quality, it is imperative to comprehend these perceptions. In order to improve brand relationships, lower operating expenses, and

[1] Associate Professor, Business School, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.
E-mail: drakondi@kluniversity.in
[2] MBA(Final), Business School, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.
E-mail: anmithatunuguntla10@gmail.com

foster loyalty, customer pleasure is essential. While dissatisfied customers are more inclined to escalate to human agents and utilise self-service less frequently, satisfied users are more likely to trust, stick with, and refer chatbots. According to research, consistent and human-like chatbot interactions promote favourable brand perception and sustained engagement (Xu et al., 2022).On the other hand, unfavourable encounters lower returns on technology investments and erode trust in AI systems. Therefore, to improve user experience and guarantee long-term organisational results, a methodical evaluation of perceived chatbot performance is required.

LITERATURE REVIEW

According to (Adamopoulou, 2020), a chatbot is a computer software created to mimic human communication through spoken or typed exchanges, giving consumers automatic access to information and help. In general, chatbots can be divided into two categories: artificial intelligence-driven systems that employ machine learning and natural language processing (NLP) to comprehend user intent and provide adaptive responses, and rule-based systems that use rewritten scripts and decision trees (Dale, 2016). A more sophisticated type of these is voice-controlled systems, often known as voice assistants, which combine natural language processing (NLP) and automatic speech recognition (ASR) to enable smooth spoken communication between people and machines (Hoy, 2018). Prominent instances including Apple Siri, Google Assistant, Amazon Alexa, and Samsung Bixby show notable advancements in ecosystem integration, contextual awareness, and usability (Lopatovska, 2018).

Voice-activated chatbots have developed over the last ten years from simple speech-to-text tools to complex conversational agents with contextual awareness and tailored responses. More natural and dynamic interaction has been achieved by developments in ASR and NLP, which have also improved speech accuracy and cross-device interoperability. Important turning points in this history are marked by platforms like Siri, Alexa, Google Assistant, and Bixby, which are now essential components of smart speakers, smart phones, and connected home ecosystems (Darda, 2020). Their position in routine digital encounters has been reinforced by their extensive integration.

Depending on systematic reviews, voice assistants have become widely used because speech-based interaction is more convenient and intuitive than text input (Darda, 2020). These systems have revolutionised digital service access by allowing users to carry out tasks like scheduling, information retrieval, navigation, and home automation through natural voice. However, a large portion of the material now in publication places more emphasis on system performance, usability metrics, and technological architecture than on users' subjective evaluations of efficacy across platforms. As a result, nothing is known about how consumers compare the performance of the most popular commercial voice assistants in realistic settings.

According to research on user perceptions, interaction quality, personalisation, and trust have a significant effect on perceived efficacy. Perceptions of utility and credibility are influenced by conversational design characteristics like tone, responsiveness, clarity, and anthropomorphic qualities. Customer satisfaction ratings are generally greater for chatbots that react quickly, precisely, and contextually. These assessments are also influenced by demographic factors. While older users place greater emphasis on task efficiency and functional dependability, younger users frequently place more value on interactive elements and convenience. These variations underline the necessity of investigating how background, age, and experience affect opinions on chatbot efficacy (Pušnik, 2025).

Comparative research also shows that user expectations and usage context affect perceived performance. Voice assistants are exploited by many users for repetitive tasks, and opinions on their responsiveness and dependability differ depending on the user's familiarity with the device and past experiences (Sziron, 2022). People often personify their assistants, giving them human characteristics that influence their propensity to trust them and delegate difficult jobs. Despite these revelations, there is always a dearth of direct comparative studies across Siri, Alexa, Google Assistant, and Bixby across various demographic groups.

There are still significant gaps in the literature, despite the fact that chatbot evolution and usability models

are covered in great detail. Instead of concentrating only on voice-activated assistants, numerous studies also examine text-based systems or overall interaction quality. There aren't many comparative studies of the main viable platforms, especially in formal academic settings. Furthermore, rather than being a direct result of perceived efficacy across platforms, demographic factors are frequently associated with adoption. In order to address these gaps, the current study integrates demographic data to provide a thorough understanding of how effectiveness is defined and assessed across platforms and user groups. It also looks at user perceptions of chatbot effectiveness across top voice assistants within a single framework.

RESEARCH GAP

The majority of studies on voice-activated chatbots, including Samsung Bixby, Apple Siri, Google Assistant, Amazon Alexa, IBM Watson, and others, have mostly concentrated on usability, user adoption, and technological advancement. Few research, nevertheless, thoroughly investigate how users perceive chatbot efficacy. It is unknown how chatbot type affects perceived performance because existing research frequently assesses specific platforms rather than comparing several assistants within a single analytical framework. While demographic factors like age, gender, and education are often associated with adoption, they are rarely examined in terms of perceived efficacy. Furthermore, neither the distinction between platform-specific and common user experiences nor the analysis of the ways in which various demographic characteristics interact to influence impressions have received ample attention. There are still insufficient comparative analyses of the leading commercial voice assistants. Therefore, a systematic assessment integrating platform comparison and demographic impact is necessary, which this study aims to provide from a consumer perception perspective.

OBJECTIVES OF THE STUDY

The following are the objectives formulated from the gaps identified by the study.

1. To assess users' perception towards chatbot effectiveness.
2. To examine the effect of type of chatbots onchatbot effectiveness from consumers' perception. (Unique Responses)
3. To analyse chatbot effectiveness among all the chatbots. (Repeated Responses)
4. To study the influence of consumers' demographics on chatbot effectiveness.
5. To examine the association between respondents' demographics and their choice of chatbots.

HYPOTHESES

The following are the alternative hypotheses formulated in-line with the above stated objectives.

$H_1$1: Users perception towards chatbot effectiveness is positive.

$H_1$2: Users' perception towards chatbot effectiveness varies by type of chatbot.(Where respondents' response to the most preferred chatbot is considered – Single Response)

$H_1$3: The preference of consumers towards choosing chatbots is not the same. (Where respondents' response for all chatbots is considered – Repeated response)

$H_1$4: There is a significant impact of demographics on chatbot effectiveness.

$H_1$5: There is an association between chosen demographic variables (Gender, Location and Residential status) and their choice of chatbots.

There are sub-hypothesis for 4th and 5th hypothesis, they are mentioned in results and discussion section.

METHODOLOGY

This section helps in meeting the objectives one by one to understand the respondents' perception on Chatbots Effectiveness. So, this includes the Research Design, Sample Design, Data Collection and Data Analysis adopted by the study.

Research Design

The study has used mixed methodological research design including Exploratory Research Design and Descriptive Research Design, wherein Literature review method of Exploratory Research Design is used for understanding the theoretical and empirical studies on the research topic and helps in identifying

the research gap & Cross-Sectional Survey method of Descriptive Research Design is used in collecting the data.

Sample Design and Data Collection

In this design of study, target population are youngsters using voice-activated chatbots to know their perception on their effectiveness, particularly MBA pupils at KL Business School, KLEF. A sample of 175 from a sample frame of 321, is collected using a structured Questionnaire with likert items along with demographic variables including Gender, Location and Residential Status through Google Forms from randomly selected students using Proportionate Stratified Random Sampling (Singh, 2014).

Table 1: Sample Determination Table using Cochran Formulae

Section	*Population*	*Population%*	*Sample%*	*Sample*
Sec A	79	24.61	24.61	43
Sec B	68	21.18	21.18	37
Sec C	47	14.64	14.64	26
Sec D	59	18.38	18.38	32
Sec E	68	21.18	21.18	37
Total	321	100	100	175

To determine the sample elements from each section, sample () in R is used, which is from the base package.

Data Analysis

The study includes two phases of analysis including pilot study and final study. Initially, Questionnaire, a survey instrument is tested for its validity and reliability as a part of Pilot study. Under validity, face validity of instrument is conducted by two industry professionals and five academic members to guarantee the questionnaire's relevance and intelligibility. In addition, Construct validity is also conducted to check the factor validity of Chatbots Effectiveness. To check the reliability of Instrument, Cronbach's alpha is analysed. Response bias is examined by using a chi-square test after data cleaning was completed in R.

As a part of final study, several statistical tools are adopted by the study. For the first objective, Wilcoxon Signed Rank Test is used to measure the Chatbots Effectiveness (i.e., Users' satisfaction scores), as the data is not-normally distributed (Nahm, 2015).

To meet the second, One-way ANOVA is used to examine the effect of type of chatbots (five – AmazonAlexa, Apple Siri, Samsung's Bixby, IBM's Watson and Google's Google Assistant) on Users' Satisfaction score towards Chatbots Effectiveness, being type of chatbots is acting as Independent variable with five categories and Chatbots Effectiveness as dependent variable (María José Blanca, 2017).

For the third, the Friedman Test is used as the data is ranked across the five chatbots (Nahm, 2015).In the fourth Objective, the influence of respondent's demographics on their Satisfaction score towards Chatbots effectiveness is assessed using respective statistical tools namely Mann-Whitney U test for Gender and Residential Status, One-Way ANOVA for location and choice of chatbots on time-spent by respondents (Nahm, 2015). Finally, the fifth objective assess the association between respondents' demographics (i.e., Gender, Residential Status and Location) and their Chatbots Selection (McHugh, 2013) using Chi-square test.

RESULTS AND DISCUSSION

After conducting pilot study analysis, the study further proceed with final study with 16 statements and the conducted reliability analysis is as follows,

Table 2: Reliability Score

Measure	*Value*	*Remarks*
Raw Alpha	0.96	>0.5, reliable.

Table 3: Threshold Values

Method	*Alpha*
Feldt	0.96
Duhachek	0.96

Source: R
Package: psych
Function: alpha ()

From the above Table 3, "Cronbach's alpha of 0.96, the questionnaire exceeded the 0.90 level for excellent internal consistency, indicating outstanding reliability" (Nunnally, 2016). The questionnaire is a very trustworthy tool because of its narrow 95% confidence intervals from Feldt's and Duhachek's methodologies (0.95–0.97), which validate the stability and accuracy of the reliability estimate (Nunnally, 2016).

Overall, the results establish that the questionnaire is a highly reliable instrument. With these results, the study further proceed for final study.

For **data cleaning** the study has usedthe following packages and functions in R:

Base Package- colSums(),sum(),is.na()
VIM Package-irmi()

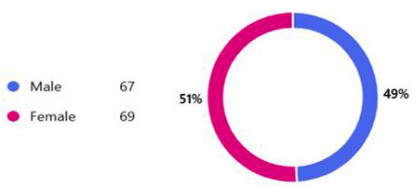
The **descriptive analysis** of data for this study is as follows:

- Categorical Data:
  Gender (2) – Male, Female
  Location (4) – Vijayawada, Guntur, Tenali, Others
  Residential status (2) – Dayscholar, Hosteler
  Chatbots (4) – Amazon_Alexa, Google_Assistant, Siri, Bixby
- Continuous Data:
  Score – Satisfaction from the Likert scale
  Time – No. of hours spent on the selected chatbot.

Bias Justification for data collected:

H_o: Proportions are equal.
H_1: Proportions are not equal.

1. Gender



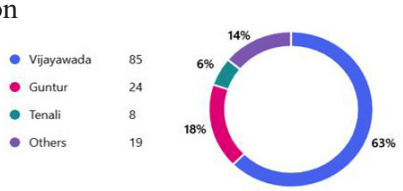

```
> chisq.test(c(51,49),p=c(1/2,1/2))

        Chi-squared test for given probabilities

data:  c(51, 49)
X-squared = 0.04, df = 1, p-value = 0.8415
```

P-value is 0.8145 greater than 0.05, acceptnull hypothesis can conclude that there is no bias (Stefanos Bonovas 1, 2023).

2. Location



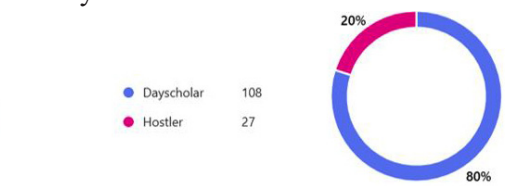

```
> chisq.test(c(63,18,6,14),p=c(1/4,1/4,1/4,1/4))

        Chi-squared test for given probabilities

data:  c(63, 18, 6, 14)
X-squared = 78.208, df = 3, p-value < 2.2e-16
```

P-value is 2.2e-16 less than 0.05, reject null hypothesis and conclude that proportions are not equal (Stefanos Bonovas 1, 2023).

3. Are you a



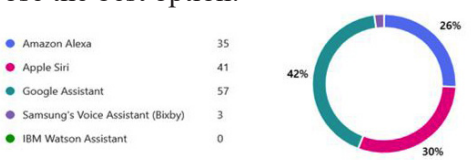

```
> chisq.test(c(80,20),p=c(1/2,1/2))

        Chi-squared test for given probabilities

data:  c(80, 20)
X-squared = 36, df = 1, p-value = 1.973e-09
```

P-value is 1.973e-09 less than 0.05, reject null hypothesis and can say that proportions are not equal (Stefanos Bonovas 1, 2023).

4. From the following Voice Activated Chatbots, choose the best option:



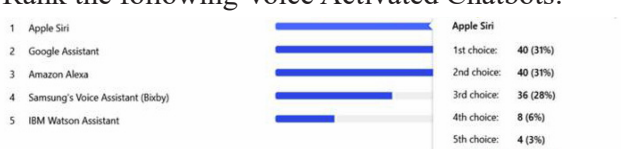

```
> chisq.test(c(30,42,26),p=c(1/3,1/3,1/3))

        Chi-squared test for given probabilities

data:  c(30, 42, 26)
X-squared = 4.2449, df = 2, p-value = 0.1197
```

For analysis the study removed Samsung's Voice Assistant (Bixby) and IBM Watson Assistant for some tests to remove bias. P-value is 0.1197 greater than 0.05,accept null hypothesisand can say that there is no biasin Voice activated chatbots which are Amazon_ Alexa, Apple_Siri, Google_Assistant. (Stefanos Bonovas 1, 2023)

5. Rank the following Voice Activated Chatbots:




```
> chisq.test(c(40,40,36,8,4),p=c(1/5,1/5,1/5,1/5,1/5))

        Chi-squared test for given probabilities

data:  c(40, 40, 36, 8, 4)
X-squared = 50.75, df = 4, p-value = 2.518e-10
```

P-value is 2.518e-10 less than 0.05,reject null hypothesis and can say that proportions are not equal (Stefanos Bonovas 1, 2023).

For analysis the study removed Samsung's Voice Assistant (Bixby) and IBM Watson Assistant for some tests to remove bias.

```
> chisq.test(c(40,40,36),p=c(1/3,1/3,1/3))

        Chi-squared test for given probabilities

data:  c(40, 40, 36)
X-squared = 0.27586, df = 2, p-value = 0.8712
```

P-value is 0.8712 greater than 0.05,accept null hypothesisand can say that there is no bias in Voice activated chatbots which are Amazon_Alexa, Siri, Google_Assistant (Stefanos Bonovas 1, 2023).

The analysis shows no bias in the selection of the main chatbots or in gender distribution. Some bias appears in location and when less prominent chatbots are included, leading to their exclusion from certain assessments. Overall findings remain consistent, and future research should use larger, more geographically balanced samples.

The following results are in-line with the above-stated objectives.

From Objective 1

$H_0$1: Users perception towards chatbot effectiveness (Score) is not positive. [mu <= 3 (shows dissatisfaction)]

$H_1$1: Users perception towards chatbot effectiveness (Score) is positive. [mu>3 (shows satisfaction)]

Code Snippet (Source: R)

> shapiro.test(Score)

p-value = 0.0002847

As the p-value is < 0.05, **proceed for wilcoxon signed rank test** (Nahm, 2015)**.**

Table 4: Wilcoxon Signed Rank Test Results

Variable_name	*Test*	*p*	*Remarks*
Score	Wilcoxon	0.123	> 0.05, accept null hypothesis

Source: Jeffreys's Amazing Statistics Program (JASP)

Interpretation

From the above table 4, as the p-value is greater than 0.05, accept null hypothesis and can say that users are dissatisfied with all the functions of chatbots (Stefanos Bonovas 1, 2023).

Murray (2024) reports a sharp fall in trust, with usage dropping from 73% to 60% in just over a year. This decline confirms that current chatbots fail to meet user expectations of reliability and performance.

From Objective 2

$H_0$2: Users' perception towards chatbot effectiveness does not vary by type of chatbot. [(meanScore(Amazon Alexa)=meanScore(AppleSiri)=meanScore(Google Assistant)]

$H_1$2: Users' perception towards chatbot effectiveness varies by type of chatbot. [(meanScore(Amazon Alexa)#meanScore(AppleSiri)#meanScore(Google Assistant)]

Table 5: ANOVA Results

Variables	*Test*	*p*	*Remarks*
Dependent Variable – Score Independent Variable – Chatbots	ANOVA	0.938	> 0.05, accept null hypothesis

Source: JASP

Interpretation

From the above table 5, as the p-value is 0.938 greater than 0.05,accept null hypothesis and conclude that Users' perception is not influenced by type of chatbot (Stefanos Bonovas 1, 2023).

Customers rated Amazon Alexa, Google Assistant, and Siri similarly for user pleasure and chatbot efficacy, suggesting that platform or brand differences have little bearing on perceived performance. Customer perceptions are influenced more by performance expectations and trust than by brand identity, and competitive advantages are instead provided by elements such ecosystem integration, marketing, and value-added features (Joshi, 2021).

From Objective 3

$H_0$3: The preference of consumers towards choosing chatbots is same. [median(Amazon_Alexa)=median(Google_Assistant)=median(Siri)=median(Bixby)=median(IBM)]

$H_1$3: The preference of consumers towards choosing chatbots is not the same. [median(Amazon_Alexa) #median(Google_Assistant)#median(Siri)#median (Bixby) #median(IBM)]

Code Snippet (Source: R)

>friedmanTest(y=outcome, groups=treatment,blocks= participants)

(Nahm, 2015)

p-value = 2.2e-16

Interpretation

As the p-value is less than 0.05, reject null hypothesis which concludes that median of all the chatbots are different (Stefanos Bonovas 1, 2023).

POST-HOC Test

Code Snippet (Source :R)

>frdAllPairsMillerTest(y=outcome,groups=treatment, blocks=participants)

(Nahm, 2015)

1 2 3 4

2 1 - - -

3 1 1 - -

4 1.6e-08 1.2e-08 1.1e-09 -

5 < 2e-16 < 2e-16 < 2e-16 2.9e-07

Interpretation

As group-wise clear differences are visible,there were no discernible differences between the Google Assistant-Amazon Alexa, Siri-Amazon Alexa, Bixby-Amazon Alexa, and Siri-Google Assistant groups.

But Bixby was very different from Siri, Google Assistant, and Amazon Alexa.

When IBM was compared to all other chatbots (Amazon Alexa, Google Assistant, Siri, and Bixby), the differences were extremely noticeable.

User evaluations of efficacy are consistent across chatbot kinds, indicating that performance expectancy and trust are more important than brand labels (Joshi, 2021).

Bixby/IBM lags behind because of chatbots that have a bad communication style or a limited social presence, as demonstrated by (Na Cai, 2024).

From Objective 4

Gender on Users' perception on chatbots effectiveness (Score)

$H_0$4.1: Gender does not affect the Score. [(median Score(f)=medianScore(m)]

$H_1$4.1: Gender affects the Score. [medianScore(f) #medianScore(m)]

Code Snippet (Source :R)

> f<-subset(Score,Gender==’Female’)

> m<-subset(Score,Gender==’Male’)

> shapiro.test(f)

p-value = 0.002011

> shapiro.test(m)

p-value = 0.05369

As the p-value for independent variable Gender in which female is 0.002011 which is less than 0.05,conclude that data is not normal, proceed for Mann Whitney U Test (Nahm, 2015).

Table 6: Mann Witney U Test Results

Variable_name	*Test*	*p*	*Remarks*
Score	Mann-Whitney	0.347	> 0.05, accept null hypothesis

Source: JASP

Interpretation: From the above table 6, as the p-value is 0.347 greater than 0.05, accept null hypothesiswhich says that Gender does not affect the Score (Stefanos Bonovas 1, 2023). Users of both sexes have comparable opinions about chatbots. Companies can concentrate on creating and promoting gender-neutral services. Language and technology traits have a bigger influence on satisfaction than demographics (Kokošková, 2023).

Location on Users' Perception on Chatbots Effectiveness (Score)

$H_0$4.2: The Score does not vary by location. [(meanScore(Vijayawada)=meanScore(Guntur)= meanScore(Tenali)]

$H_1$4.2: The Score varies by location. [(meanScore (Vijayawada)#meanScore(Guntur)#meanScore (Tenali)]

Table 7: ANOVA Results

Variables	*Test*	*p*	*Remarks*
Dependent Variable – Score Independent Variable –Location	ANOVA	0.939	> 0.05, accept null hypothesis

Source: JASP

Interpretation: From the above Table 7, as the p-value is 0.939 greater than 0.05,acceptnull hypothesiswhich

saysthat the users' perception does not vary by location (Malcolm Koo 1, 2025).Chatbots received similar ratings from users in Vijayawada, Guntur, Tenali, and other areas. Although local language integration may improve inclusivity, businesses can grow chatbots geographically without losing perception.Adoption is more influenced by trust and transparency than by geography or demographics (John, 2025).

Residential Status on Users' perception on chatbots effectiveness (Score).

$H_0$4.3: There is no impact of Residential Status on Score [medianScore(D)=medianScore(H)]

$H_1$4.3: There is no impact of Residential Status on Score [medianScore(D)#medianScore(H)]

Code Snippet (Source :R)

> D<-subset(Score,DH==’Dayscholar’)

> H<-subset(Score,DH==’Hostler’)

> shapiro.test(D)

p-value = 0.00126

> shapiro.test(H)

p-value = 0.05232

As one of the groups data which is Dayscholar p-values less than 0.05 reject null hypothesis and conclude that data is not normal, proceed with Mann Whitney U Test (Nahm, 2015).

Table 8: Mann Witney U Test Results

Variable_name	*Test*	*p*	*Remarks*
Score	Mann-Whitney	0.239	> 0.05, accept null hypothesis

Source: JASP

Interpretation: From the above table 8, as the p-value is greater than 0.05,acceptnull hypothesiswhich says that Residential Status has no influence on the perception of the users (Stefanos Bonovas 1, 2023).

There is no discernible difference between day students' and hostelers' opinions of chatbot effectiveness, suggesting that student demographics have adopted them consistently. This implies that rather than segment-specific modification, chatbot development can concentrate on functional perfection.Linguistic and technological characteristics, not demographic factors like living arrangement, are what influence user happiness with chatbots (Kokošková, 2023).

Choice of Chatbots onNo. of Hours Spent by the User (Time Spent -TI)

$H_0$4.4: The choice of chatbot does not affect the no. of hours spent by User. [(meanTimeSpent(Amazon Alexa)=meanTimeSpent(AppleSiri)=meanTime Spent(Google Assistant)]

$H_1$4.4: The choice of chatbot affects the no. of hours spent by User. [(meanTimeSpent(Amazon Alexa)#meanTimeSpent(AppleSiri)#meanTime Spent(Google Assistant)]

Code Snippet (Source :R)

Test Homogeneity:

H_0: Homogeneity exists

H_1: Heterogenity exists

>bartlett.test(TI~Chatbots,data=dt)

p-value = 7.152e-08

\#As the p-value is very much less than0.05,reject null hypothesis and hence conclude that heterogeneity exists (Stefanos Bonovas 1, 2023).

Independence of observations presumed.

Table 9: ANOVA Results

Variables	*Test*	*p*	*Remarks*
Dependent Variable- No. of Hours Spent by the User (Time Spent – TI) Independent Variable – Chatbots	ANOVA	0.328	> 0.05, accept null hypothesis

Source: JASP

Interpretation: From the above table 9, as the p-value is greater than 0.05,accept the null hypothesiswhich concludes that the choice of chatbot does not affect the no. of hours spent (Stefanos Bonovas 1, 2023). Individuals use Google Assistant, Siri, and Alexa for comparable periods of time, demonstrating functional parity in usage duration. Instead of investing effort in differentiating chatbot performance, businesses can concentrate on task fulfillment, engagement quality, and personalization.

Residential Status on No. of Hours Spent by the User (Time Spent – TI)

$H_0$4.5: Residential Status does not affect the no. of time spent [medianTimeSpent(D)=medianTimeSpent(H)]

$H_1$4.5: Residential Status affects the no. of time spent [medianTimeSpent(D)#medianTimeSpent(H)]

Code Snippet (Source :R)

>D<-subset(TI,DH==’Dayscholar’)

> H<-subset(TI,DH==’Hostler’)

> shapiro.test(D)

p-value = 9.678e-06

> shapiro.test(H)

p-value = 1.982e-08

> #As the p-values of groups are less than 0.05,data is not normal.

Proceed for Mann Whitney U Test (Nahm, 2015).

Table 10: Mann Witney U Test Results

Variables	*Test*	*p*	*Remarks*
Independent Variable – Residential Status Dependent Variable – No. of Hours Spent by the User (Time Spent – TI)	Mann-Whitney	0.923	> 0.05, accept null hypothesis

Source: JASP

Interpretation: From the above table 10, as the p-value is 0.923 which is greater than 0.05,acceptnull hypothesis which says that Residential Status does not have any effect on time spent (Stefanos Bonovas 1, 2023).

Students use chatbots in the same way regardless of their living arrangement, thus whether they are day students or hostelers has no bearing on how much time they spend using them.Human-like characteristics, trust, and interaction have a greater influence on chatbot adoption in e-commerce than demographic characteristics (Najaf & Ding, 2024).

Gender on No. of Hours Spent by the User (Time Spent – TI)

$H_0$4.6: Genderdoes not affectthe no. of hours time spent. [mediandTimeSpent(f)=medianTimeSpent(m)]

$H_1$4.6: Gender affects the no. of hours time spent. [mediandTimeSpent(f)#medianTimeSpent(m)]

Code Snippet (Source :R)

> f<-subset(TI,Gender==’Female’)

> m<-subset(TI,Gender==’Male’)

> shapiro.test(f)

p-value = 2.944e-11

> shapiro.test(m)

p-value = 0.00058

> #As the p-values of groups are less than 0.05, data is not normal, proceed for Mann Whitney U Test (Nahm, 2015).

Table 11: Mann Witney U Test Results

Variables	*Test*	*p*	*Remarks*
Independent Variable – Gender Dependent Variable – No. of Hours Spent by the User (Time Spent – TI)	Mann-Whitney	0.168	> 0.05, accept null hypothesis

Source: JASP

Interpretation: From the Table 11, as the p-value is greater than 0.05, accept null hypothesiswhich says that the timespent on the selected chatbot does not differ by gender (Stefanos Bonovas 1, 2023).Instead of gender-specific customisation, developers should focus on universal attributes (easiness, clarity,and trust). Trust and past experience outweigh demographic factors like gender (Kozak & Fel, 2024)

Location on No. of Hours Spent by the User (Time Spent – TI)

$H_0$4.7: Location does not make any difference on time spent with respective to chosen chatbot. [(meanTimeSpent(Vijayawada)=meanTimeSpent (Guntur)=meanTimeSpent(Tenali)]

$H_1$4.7: Location makes difference on time spent with respective to choosen chatbot. [(meanTimeSpent (Vijayawada)#meanTimeSpent(Guntur)#meanTime Spent(Tenali)]

Code Snippet (Source :R)

>#Ho: Homogenity exists

>#H1:Heterogenity exists

>bartlett.test(TI~Location,data=dt)

p-value = 0.004924

> #As the p-value is less than 0.05, reject null hypothesisand hence Heterogenity exists. (Stefanos Bonovas 1, 2023)

> #Independence of observations presumed.

Table 12: ANOVA Results

Variables	*Test*	*p*	*Remarks*
Independent Variable – Location Dependent Variable – No. of Hours Spent by the User (Time Spent – TI)	ANOVA	0.835	> 0.05, accept null hypothesis

Source: JASP

Interpretation: From the Table 12, as the p-value is 0.835which is greater than 0.05, accept the null hypothesis which concludes that Location does not make any difference on time spent with respective to selected chatbot (Stefanos Bonovas 1, 2023).

Chatbots were used for comparable amounts of time by students in various cities. This implies that chatbot adoption is stable across regions in comparable socio-cultural contexts, enabling developers to put functionality ahead of regional customisation. Adoption is more strongly influenced by trust and transparency than by geography (Colpan, 2024).

From Objective 5

Gender on Choice of Chatbots

$H_0$5.1: Users' Gender has no connection with their choice of chatbots.

$H_1$5.1: Users' Gender has no connection with their choice of chatbots.

Table 13: Contingency Table Results

Variables	*Test*	*p*	*Remarks*
Variable 1 – Gender Variable 2 – Chatbots	Chi-Squared Test (χ^2)	0.791	> 0.05, accept null hypothesis

Note: Continuity correction is available only for 2×2 tables.
Source: JASP

Interpretation: From the above table 16, as the p-value is 0.0791, which is greater than 0.05, accept null hypothesis which concludes that gender has no connection with the chatbot chosen by the users (Stefanos Bonovas 1, 2023).

Residential Status on Choice of Chatbots

$H_0$5.2: Choice of Chatbots does not depend on Users' Residential Status.

$H_1$5.2: Choice of Chatbots depends on Users' Residential Status.

Table 14: Contingency Table Results

Variables	*Test*	*p*	*Remarks*
Variable 1 – Dayscholar/ Hosteler Variable 2 – Chatbots	Chi-Squared Test (χ^2)	0.926	> 0.05, accept null hypothesis

Note: Continuity correction is available only for 2×2 tables.
Source: JASP

Interpretation: From the above table 17, as the p-value is 0.926, which is greater than 0.05, accept null hypothesis which says that chatbots choice does not depend on day scholar or hosteler (Stefanos Bonovas 1, 2023).

Chatbots were similarly selected by day scholars and hostel residents, demonstrating that brand preference is not influenced by residential setting. Chatbot adoption patterns are explained by trust and interaction rather than demographics (Najaf & Ding, 2024).

Location on Choice of Chatbots

$H_0$5.3: Users' Location and their Choice of Chatbots are not related.

$H_1$5.3: Users' Location and their Choice of Chatbots are related.

Table 15: Contingency Table Results

Variables	*Test*	*p*	*Remarks*
Variable 1 – Location Variable 2 – Chatbots	Chi-Squared Test (χ^2)	0.397	> 0.05, accept null hypothesis

Note: Continuity correction is available only for 2x2 tables.
Source: JASP

Interpretation: From the above table 18, as the p-value is 0.397 which is greater than 0.05accept null hypothesis, concludes that location and selection of chatbots are not related (Stefanos Bonovas 1, 2023).

Similar chatbot preferences were displayed by students in Vijayawada, Guntur, Tenali, and other locations, suggesting that brand adoption is consistent across geographic boundaries. Adoption is more strongly influenced by trust and transparency than by demographics based on geography (Colpan, 2024).

FINDINGS OF THE STUDY

1. Users' perception towards chatbots effective is not much positive.

2. Users' perception on chatbots effectiveness is same irrespective of the type of chatbots (For the best Chosen Chatbots from all the respondents).
3. Users' preferences towards choosing chatbots vary from those using all the chatbots in terms of efficacy, trust and communication style aspects.
4. Users' Gender, Location and their Residential Status are not affecting Chatbots Effectiveness as they have no role in the choice of those chatbots.
5. Users' Time spent on Chatbots does not depend on the choice of chatbots.
6. Even, User's demographics namely Gender, Location and their Residential Status have no role in their time spent on their chosen chatbots.

IMPLICATIONS OF THE STUDY

Companies can use these findings to improve chatbot functions and increase overall user satisfaction. Since Bixby and IBM Watson show lower user engagement compared to Google Assistant, Amazon Alexa, and Siri, these companies should focus on improving brand awareness and device integration. As user perceptions do not vary by gender or location, chatbot improvements should follow gender-neutral and location-neutral approaches.

FUTURE SCOPE OF RESEARCH

This paper provides a foundational overview of chatbots and users' perception towards voice activated chatbots. For further exploration, consider delving deeper (Yuan et al., 2024) (Yuan et al., 2024) into specific areas like the Impact of Voice-Activated Chatbots on Elders. Change the target audience and analyse the users' perception to obtain most robust results. Use other statements in Likert scale for more variables and analyse deeper.

REFERENCES

Adamopoulou, E. (2020). *Chatbots: History, technology, and applications* (Vol. 584). Springer.

Bertaria Sohnata Hutauruk, R. P. (2024). The Effectiveness of Artificial Intelligence by Chatbot in Enhancing the Students' Vocabulary. *ResearchGate*.

Colpan, E.W. (2024). Reconciling the tension between contextualization and generalization in qualitative international business research. *Journal of International Business Studies, 34*(2), 1–15.

Dale. (2016). The return of the chatbots. *Cambridge University Press, 22*(5), 811–817.

Daniel Maxwell, B. O. (2025). Generative AI in Higher Education: Demographic Differences in Student Perceived Readiness, Benefits, and Challenges. *Springer Nature*.

Darda, P. (2020). Voice Assistant: A Systematic Literature Review. *ResearchGate*, 1–8.

Dutta, S. (2024). *Chatbot Effectiveness in Enhancing Guest Communication: Insights from Secondary Data.* Tuijin Jishu.

Escobar-Grisales, D. (2023). Evaluation of effectiveness in conversations between humans and chatbots using parallel convolutional neural networks with multiple temporal resolutions. *Springer Nature*.

Ghazala Bilquise, S. I. (2022). Emotionally Intelligent Chatbots: A Systematic Literature Review. *Human Behavior and Emerging Technologies*, 1-23.

Gomes, S. L. (2025). Anthropomorphism in artificial intelligence: a game-changer for brand marketing. *Volume 11, article number 2*.

Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Taylor & Francis, 37*(1), 81–88.

Hussain, K. (2025). Humanizing generative AI Brands: How brand anthropomorphism converts customers into brand heroes. *ScienceDirect, Volume 19*.

Hutauruk. (2024). *The effectiveness of artificial intelligence by chatbot in enhancing the students' vocabulary* (Vol. 6). Journal of English Teaching and Applied Linguistics.

Jalawadai, M. (2024). Chatbots: Conversational AI Shaping Our World. *International Journal of Creative Research Thoughts(IJCRT), 12*(4), 142–146.

John, B. (2025). The Role of Trust in AI-Powered Chatbots: Impact on User Adoption and Retention. *Computers in Human Behavior* .

Joshi, H. (2021). Perception and Adoption of Customer Service Chatbots among. *Proceedings of the 17th International Conference on Web Information Systems and Technologies - Volume 1: WEBIST, 10*(3).

Jyosthna, M., Venkata Subbaiah, P., & Natalia, K. (2024). Exploring the Chatbot usage intention-A mediating role of Chatbot initial trust. *Research Gate*.

Kikani, D., & Ramchandani, S. M. (2025). *The Impact of AI Chatbots on Customer Service: Efficiency VS Human Touch* (Vols. volume10, Issue3). IJRTI.

Kokošková, J. B.-M. (2023). Integrating chatbots in education: insights from the Chatbot-Human Interaction Satisfaction Model (CHISM). *SpringerOpen, 20*(62), 1-14.

Kozak, J., & Fel, S. (2024). How sociodemographic factors relate to trust in artificial intelligence among students

in Poland and the United Kingdom. *Scientific Reports, 14*(28776), 28776.

Lopatovska, I. (2018). Personification of the Amazon Alexa: BFF or a Mindless Companion. *ACM Association for Computing Machinery corporate, 55*(1), 1–10.

Magno, F. (2023). *The effects of chatbots' attributes on customer relationships with brands: PLS-SEM and importance–performance map analysis* (Vols. Volume 35, Issue 5). Emerald Publishing.

Mahima Anna Varghese1, O. I., Poonam Sharma1, O. I., & Maitreyee Patwardhan2, O. I. (2024). *Public Perception on Artificial Intelligence–Driven Mental Health Interventions: Survey Research* (Vol. 8). JMIR Publications.

Malcolm Koo 1, 2. a.-W. (2025). Likert-Type Scale. *SN Comprehensive Clinical Medicine*.

María José Blanca, R.A. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema, 29*(4), 552–557.

Martin Adam, M. W. (2020). AI-based chatbots in customer service and their effects on user compliance, (pp. 1–6). Association for Information Systems.

McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica, 23*(2), 143–149.

Murray, S. (2024). *GenAI and Voice Assistants: Adoption and Trust Across Generations.* Deloitte.

Na Cai, S. G. (2024). How the communication style of chatbots influences consumers' satisfaction, trust, and engagement in the context of service failure. *Computers in Human Behavior, 11*(687), 1–15.

Nahm, F. S. (2015). Nonparametric statistical tests for the continuous data: the basic concept and the practical use. *Korean Journal of Anesthesiology, 68*(1), 8–14.

Najaf, M., & Ding, Y. (2024). *Interactivity, humanness, and trust: a psychological approach to AI chatbot adoption in e-commerce* (Vol. 146). Elsevier.

Nunnally, J. C. (2016). *Understanding Coefficient Alpha: Assumptions and Interpretations.* McGraw-Hill.

Pušnik, M. (2025). Differences in User Perception of Artificial Intelligence-Driven Chatbots and Traditional Tools in Qualitative Data Analysis. *Computers in Human Behavior Reports*.

Reddy, N. T. (2025). A Modular Retrieval-Augmented Conversational AI Chatbot System with Integrated Recommender Engine Using Local LLMs. *2*.

Singh. (2014). Sampling techniques and determination of sample size in applied statistics research: An overview. *International Journal of Economics, Commerce and Management, 2*(11), 1–7.

Singh, A.S. (2014). *Sampling Techniques & Determination of Sample* (Vols. vol. II, Issue 11). International Journal of Economics, Commerce and Management.

Stefanos Bonovas 1, 2. a. (2023). On p-Values and Statistical Significance. *Hippokratia, 27*(1), 1–4.

Stöhr, C. (2024). *Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study.* Computers and Education: Artificial Intelligence.

Sziron, M. (2022). Investigating user perceptions of commercial virtual assistants: A qualitative study. *SN Computer Science, 3*(6), 1–12.

Trupti Rathi1, D.B. (2022). *Questionnaire as a Tool of Data Collection in Empirical Research* (Vol. 6). Journal of Positive School Psychology.

Xu, Y., Zhang, J., Chi, R., & Deng, G. (2022). Enhancing customer satisfaction with chatbots: The influence of anthropomorphic communication styles and anthropomorphised roles. Volume 14 (Issue 2).

Ying Xu, J. Z. (2022). Enhancing customer satisfaction with chatbots: The influence of communication styles and consumer attachment anxiety (pp. 1-10). Association for Information Systems.

Yuan, L., Sun, T., Dennis, A., & Zho, M. (2024). *Perception is reality? Understanding user perceptions of chatbot-inferred versus self-reported personality traits.* Computers in Human Behavior: Artificial Humans.

ANNEXURE – QUESTIONNAIRE

Questionnaire on Voice-Activated Chatbots

1. Gender
   (a) Male
   (b) Female
2. Location
   (a) Vijayawada
   (b) Guntur
   (c) Tenali
   (d) Others
3. Provide your Residential Status
   (a) Dayscholar
   (b) Hosteller
4. Number of hours spent on these voice-activated chatbots. (e.g.: 1 hour spent = 1) ...
5. From the following voice-activated chatbots, choose the best option.
   (a) Amazon Alexa
   (b) Apple Siri
   (c) Google Assistant
   (d) Samsung's Voice Assistant (Bixby)
   (e) IBM Watson Assistant
6. Provide your agreement with the following statements that focus on usage of the chosen product that measures Chatbots Effectiveness.

Likert Scale used in Questionnaire.

S. No.	*Statements*	*1*	*2*	*3*	*4*	*5*
1	More voice clarity					
2	Takes less instructions					
3	Takes less response time					
4	Provides more accurate results					
5	Validates or accepts maximum results					
6	Can easily operate other devices					
7	Can perform multitask routines					
8	Compatible with every device					
9	It remembers instructions given					
10	Hands-free calling and messaging					
11	Sets reminders effortlessly					
12	Schedules meetings effortlessly					
13	Gets directions and traffic live reports					
14	Online shopping and orders					
15	Health and fitness assistance.					
16	Helps in language translation.					

1 – Strongly Disagree; 2 – Disagree; 3 – Neutral; 4 – Agree; 5 – Strongly Agree.

7. Rank the following voice activated chatbots.
   (a) Amazon Alexa
   (b) Apple Siri
   (c) Google Assistant
   (d) Samsung's Voice Assistant (Bixby)
   (e) IBM Watson Assistant

Thank you for sparing your valuable time in filling this survey form.