AI Chatbot that Scrapes Website Information for

New Answers and Creates Articles

by

Junayd Lateef

Thesis Director: Dr. Maria Elena Chavez Echeagaray

Second Committee Member: Mr. Dave Burnett

Ira A. Fulton School of Engineering

Barrett, the Honors College

Arizona State University

December 2024

# Abstract

During the course of this study, we were curious about what some potential factors may be when it comes to an individual rating the performance of an AI chatbot. This curiosity stems from the idea of implementing a chatbot to help users to quickly get information or to navigate a website they are on. The sponsor of the project, Mr. Dave Burnett, also wanted to test out responses that were article length to see if there was an increase in descriptiveness as well. When planning this study, I developed a fully-functioning chatbot that utilizes OpenAI’s assistant APIs, the Scrapy library, the Tkinter library, and the WordPress library to develop code for the graphical user interface of the software along with the functions that generate the responses and format the responses based on the content scraped from the sponsor’s website AOKmarketing.com. For the study, we asked participants to test out the chatbot with a variety of prompts. For each of these prompts, the participants would rate the chatbots responses, which would then lead to the exit-survey. The exit-survey contains questions used to rate the chatbot along with demographic questions to test out a variety of demographic factors to see if they have an impact on the participants rating on the chatbot. After conducting the survey with a group of 40 people, we found that none of the demographic factors had any significant impact on the participants rating of the chatbot. The only factor that was close to having a significant impact was a person's familiarity with generative AI. At the end of the study, the importance of having a diverse group of participants, having quantitative and qualitative data, and making survey instructions as clear as possible. We were limited to only computer science students and did not have a big budget to improve the quality of the software. In future studies, we would want to have a vast demographic audience range and improve the software with up-to-date technologies.

# Table of Contents

[**Table of Contents 3**](#_jncslxkus159)

[**Introduction 4**](#_h372buoab4tk)

[**Background & Related Work 6**](#_rhaocd6b0y93)

[Background 6](#_10w8bagps95k)

[Related Work 7](#_a51pmlxn261y)

[**Approach 8**](#_otlqf1d0h8bt)

[**Experimental Study 11**](#_oaphlie4rz9z)

[Description 11](#_9mgbuv8k9gj8)

[Methods 12](#_302vyqitd5ti)

[Participants 12](#_f12a2r2bxmjj)

[Instruments 15](#_thc69a8p2ymb)

[Measurements 16](#_wiqaptei2cfv)

[**Results 18**](#_pk07e8nmigqn)

[Measurement Analysis 18](#_o7k1jz1y8r7u)

[Data for Results 19](#_uu18po2edna1)

[**Discussion 23**](#_dquq9mnx9p82)

[**Conclusion 25**](#_yh1iq4riuc04)

[Lessons Learned 25](#_5a2kvpjc422l)

[Limitations 26](#_etpodihzlc4v)

[Future Work 26](#_l0jfx3gi6agi)

[**Appendix 28**](#_9ijf018uaqxw)

[**A: IRB Approval 28**](#_2prgtfyz9bae)

[**B: Survey Questions 31**](#_uppddbjd39i5)

[**C: Results Table of ANOVA Single Factor Analysis 34**](#_jujlnwgxy64f)

[**References 47**](#_wy4uhenl62uq)

# Introduction

The rise of Artificial Intelligence has caused the growth of various solutions used for health, surveillance, marketing, and more. One of these solutions that multiple companies use to boost the number of interactions with their website is a chatbot. Chatbots have been utilized to help new users to navigate the site and to answer common questions about the information the website has. When meeting with the sponsor for the first time, I was introduced to his company called AOKMarketing which had a great number of articles about different marketing topics. However, the main problem of the situation is found when the user asks a question that may not be on the site, there needs to be an alternative method to quickly scrape through the website's contents to generate an appropriate response. In our case, having a chatbot that can provide as much information as possible on marketing topics and creating thorough articles based on topics that are not currently in the website will allow for new ideas to be pitched and discussed by the website publishers.

Based on a Forbes Advisor study, they found that “97% of business owners believe [that] ChatGPT will help their business” whether it be for their website or for other work tasks (Haan, 2024). Adding a chatbot to the AOKMarketing website, may allow for more people to become interested in marketing or use the easy access to website information that may give the impression that the website content is trustworthy. Improving the quality and professionalism of the site can possibly help gain more clients for their marketing services. The AOKMarketing team also believes that the article generator will prevent writer’s block when it comes to developing blog posts. As of 2024, there was an increase in AI use in business by around 64% as it “increased productivity” among the employees (Haan, 2024). Creating a feature that develops articles based on prompts that are not currently in the blog database can lead to a variety of ideas being proposed that they may not have thought of before. This is where the idea of creating a chatbot with URL recommendation and article generation based on current blog posts stemmed from.

However, there is still the problem of how valid these AI chatbot responses are in terms of trustworthiness as these AI models mainly look for the words that are more likely to appear based on the input, the previous word used, and the data used to train the model (Caulfield, 2023). This possibility piqued my curiosity of how newer users of the chatbot would rate the software after asking it a few questions. It also got me thinking about any potential factors that may lead to people having less trust in the AI chatbot or how they rate the overall performance of the chatbot.

# 

# Background & Related Work

## Background

When considering the purpose and the design purposes, the AOKMarketing founder was looking to add a chatbot to improve the quality of their website and use the article generation assistant API to find new blog post ideas. This led him to think about having a chatbot that would be able to quickly generate new and accurate answers based on the blog post on AOKMarketing.com. Along with that, if an article already exists on that topic the user wants they should be directed to that specific page on AOKMarketing.com. If not then he wanted Chattable to create a new article on that topic based on the current blog posts on the AOKMarketing site. From there, the articles would be saved as a draft on the AOKMarketing WordPress site so that employees at AOKMarketing may look over these articles to ensure they are appropriate to post and the content of the article is accurate. Besides the development of this chatbot, this research work includes the testing of the chatbot to see how users would feel about its quality (trustworthiness and descriptiveness) and to see if there are any trends based on a specific demographic of users: such as gender, major or occupation, etc.

When creating chatbots like ChatGPT, libraries and application programming interfaces used for creating these kinds of chatbots vary depending on the programming language used, however, in general the ‘openai’ library is always used (OpenAI). Along with that, technologies are used to create a graphical user interface (GUI) that will efficiently update based on what the user inputs and the chatbot responds with: such as extra libraries (PySimpleGUI and Tkinter) or more adequate programming languages for software development (JavaScript and CSS). Python libraries are more simple to implement into the software, while the JavaScript and CSS languages give a more smooth and professional look to the software. However, with the JavasScript and CSS languages, the implementation time takes a lot longer in comparison to Python, which we did not have at the start of this study.

## Related Work

When looking into AI chatbots, it is found that they use a natural language processing model to understand the user’s input and create responses that answers any questions a user may have (ChatGPT). It allows for summarized responses to be made based on the user’s input and the blog posts AOKMarketing has in their site. However, when it comes to more specific topics the ChatGPT model has struggles thoroughly answering these kinds of questions. A little while after the release of ChatGPT, assistant APIs were soon introduced by OpenAI to allow companies to utilize a machine learning technique called “fine-tuning,” which allows a model to be trained on both a large set of data to understand basic tasks that it will perform and a smaller dataset to train the model on more specific tasks that the model will do (GeeksforGeeks, 2024). These models are used more in software development as companies add these kinds of models to their website to answer questions based on the information their website contains. Some of these chatbot implementations may include additional features that can help in the website navigation or information scraping: such URL recommendation or adding FAQ buttons at the beginning of the chatbot.

# Approach

During the development, OpenAI’s ChatGPT Assistant API created a more specialized version of a chatbot that can recognize certain commands and respond differently depending on the user’s input. This way the chatbot can respond to general questions similar to how ChatGPT would respond to those types of questions and respond in a more informative and descriptive manner when needed. The AOKMarketing blog posts could also be used to train these assistant models to help in drafting a more descriptive and accurate response. Along with this, there was one other major library that I utilized which was Tkinter, and a few “.tcl” files to improve the look of the GUI. Tkinter gave the overall software a pleasant look that would not scare users away from interacting with the chatbot in fear of it malfunctioning and to prevent users from getting a bad reputation of AOKmarketing.com in their minds. Along with that, the library was cost effective, as we did not have any funds for software development outside of the OpenAI API key, and simple to understand in contrast to the other libraries looked at. Lastly, the Scrapy Python library also came in useful as it was used to gather crucial marketing information and links within AOKMarketing.com’s blog page to give to users who are new to the site. Scrapy paired with the ChatGPT Assistant model will provide the user with thorough responses on topics that have been written about and will generate articles on marketing topics that are not featured on the website. It also allows us to scrape for only the necessary information to train the assistant models: such as article title, URL, and article content.

When developing a fully functional chatbot to support the website for marketing by adding an article generator feature that can be used in the scenario that no relevant sources can be found. Allowing for articles to be generated and kept as drafts for the website publishers leads to them learning about new questions that may be asked over time. This provides user feedback to the publishers and will give them new article ideas when they may be experiencing writer's block. For users, it allows them to get immediate feedback to questions that do not have a specific article further discussing the topic rather than waiting for an article related to the topic or having to switch to another site. On top of that, the addition of URL recommendation allows users to get familiar with the website rather than solely relying on the chatbot to answer questions. The more the user interacts with these recommended URLs, the more they will understand where they are in the site and will click on other website links to learn about other marketing topics or even learn about the services AOKMarketing provides.

Along with that, the usability and performance of the Chattable AI software was tested to ensure that the chatbot creates concise and descriptive responses when similar articles are found and should create a 1000 word article if no articles relate to the inputted question. The chatbot should display thorough responses related to the user's input regardless of whether or not there are articles related to this subject. These responses are mainly developed based on the AOKMarketing blogs it is trained on, however, in the case where the information does not answer the question then it will default to using ChatGPT’s database for help.

While the study focuses on the performance of the GUI, the time limitations that prevented me from implementing a smooth graphical user interface may lead to a lower impression of the overall performance of the software. Along with implementing the software, I needed to get confirmation for the study by the IRB and to develop the appropriate forms and documentation needed to conduct the study and recruitment (Appendix A).

As part of the development, a user experience study was performed to test the responses and the ease of use for the chatbot. When looking for information about a certain topic, it is crucial for responses to be descriptive and trustworthy, so the users asking questions will use the information and ask more questions.

When creating the application for this solution, the GUI was created first to properly connect the inputs and outputs of the AI’s responses to the correct places. On top of that, I spent time learning more about the Scrapy library so I could work on the development of the web crawler that could gather data from AOKMarketing.com respective to each of the blog pages. Once the inputs and outputs were correctly connected within the GUI and had been tested with a practice ChatGPT API key, I then moved on to the development of the ChatGPT assistant model and the web crawler to learn about how to connect the two aspects of the project to create a responsive chatbot that answers questions similar to a real human being. The web scraper development was worked on first to ensure that all the correct data was taken from the blog pages along with their respective URL. Then, after spending time studying how the assistant model works, the custom assistant model for the AOKMarketing website was developed and attached to the GUI so the answers and sources could be formulated and output onto the chat history. If no sources are found, then the assistant model will generate an article with the information on the AOKMarketing website. Once the first implementation was completed, the chatbot was then tested and then adjusted to match the participants’ feedback. The chatbot was then tested again to ensure that it was well-liked by the users.

# Experimental Study

## Description

After the completion of the implementation, a group of possible participants to test out the software will be recruited, so that I may be informed of any bugs or fixes that need to be made. These users will be both students and adults who know the business field and are looking to learn more about improving their businesses. This way I can fully understand the guidelines I must follow when creating this study and how I should plan my research so no risks will be present for the participants. I can also define the criteria that each person in the test group will be ranking the software on. From there I will communicate with my sponsor and the director about the results and discuss possible additions or solutions for the software that I see fit. This information will be used to determine which demographic we should aim to improve their view on the chatbot and improve any general software development critiques participants may have.

The research and user testing procedure consists of two main parts which are the during-testing questions along with the post-testing questions. This will be conducted during the participant's free time and this test should take between 25-40 minutes. This meeting will either take place on Zoom to ensure that if the participants have any questions about the testing process they can be quickly answered. For the during-testing portion of the study, participants will test the software on Replit.com so that we can have participants quickly start their study rather than having to wait for participants to download the entire code. The user will share their screen and the outputs of the chatbot will be recorded. The participant will then follow the instructions writing a certain kind of prompt to the chatbot and take time to learn about the types of information AOKMarketing provides in its blogs. Once given the response they will judge the quickness, descriptiveness, and naturalness of the response. After that part of the testing, the participant will answer a brief survey that has questions based on how well the chatbot performed. This ranges from questions rating the different aspects of the chatbot to short answer questions asking for feedback on the chatbot. The during-testing portion will take 10-15 minutes while the post-testing portion should take 5-10 minutes. The data from these tests will be stored in a private Google Drive folder along with its respective Google form.

## Methods

### Participants

For the testing phase of this project, users who most resemble the main demographic would be using the software to get more beneficial information on what features can be added to improve the quality of the chatbot and to make the product as user-friendly as possible. The recruitment and consenting of participants was conducted by me once the project was fully developed. This was done during October and November of 2024 as I asked Computer Science students around ASU Tempe campus.

For testing, 40 people participated in the study and all were Computer Science students at ASU. Out of all the students in the study, 97.5% of the participants were in the age range 18-24 and 2.5% of the participants were in the age range 25-34 (Figure 1).

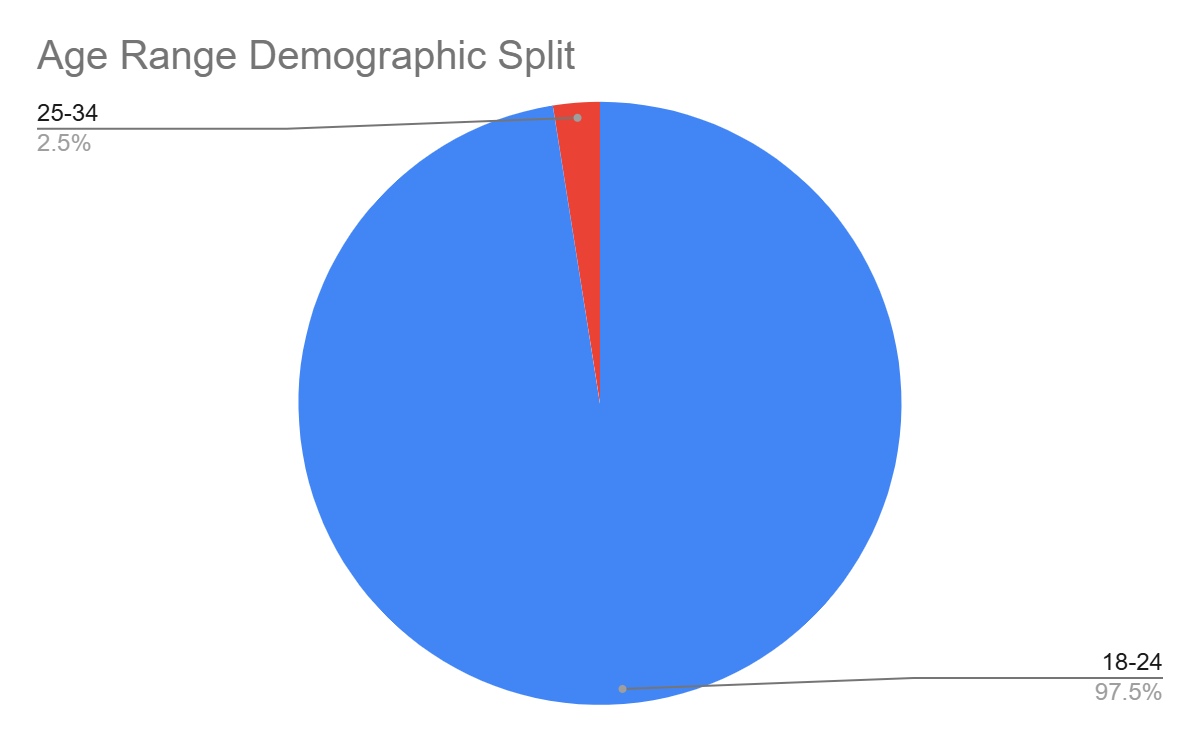


Figure 1) Pie graph that shows the age demographic split

The gender split was 20% female, 75% male, and 5% preferred not to say (Figure 2).

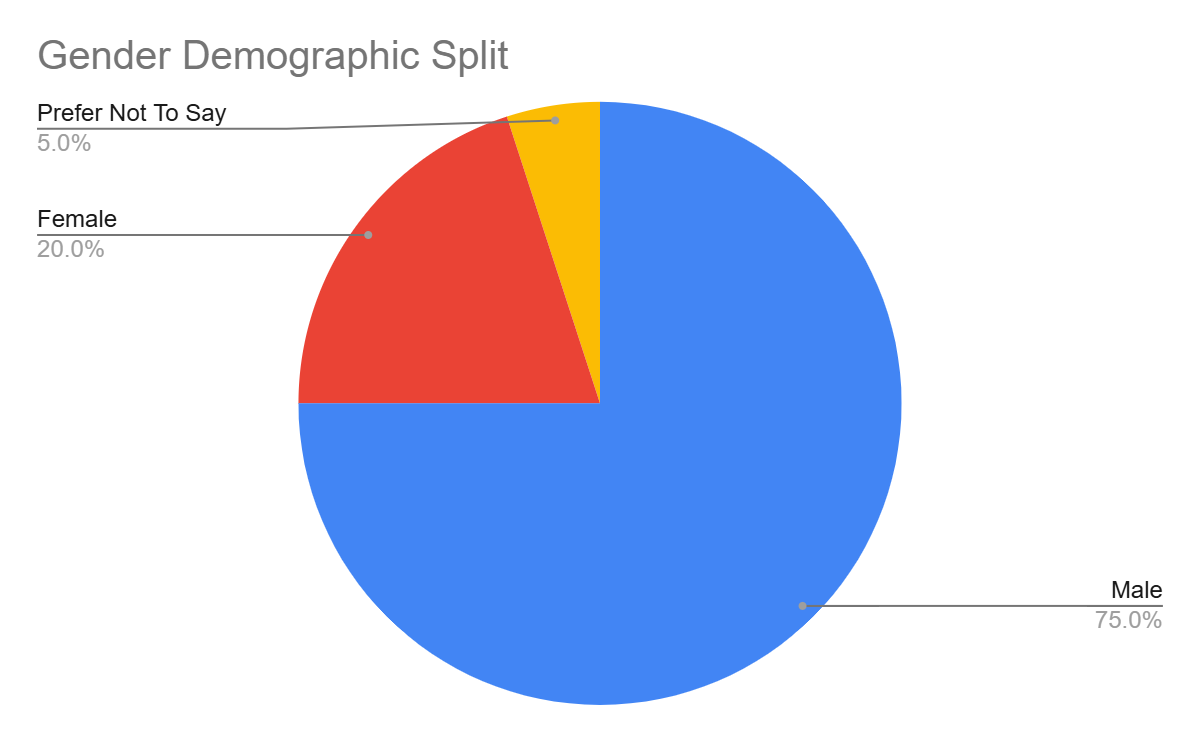


Figure 2) Pie graph that shows the gender demographic split

On top of that, around 77.5% of students were seniors, around 17.5% of students were Juniors, and around 5% were sophomores (Figure 3).

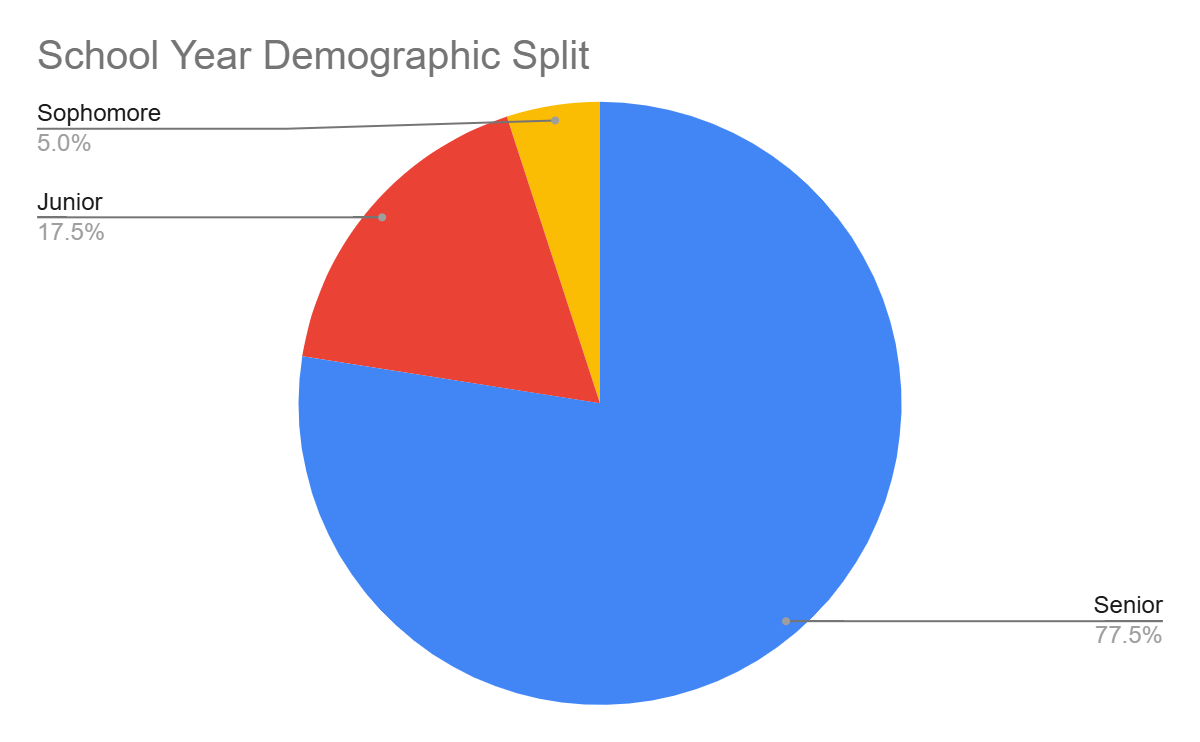


Figure 3) Pie graph that shows the school year demographic split

For the participants’ familiarity with generative AI, 47.5% of participants were familiar with the topic and 52.5% of participants were very familiar with the topic (Figure 4).

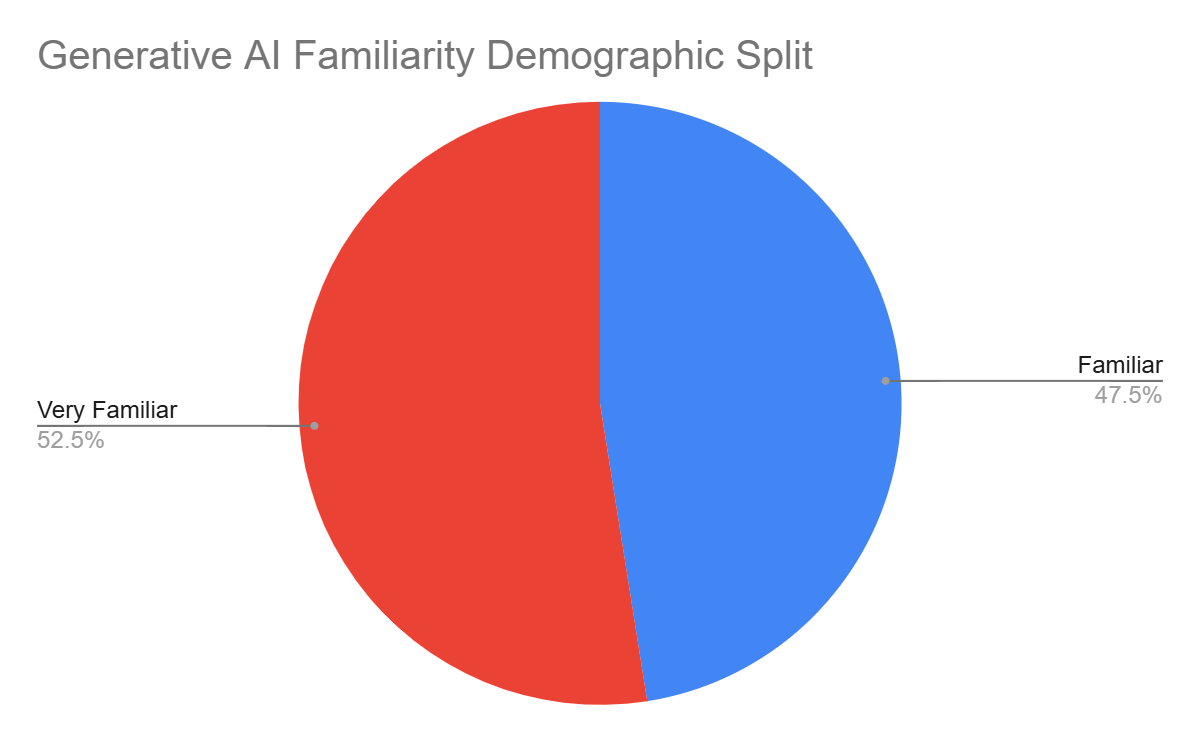


Figure 4) Pie graph that shows the generative AI familiarity demographic split

For the participants’ familiarity with marketing and business topics, 10% of the participants were very unfamiliar with the topic, 25% of the participants were unfamiliar with the topic, 30% of the participants were neutral with the topic, 25% of the participants were familiar with the topic, and 10% of the participants were very familiar with the topic (Figure 5).

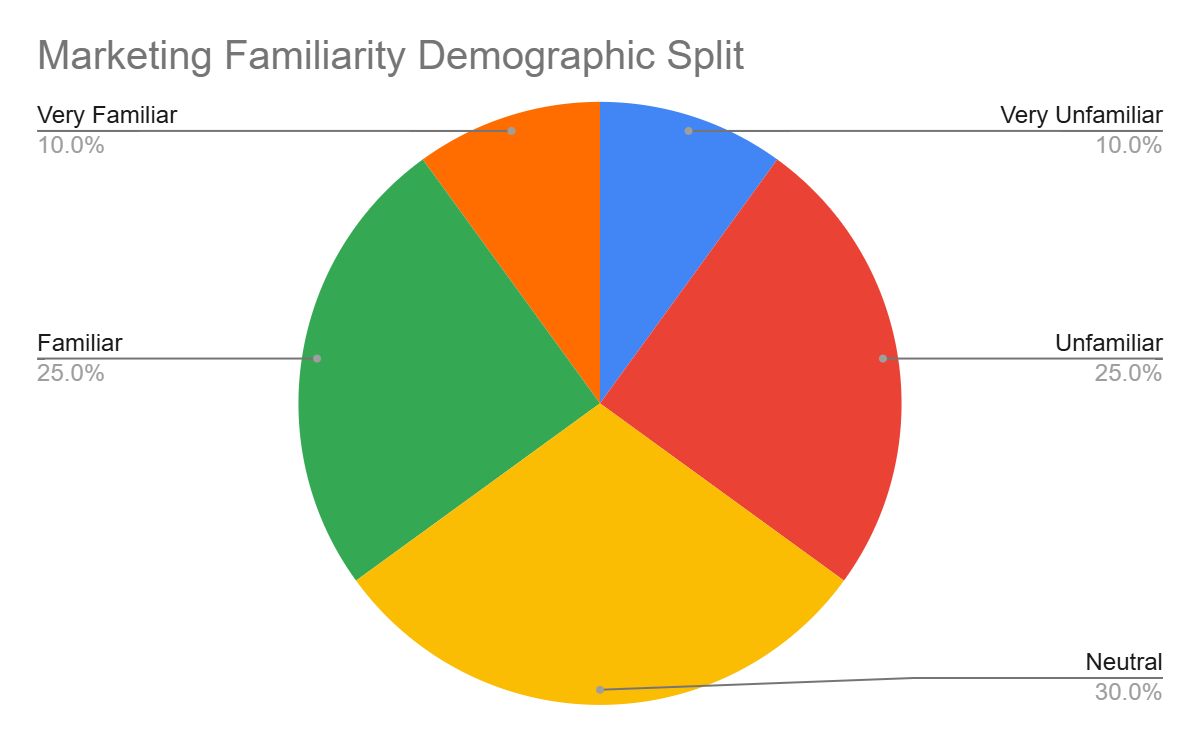


Figure 5) Pie graph that shows the marketing familiarity demographic split

For the participants’ familiarity with chatbots, 2.5% of the participants were very unfamiliar with the topic, 5% of the participants were neutral with the topic, 37.5% of the participants were familiar with the topic, and 55% of the participants were very familiar with the topic (Figure 6).

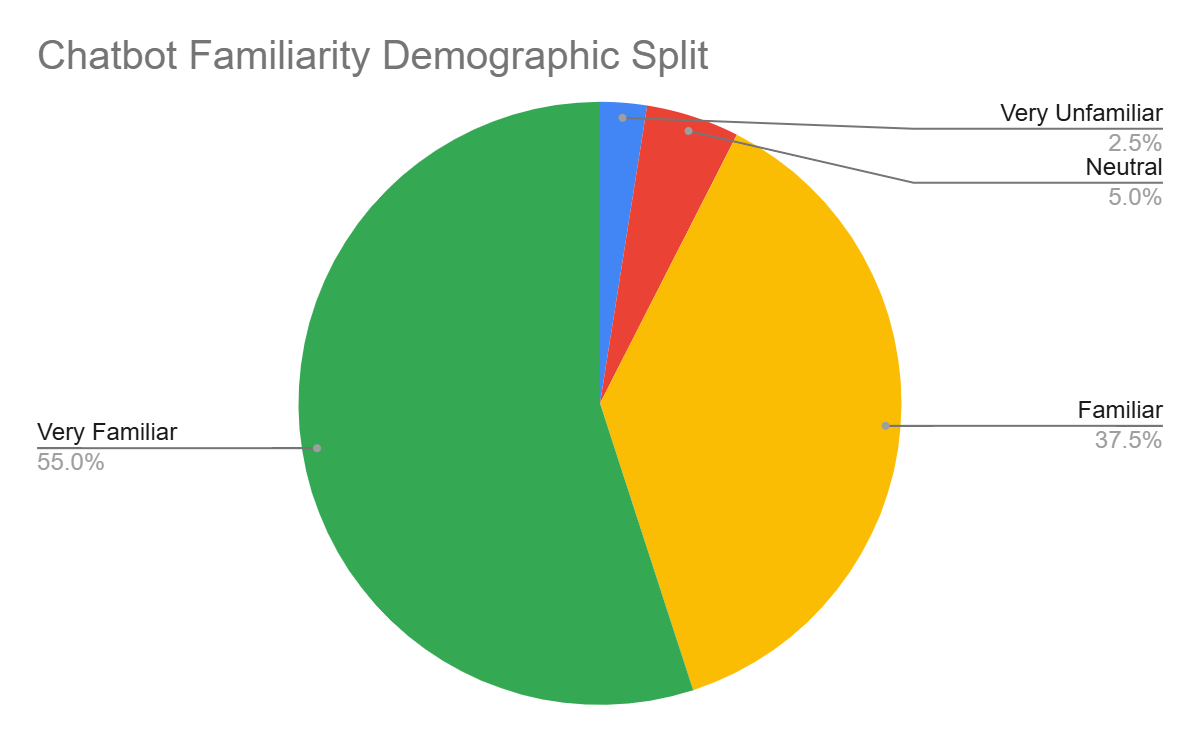


Figure 6) Pie graph that shows the chatbot familiarity demographic split

### Instruments

When determining how to record the participants’ reactions of the chatbot, I decided to record the screen of the participant to see how they interact with the chatbot, and to use a Google Form to have more quantifiable data. For the Google Form, there were a variety of questions ranging from evaluation of the chatbot based on its responses to the prompt the user inputted to demographic questions that will be (Appendix B). The website Replit.com was used to run the software so that the participants can interact with it.

When interacting with the chatbot, the user would answer the following questions after seeing the response of the chatbot: Rate the quickness of the chatbot’s response with 1 being slow and 5 being fast, Rate the descriptiveness, the amount of detail/thoroughness, of the chatbot’s response, and Rate how natural the response of the chatbot was. This way we can determine if there are differences in opinion based on the different ways the chatbot responds to a user input.

After interacting with the chatbot, we asked both multiple choice and open-ended questions based on the chatbot’s overall performance so that we could analyze the opinions of the participants and learn about any smaller issues users had. Once those questions were answered, participants would finish up the survey by answering demographic questions to help us determine if there are any trends that exist between various demographic groups (Age, Gender, Familiarity with AI, etc.).

### Measurements

For this study, we measured how the participants felt about the overall performance, trustworthiness, informativeness, and the positivity of the feedback related to our chatbot. These data points are considered as the dependent variables in the study. For the categories we looked at, we measured the major of the participants’ if they were students, the age range of the participants, the participants’ familiarity with generative AI, the participants’ familiarity with marketing and business topics, and the participants’ familiarity with chatbots. These are considered as the independent variables in this study as we are trying to determine if these independent variables impact the results of the dependent variables. Once these conversions were made and the data was split up based on demographic category, we conducted ANOVA: Single Factor analysis on each of these demographic groups to see if there are any potential trends that we could find. We conducted this ANOVA: Single Factor test with each demographic factor against each of the observational data categories we collected. The means of the data were utilized to see if there were any differences in means among the data.

# Results

### Measurement Analysis

When preprocessing the data collected, the answers were converted from the likert scale questions to discrete numbers (higher number meaning better performance on a scale from 1 to 5) and used the open-ended question feedback we got to analyze if the feedback was positive or negative (1 = negative, 2 = neutral, 3 = positive). These conversions allowed us to conduct ANOVA analysis on the performance data that we recorded.

For the overall performance data, it asked participants to rate the overall performance of the chatbot on the following options: Very Poor, Poor, Neutral, Good, Very Good. When converting the data into discrete numbers, we assigned values based on the positivity of the choice. This means that Very Poor = 1, Poor = 2, Neutral = 3, Good = 4, and Very Good = 5.

When converting the trustworthiness data, it asked participants to rate the overall performance of the chatbot on the following options: Very Untrustworthy, Untrustworthy, Neutral, Trustworthy, Very Trustworthy. The data was converted similarly to the overall performance data: Very Untrustworthy = 1, Untrustworthy = 2, Neutral = 3, Trustworthy = 4, Very Trustworthy = 5.

When converting the informativeness data, it asked participants to rate the overall performance of the chatbot on the following options: Very Uninformative, Uninformative, Neutral, Informative, Very Informative. The data was converted similarly to the overall performance data: Very Uninformative= 1, Uninformative= 2, Neutral = 3, Informative= 4, Very Informative= 5.

Lastly, when converting the feedback data, I looked into the open-ended questions that were asked and rated the questions based on the overall positivity of all of the open-ended responses. The data was converted into one of three options: Negative = 1, Neutral = 2, Positive = 3.

### Data for Results

The firstfactor we wanted to look at was the gender of the participants. A one-way ANOVA was performed to compare the effect of gender on the chatbots overall performance. A one-way ANOVA revealed that there was not a statistically significant difference overall with the F-value = 0.1206 and the p-value = 0.7304 (Appendix C - Gender Significance with Overall Performance). For the chatbot’s trustworthiness, there was not a statistically significant difference as the F-value = 0.488 and the p-value = 0.4893 (Appendix C - Gender Significance with Trustworthiness). For the chatbot’s informativeness, we did not find a statistically significant difference with the F-value = 0.1205 and the p-value = 0.9132 (Appendix C - Gender Significance with Informativeness). Lastly, for the feedback on our chatbot, there was no statistical significant difference as the F-value = 0.0109 and the p-value = 0.9175 (Appendix C - Gender Significance with Overall Feedback; Figure 7).

The next factor we wanted to look at was the school year of the participants. For this section, participants were only in their Sophomore, Junior, or Senior year. When looking at the overall performance, it was revealed that there was no difference as the F-value = 0.0723 and the p-value = 0.9304 (Appendix C - School Year Significance with Overall Performance). For the chatbot’s trustworthiness, there was no statistically significant difference with the F-value = 0.397 and the p-value = 0.6749 (Appendix C - School Year Significance with Trustworthiness). For its relation to the chatbot’s informativeness, there was no statistically significant difference in the participants' school year with the F-value = 0.024 and the p-value = 0.9763 (Appendix C - School Year Significance with Informativeness). Lastly, when looking at the feedback about our chatbot, there was no statistically significant difference found in the participants' school year with the F-value = 0.7514 and the p-value = 0.4787 (Appendix C - School Year Significance with Overall Feedback; Figure 7).

Following the school year factor, we measured for any relationships in familiarity with generative AI. For this section, participants were only Familiar or Very Familiar with generative AI. When looking at the overall performance, there was no difference found with the F-value = 1.5679 and the p-value = 0.2182 (Appendix C - Familiarity with Generative AI Significance with Overall Performance). For the trustworthiness, there was not a statistically significant difference with the F-value = 1.0784 and the p-value = 0.3056 (Appendix C - Familiarity with Generative AI Significance with Trustworthiness). There was also no statistically significant difference in the informativeness ratings as the F-value = 0.6678 and the p-value = 0.4189 (Appendix C - Familiarity with Generative AI Significance with Informativeness). Lastly, when looking at the feedback, there was no statistical difference as the F-value = 0.5571 and the p-value = 0.46 (Appendix C - Familiarity with Generative AI Significance with Overall Feedback; Figure 7).

We then measured the data’s relationship with the familiarity with marketing and business topics. The groups for this section were Very Unfamiliar, Unfamiliar, Neutral, Familiar, and Very Familiar. The data overall performance did not have a statistically significant difference with the F-value = 0.1127 and the p-value = 0.9772 (Appendix C - Familiarity with Marketing Significance with Overall Performance). For the trustworthiness data, no statistical difference was found as with the F-value = 0.1178 and the p-value = 0.9753 (Appendix C - Familiarity with Marketing Significance with Trustworthiness). When looking at the data for informativeness, there was no statistically significant difference with the F-value = 1.1323 and the p-value = 0.3573 (Appendix C - Familiarity with Marketing Significance with Informativeness). Lastly, the data on the feedback for our chatbot did not yield a statistically significant difference with the F-value = 0.5107 and the p-value = 0.7282 (Appendix C - Familiarity with Marketing Significance with Overall Feedback; Figure 7).

Finally, we compared the data to their correlation to the participants’ familiarity with chabots. For this report, the only groups that were measured were Very Unfamiliar, Neutral, Familiar, and Very Familiar. When measuring this with the overall performance, there was no significant difference found with the F-value = 1.254 and the p-value = 0.3047 (Appendix C - Familiarity with AI Chatbots Significance with Overall Performance). When looking at it with the trustworthiness data, there was not a statistically significant difference as the F-value = 0.9198 and the p-value = 0.4411 (Appendix C - Familiarity with AI Chatbots Significance with Trustworthiness). For the data about our chatbot’s informativeness, there was not a statistically significant difference in the participants' school year between the Sophomore, Junior, and Senior groups with the F-value = 0.1963 and the p-value = 0.8982 (Appendix C - Familiarity with AI Chatbots Significance with Informativeness). Lastly, when looking at how positive the feedback was for that chatbot, no statistical difference was found as the F-value = 0.548 and the p-value = 0.6527 (Appendix C - Familiarity with AI Chatbots Significance with Overall Feedback; Figure 7).

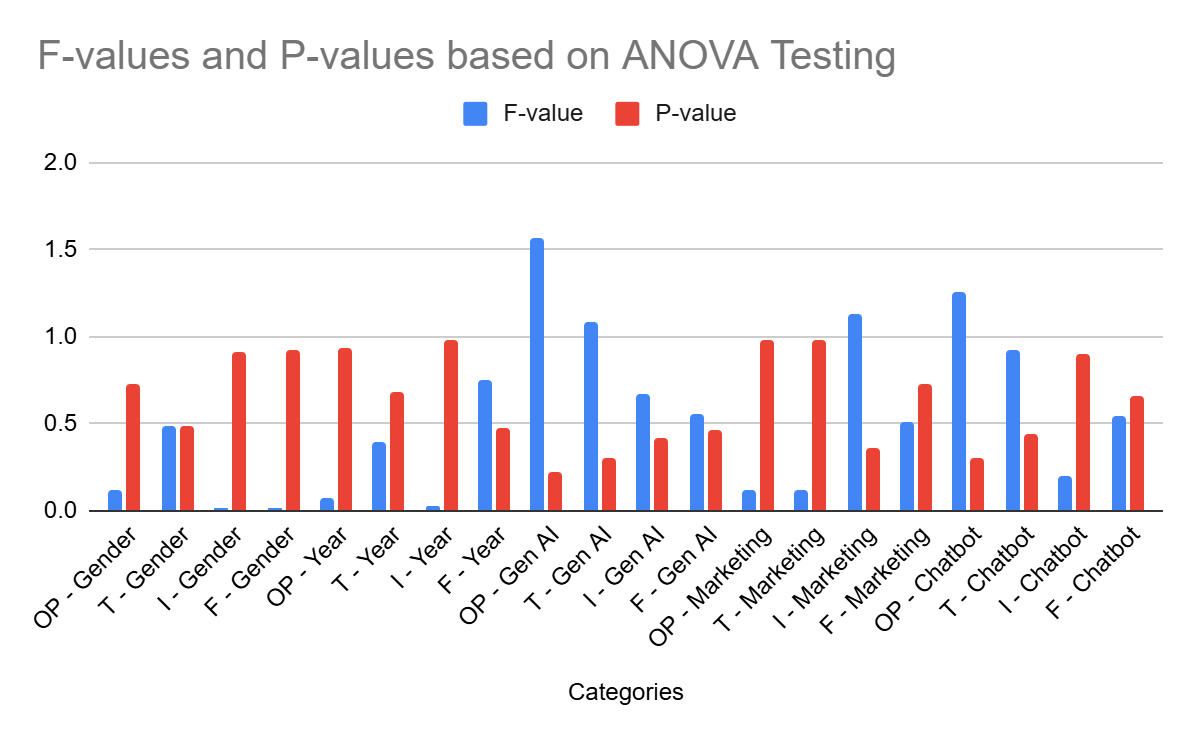


Figure 7) The diagram above displays the F-values and P-values for each of the factors we tested. (OP = Overall Performance, T = Trustworthiness, I = Informativeness, F = Feedback)

# 

# Discussion

After looking through the p-values of the various ANOVA: Single Factor tests that were conducted, none of the factors impacted the participants opinion on the overall performance, trustworthiness of the chatbot, informativeness of the chatbot, and the positivity of the feedback on the chatbot. (Figure 3) In order for the factor to be classified as having a significant effect on the dependent variables we tested for, the p-value from the test needs to be below the value of 0.05. Each of these categories (independent variables) were individually tested against each of the observation data (dependent variable data) recorded. However, all of these tests yielded p-values above the value of 0.05 meaning that none of them had a large enough difference in means to be considered as having a significant effect on any of the dependent variable data we collected.

**Table I: Generalized Results for each ANOVA Single Factor Analysis**

| **Categories / Factors** | Is Factor Significant to Overall Performance? | Is Factor Significant to the chatbot’s Trustworthiness? | Is Factor Significant to the chatbot’s informativeness? | Is Factor Significant to Overall Feedback? |
| --- | --- | --- | --- | --- |
| Gender | NO | NO | NO | NO |
| School Year | NO | NO | NO | NO |
| Familiarity with Generative AI | NO | NO | NO | NO |
| Familiarity with Marketing | NO | NO | NO | NO |
| Familiarity with Chatbots | NO | NO | NO | NO |

One thing that did stick out was how close all of the P-values in the generative AI familiarity section were to 0.05. This displays a strong possibility that a participant’s familiarity in the generative AI subject and in using it may have a strong impact in determining how a participant rates our chatbot. Along with that, the P-value for informativeness in correlation with marketing and business familiarity and overall performance in correlation with chatbot familiarity also have low P-values, which could imply that familiarity in these fields can greatly impact how a participant rates these data points. For example, if someone frequently uses ChatGPT to get information about a subject, they will know the descriptiveness of the chatbot, speed of the responses, and how reliable these answers are, which gives them expectations of how a proper chatbot should function. These baseline expectations allow for a more fair judgement of the chatbot. On the other hand, a person who has not used ChatGPT at all or has used it very little will not know what to expect and may not expect much leading to a more positive rating of the chatbot.

# Conclusion

### Lessons Learned

One lesson taken into account was to get as diverse of an audience as possible as it helps generalize the data trends found. Even though the data was taken from Computer Science students at ASU along with people in the age range 18-24, getting people from various backgrounds increases the validity of the trends derived. If diversity is not taken into account when recruiting participants, it may cause doubts that the observations found apply to all people and may associate the trend to a specific group.

It was also realized that adding more multiple choice questions to the survey allows the results to be easily quantifiable. For the first draft of the Google form used to record the participants feedback, there were more open-ended questions rather than multiple choice questions that could be quantified. In the second draft, more Likert-Scale questions were added so I could convert the responses to numerical values. Adding more of these types of questions eased the process of analyzing the data with the ANOVA analysis and better allowed me to draw more concrete conclusions from the data.

Lastly, making the instructions for the survey as simple as possible for participants reduced the amount of impact human error has on a person’s rating of the chatbot. During the process of surveying the participants, there were moments where the participants were confused with what to ask the chatbot due to the lack of labels or descriptiveness of instructions in the survey. Adding more descriptiveness to the instructions or adding labels to sections students need to look at will prevent any mistakes from being made in the survey and prevent any questions from being improperly answered.

### Limitations

In retrospect of this study, we found a few limitations when it comes to both the recruitment portion and chatbot implementation of the study. Firstly, the lack of variety of participants is a major limitation since we are only conducting a survey with people who are a part of the Computer Science program at Arizona State University. While there was a good variety in the age demographic, numerous fields and majors were missing that could add to our study and prevent making assumptions that apply more to computer science students. We only are able to conduct this study on people in the marketing field and the Computer Science major which is only a select few in the variety of fields there are.

Another limitation is the lack of frontend development done on the AI chatbot. Due to time limitations, we were only able to do so much work on the software that prevented us from implementing a smooth graphical user interface. This may lead to more confusion about the software leading more people to have a more negative perception of AI chatbots than they usually do. The lack of development with the graphical user interface makes the software function appear stagnant at times which may give the user the belief that the software is crashing or not functioning.

### Future Work

In the future, expanding the vastness of our study participants to various age ranges as well as occupations. All of the data recorded for this are from computer science undergraduate students possibly meaning that all takeaways from this study may only apply to people who are in the process of their computer science degree. Opening the survey to a wider range of demographics will create more generalized observations rather than observations specific to one major or one age range.

Along with that, we may want to upgrade the software to use a more professional and aesthetically pleasing look for the chatbot. On top of that, we want to consider using software that allows for more smooth interactions. The minimalist and little time working on the frontend aspects of the design may lead people to an increase in confusion when operating the software. Improving these aspects will allow users to connect it with other professional AI chatbots, such as ChatGPT, and will allow participants to better understand how the system works, possibly leading to participants having more trust in the software and adding more validity to our observations on what factors contribute to our chatbot’s ratings. This could be done with more modern languages for frontend development (JavaScript and CSS) or by using more expansive and customizable python libraries (PyQt or Kivy).

# Appendix

# **A: IRB Approval**

# **B: Survey Questions**

1. Pre-Survey
   1. Warm-up: Instructions: Before using the chatbot, also known as Chattable, please take about 3-5 minutes to familiarize yourself with the content that AOKMarketing’s blog page has to offer. This way you can get a general understanding of what kind of questions to ask the chatbot. (<https://aokmarketing.com/blog>) Once this is done, start up the software on the Replit site and enter in the API keys needed to get the chatbot responses. These will be given by the survey conductor (Junayd Lateef).
2. Survey
   1. Instructions: Once you have looked over the AOKMarketing blogs page, you can now begin interacting with the chatbot by creating prompts similar to the questions below (Use these questions as a guide of what you should ask and what you should NOT ask it). Please click on the “Start” Button at the top of the Replit page to run the program. You will be given a few topics that you can ask Chattable about and once you have received a response you may answer the survey questions that follow it.
      1. Good Questions: What are some beginner tips for marketing?, What are some common marketing mistakes?, What are the best places to market a product at?
      2. Bad Questions: How is the weather outside today?, Do you recommend I go into marketing?, What is the eBay marketing company?
   2. Ask Chattable for some beginner tips about marketing and open the URLS it recommends. (These will be asked for each of the survey questions)

* Rate the quickness of the chatbot’s response with 1 being slow and 5 being very fast. The average time for chatbot responses is 5-15 seconds based on your question (1-5)
* Rate the descriptiveness, the amount of detail/thoroughness, of the chatbot’s response (1-5)
* Rate how natural the response of the chatbot was (did the responses sound like a human or more like a robot) (1-5)
  1. Ask Chattable about a question you have about marketing relating to the following topics (Marketing Mistakes, Online Engagement, Marketing Strategies).
  2. Ask Chattable a question relating to a marketing topic that is not relating to the following topics (Marketing Mistakes, Online Engagement, Marketing Strategies).
  3. Ask Chattable about a question not related to marketing or business.

1. Exit-Survey
   1. Chatbot Questions
      1. Rate the chatbot’s overall performance. Please consider the time of the response, the level of detail in the response, how natural the response was, and how easy it was to use the chatbot. (1-5)
      2. Why did you give the rating above?
      3. What did you like about the chatbot?
      4. How did you feel about the personality and tone of the response from the chatbot?
      5. Did you find the chatbot trustworthy?
         1. 1 = Not Trustworthy
         2. 5 = Very Trustworthy
      6. Did the article, created by the chatbot, contain a thorough/detailed response on the topic you were interested in?
         1. 1 = Uninformative (Low detail and boring)
         2. 5 = Informative (High detail and Interesting)
      7. What would you improve about the chatbot?
      8. Do you have any other comments or feedback about the chatbot?
   2. Demographic Questions
      1. What is your age? (Can answer with ‘Prefer not to say’)
      2. What is your gender? (Can answer with ‘Prefer not to say’)
      3. What is your current occupation? (Can answer with ‘Prefer not to say’)
      4. How familiar are you with Generative AI?
         1. 1 = Not Familiar
         2. 5 = Very Familiar
      5. How familiar are you with Marketing and Business?
         1. 1 = Not Familiar
         2. 5 = Very Familiar
      6. How familiar are you with AI Chatbots?
         1. 1 = Not Familiar
         2. 5 = Very Familiar
   3. If the person’s occupation is a employee
      1. What is your major? (Can answer with ‘Prefer not to say’)
      2. What is your year?
         1. Freshman
         2. Sophomore
         3. Junior
         4. Senior
         5. Prefer not to say
   4. If the person’s occupation is a student
      1. What field do you work in? (Can answer with ‘Prefer not to say’)

# 

# **C: Results Table of ANOVA Single Factor Analysis**

Gender Significance with Overall Performance

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Female | 8 | 31 | 3.875 | 0.125 |  |  |
| Male | 30 | 119 | 3.966666667 | 0.516091954 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.05307017544 | 1 | 0.05307017544 | 0.1206013456 | 0.7304053134 | 4.113165219 |
| Within Groups | 15.84166667 | 36 | 0.4400462963 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 15.89473684 | 37 |  |  |  |  |

Gender Significance with Trustworthiness

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Female | 8 | 32 | 4 | 0.2857142857 |  |  |
| Male | 30 | 113 | 3.766666667 | 0.8057471264 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.3438596491 | 1 | 0.3438596491 | 0.4880005533 | 0.4893090691 | 4.113165219 |
| Within Groups | 25.36666667 | 36 | 0.7046296296 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 25.71052632 | 37 |  |  |  |  |

Gender Significance with Informativeness

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Female | 8 | 32 | 4 | 0.8571428571 |  |  |
| Male | 30 | 119 | 3.966666667 | 0.516091954 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.00701754386 | 1 | 0.00701754386 | 0.0120492009 | 0.9132023253 | 4.113165219 |
| Within Groups | 20.96666667 | 36 | 0.5824074074 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 20.97368421 | 37 |  |  |  |  |

Gender Significance with Overall Feedback

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Female | 8 | 21 | 2.625 | 0.2678571429 |  |  |
| Male | 30 | 78 | 2.6 | 0.3862068966 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.003947368421 | 1 | 0.003947368421 | 0.01086847137 | 0.9175480589 | 4.113165219 |
| Within Groups | 13.075 | 36 | 0.3631944444 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 13.07894737 | 37 |  |  |  |  |

School Year Significance with Overall Performance

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Senior | 31 | 121 | 3.903225806 | 0.3569892473 |  |  |
| Junior | 7 | 28 | 4 | 1 |  |  |
| Sophomore | 2 | 8 | 4 | 0 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.06532258065 | 2 | 0.03266129032 | 0.07232142857 | 0.9303630115 | 3.251923846 |
| Within Groups | 16.70967742 | 37 | 0.4516129032 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 16.775 | 39 |  |  |  |  |

School Year Significance with Trustworthiness

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Senior | 31 | 119 | 3.838709677 | 0.5397849462 |  |  |
| Junior | 7 | 25 | 3.571428571 | 1.619047619 |  |  |
| Sophomore | 2 | 7 | 3.5 | 0.5 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.5671658986 | 2 | 0.2835829493 | 0.3973278946 | 0.6749468872 | 3.251923846 |
| Within Groups | 26.4078341 | 37 | 0.713725246 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 26.975 | 39 |  |  |  |  |

School Year Significance with Informativeness

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Senior | 31 | 121 | 3.903225806 | 0.6903225806 |  |  |
| Junior | 7 | 27 | 3.857142857 | 0.8095238095 |  |  |
| Sophomore | 2 | 8 | 4 | 0 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.0331797235 | 2 | 0.01658986175 | 0.02400865177 | 0.9762924594 | 3.251923846 |
| Within Groups | 25.56682028 | 37 | 0.6909951426 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 25.6 | 39 |  |  |  |  |

School Year Significance with Overall Feedback

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Senior | 31 | 81 | 2.612903226 | 0.311827957 |  |  |
| Junior | 7 | 17 | 2.428571429 | 0.619047619 |  |  |
| Sophomore | 2 | 6 | 3 | 0 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.530875576 | 2 | 0.265437788 | 0.7514809591 | 0.4787296338 | 3.251923846 |
| Within Groups | 13.06912442 | 37 | 0.353219579 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 13.6 | 39 |  |  |  |  |

Familiarity with Generative AI Significance with Overall Performance

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Familiar | 19 | 72 | 3.789473684 | 0.3976608187 |  |  |
| Very Familiar | 21 | 85 | 4.047619048 | 0.4476190476 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.6647243108 | 1 | 0.6647243108 | 1.567913815 | 0.2181625455 | 4.098171661 |
| Within Groups | 16.11027569 | 38 | 0.4239546234 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 16.775 | 39 |  |  |  |  |

Familiarity with Generative AI Significance with Trustworthiness

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Familiar | 19 | 69 | 3.631578947 | 1.023391813 |  |  |
| Very Familiar | 21 | 82 | 3.904761905 | 0.3904761905 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.7444235589 | 1 | 0.7444235589 | 1.07843971 | 0.305609907 | 4.098171661 |
| Within Groups | 26.23057644 | 38 | 0.6902783274 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 26.975 | 39 |  |  |  |  |

Familiarity with Generative AI Significance with Informativeness

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Familiar | 19 | 72 | 3.789473684 | 0.730994152 |  |  |
| Very Familiar | 21 | 84 | 4 | 0.6 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.4421052632 | 1 | 0.4421052632 | 0.6677824268 | 0.4189199578 | 4.098171661 |
| Within Groups | 25.15789474 | 38 | 0.6620498615 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 25.6 | 39 |  |  |  |  |

Familiarity with Generative AI Significance with Overall Feedback

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Familiar | 19 | 48 | 2.526315789 | 0.3742690058 |  |  |
| Very Familiar | 21 | 56 | 2.666666667 | 0.3333333333 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.1964912281 | 1 | 0.1964912281 | 0.5570680628 | 0.4600368567 | 4.098171661 |
| Within Groups | 13.40350877 | 38 | 0.3527239151 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 13.6 | 39 |  |  |  |  |

Familiarity with Marketing Significance with Overall Performance

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Very Unfamiliar | 4 | 16 | 4 | 0 |  |  |
| Unfamiliar | 10 | 38 | 3.8 | 0.1777777778 |  |  |
| Neutral | 12 | 47 | 3.916666667 | 0.6287878788 |  |  |
| Familiar | 10 | 39 | 3.9 | 0.5444444444 |  |  |
| Very Familiar | 4 | 15 | 3.75 | 0.9166666667 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.2083333333 | 4 | 0.05208333333 | 0.112757732 | 0.977206637 | 2.641465186 |
| Within Groups | 16.16666667 | 35 | 0.4619047619 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 16.375 | 39 |  |  |  |  |

Familiarity with Marketing Significance with Trustworthiness

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Very Unfamiliar | 4 | 15 | 3.75 | 0.25 |  |  |
| Unfamiliar | 10 | 38 | 3.8 | 0.4 |  |  |
| Neutral | 12 | 44 | 3.666666667 | 0.9696969697 |  |  |
| Familiar | 10 | 38 | 3.8 | 0.6222222222 |  |  |
| Very Familiar | 4 | 16 | 4 | 2 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.3583333333 | 4 | 0.08958333333 | 0.1177989981 | 0.9753028359 | 2.641465186 |
| Within Groups | 26.61666667 | 35 | 0.7604761905 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 26.975 | 39 |  |  |  |  |

Familiarity with Marketing Significance with Informativeness

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Very Unfamiliar | 4 | 18 | 4.5 | 0.3333333333 |  |  |
| Unfamiliar | 10 | 36 | 3.6 | 0.2666666667 |  |  |
| Neutral | 12 | 47 | 3.916666667 | 0.446969697 |  |  |
| Familiar | 10 | 38 | 3.8 | 1.288888889 |  |  |
| Very Familiar | 4 | 17 | 4.25 | 0.9166666667 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 2.933333333 | 4 | 0.7333333333 | 1.132352941 | 0.357239281 | 2.641465186 |
| Within Groups | 22.66666667 | 35 | 0.6476190476 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 25.6 | 39 |  |  |  |  |

Familiarity with Marketing Significance with Overall Feedback

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Very Unfamiliar | 4 | 10 | 2.5 | 0.3333333333 |  |  |
| Unfamiliar | 10 | 28 | 2.8 | 0.1777777778 |  |  |
| Neutral | 12 | 30 | 2.5 | 0.6363636364 |  |  |
| Familiar | 10 | 25 | 2.5 | 0.2777777778 |  |  |
| Very Familiar | 4 | 11 | 2.75 | 0.25 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.75 | 4 | 0.1875 | 0.5107003891 | 0.7282033031 | 2.641465186 |
| Within Groups | 12.85 | 35 | 0.3671428571 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 13.6 | 39 |  |  |  |  |

Familiarity with AI Chatbots Significance with Overall Performance

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Very Unfamiliar | 1 | 5 | 5 | #NUM! |  |  |
| Neutral | 2 | 7 | 3.5 | 0.5 |  |  |
| Familiar | 15 | 58 | 3.866666667 | 0.4095238095 |  |  |
| Very Familiar | 22 | 87 | 3.954545455 | 0.4264069264 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 1.587121212 | 3 | 0.529040404 | 1.253990423 | 0.304681376 | 2.866265557 |
| Within Groups | 15.18787879 | 36 | 0.4218855219 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 16.775 | 39 |  |  |  |  |

Familiarity with AI Chatbots Significance with Trustworthiness

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Very Unfamiliar | 1 | 3 | 3 | #NUM! |  |  |
| Neutral | 2 | 7 | 3.5 | 0.5 |  |  |
| Familiar | 15 | 54 | 3.6 | 0.9714285714 |  |  |
| Very Familiar | 22 | 87 | 3.954545455 | 0.5216450216 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 1.920454545 | 3 | 0.6401515152 | 0.9198113208 | 0.441066128 | 2.866265557 |
| Within Groups | 25.05454545 | 36 | 0.695959596 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 26.975 | 39 |  |  |  |  |

Familiarity with AI Chatbots Significance with Informativeness

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Very Unfamiliar | 1 | 4 | 4 | #NUM! |  |  |
| Neutral | 2 | 7 | 3.5 | 4.5 |  |  |
| Familiar | 15 | 58 | 3.866666667 | 0.4095238095 |  |  |
| Very Familiar | 22 | 87 | 3.954545455 | 0.7121212121 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.4121212121 | 3 | 0.1373737374 | 0.1963426372 | 0.8981991813 | 2.866265557 |
| Within Groups | 25.18787879 | 36 | 0.6996632997 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 25.6 | 39 |  |  |  |  |

Familiarity with AI Chatbots Significance with Overall Feedback

| Anova: Single Factor |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Very Unfamiliar | 1 | 3 | 3 | #NUM! |  |  |
| Neutral | 2 | 5 | 2.5 | 0.5 |  |  |
| Familiar | 15 | 37 | 2.466666667 | 0.4095238095 |  |  |
| Very Familiar | 22 | 59 | 2.681818182 | 0.3225108225 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 0.5939393939 | 3 | 0.197979798 | 0.5479962721 | 0.6527092569 | 2.866265557 |
| Within Groups | 13.00606061 | 36 | 0.3612794613 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 13.6 | 39 |  |  |  |  |

# References

ChatGPT. (n.d.). https://openai.com/chatgpt/overview

Caulfield, J. (2023, November 16). *Is CHATGPT trustworthy?: Accuracy tested*. Scribbr.

https://www.scribbr.com/ai-tools/is-chatgpt-trustworthy/

GeeksforGeeks. (2024, February 18). *Transfer learning with fine-tuning*. GeeksforGeeks.

https://www.geeksforgeeks.org/transfer-learning-with-fine-tuning-in-nlp/

Haan, K. (2024, October 16). *22 top AI statistics and Trends in 2024*. Forbes.

https://www.forbes.com/advisor/business/ai-statistics/

OpenAI. (n.d.). *Assistants API overview beta*. Assistant API Overview.

https://platform.openai.com/docs/assistants/overview

Python. (n.d.). *Tkinter - Python interface to TCL/TK*. tkinter - Python documentation

https://docs.python.org/3/library/tkinter.html

Scrapy. (n.d.). *A fast and powerful scraping and web crawling framework*. Scrapy.

https://scrapy.org/