Exploring Uses' Customization and Personalization Preferences for Conversational AI

Catherine Baker
Computer Science and Mathematics Major
Emory University
Atlanta, GA, 30322, USA
cwbake2@emory.edu

Thomas Gu
Department of Computer Science
Emory University
Atlanta, GA, 30322, USA
tgu24emory.edu

Abstract

While there are many works related to personality preference of conversational AI and whether or not a chatbot's personality would impact user's experience, the knowledge for merely text-based chatbots are rare. In this study, we conducted two information-gathering experiments aimed at helping improve the knowledge around chatbot personality preference. To do this, we conducted a survey (N=34 participants) analyzing emoji use in college students based on their levels of extroversion. We then used this data to create 16 chatbots with different personality traits based on the Myers-Briggs Type Indicator who used emojis based on their extroversion or introversion. We were then able to conduct an ethical survey (N=10 participants) to interact with these 16 chatbots. We systematically differed the chatbots' response in both text and emoji use based on the designated personality traits possessed. We found that participants prefer extroverted chatbots due to the higher rates of agreeableness that they possess in contrast to introverted chatbots. Participants felt that introverted chatbots do not provide users a full experience due to the lack of personable behavioral traits. Ultimately, we hope this data can be used to further the realm of knowledge around conversational AI and potentially aide in development of a chatbot with a dynamic personality, able to be changed based on user preference.

1 Introduction

Conversational AI (CAI) has become part of daily life. When calling customer service lines, asking Siri to send a text, or even receiving an automated order confirmation, we are unthinkingly communicating with AI. And when interacting with CAI, users will subconsciously assign a personality to the bot, regardless of whether this was intended by the conversation designer or not (Sarah Theres Völkel and Mayer, 2020). Our natural human instinct demands we form an impression of one personality when we interact with others, and this does not stop at humans. The impression that we assume affects, often subconsciously, our attitudes and behaviors toward the person or CAI presented. In translating this human instinct to CAI, we can understand how directly CAI personality presentation affects and influences users' engagement.

Many conversation agents, such as voice assistants and chatbots, are perceived as social actors and, therefore, elicit similar personality judgements. They are mostly made to appear friendly and agreeable to increase user engagement and enjoyment of experience. And as technology advances, home assistants and chatbots such as Amazon Alexa, Google Assistant, and Siri are becoming a more and more popular device to own. These Voice Control Assistants are created to help users handle various daily tasks such as light control, web services, scheduling, navigation, and more, but they also include many anthropomorphic features like telling jokes and stories, playing games, and holding conversations. For the most part, these conversation agents are communicated with audibly, but in many cases like assisted voice agents for the hearing impaired, chatbots, and automated response programs, we communicate with the CAIs over text.

1.1 Personalities in Conversational AI

Despite their convenience and helpfulness, conversational AI still adapts to the idea of a one-size-fits all approach. This is partially due to the limitations in speech recognition and as a result, these systems do not generally fulfill the user's expectation (Martin Jentsch, 2019). With the advancement of technology, the voice that advocates to integrate more comprehensive personality differences into AI has grown and a more comprehensive examination of such personality differences is essential for a deepened understanding of user diversity regarding experience and preferences in human-technology interaction (Szalma, 2009). Furthermore, according to Sean Andrist and Tapus (2015), pairing the user's personality to AI does have a positive effect in Human-Computer Interaction (HCI) and users become more willing to interact with conversational AI when their personality preferences are taken into account. Several studies have proposed approaches to overcome the issue of personality-less AI, but none, thus far, appropriately address user needs and preferences dynamically (Cliford Nass and Tauber, 1994).

In this paper, we present a designed study based on the Myers-Briggs Type Indicator (MBTI) research (N=10) to see how different text-based CAI responses based on MBTI personality traits affect user experience. Our approach is evaluated on three different levels of extroversion in chatbots: introverted, average, and extroverted. Because extroversion and introversion may be hard to represent over a text-based conversation, we will use emojis in our chatbots language to better communicate this trait. There has been shown to be a correlation between emoji use and extroversion versus introversion in humans Marengo et al. (2017), but to accurately represent emoji use in our chatbots, we will be conducting a survey of college students to assess the difference in emoji use based on a participants percent extroversion or introversion. Once finished, we will use the results of this survey to teach our chatbots to use emojis based on their own extroversion and introversion. We will also incorporate the use of the Big Five Theory's 5 characteristics of personality: openness, conscientiousness, extroversion, agreeableness, and neuroticism, into our chatbots to represent different Myers-Briggs personality types. By systematically varying the chatbots' use of language, we will allow the user's to experience different calculated personalities in chatbots (Sarah Theres Völkel and Mayer, 2020).

1.2 Development of Conversational AI

We hypothesize the following:

1. When analyzing emoji use, we will find that more extroverted users will have a tendency to use emojis conveying a higher emotional intensity than more introverted users.

2. When analyzing emoji use, we will find that more introverted users will have a tendency to use emojis conveying a lower emotional intensity than more extroverted users.

2. When interacting with personality-assigned text based conversational AI, users will prefer interacting AI with higher levels of extroversion.

To do this we have conducted a survey analyzing emoji use by extroverts and introverts and then created 16 chatbots. Our design focuses on a text-based chatbot to standardize variables such as voice gender, dialects, and usage of tones. Each chatbot receives a different MBTI personality and based on their different MBTI personality traits, they use emojis and speech differently in conversation as expanded on in section 3 Approach. We then enlist volunteers to interact with these chatbots and report their preferences and opinions following each conversation. This data is then collated and reported to analyze user preferences in personality types of conversational AI as shown in section 4 Experiments. We have found evidence to support the hypothesis that users prefer communicating with chatbots with higher levels of extroversion which we explain and define in section 5 Analysis.

To the best of our knowledge, this is the first work that considers MBTI as a factor for the home assistant and that develops AI emoji use based on just extroversion and introversion. This work will help facilitate the development of better and more human-like conversational AI.

2 Related Work

Personality is the characteristic sets of behaviors, cognition, and emotional patterns that are formed from biological and environmental factors which change over time (Corr, 2009). In this section, we briefly discuss the research landscape and background of the MBTI, a well known personality typology constructed by two Americans, Katharine Cook Briggs and her daughter Isabel Briggs Myers and the Big Five Theory, a more recently created construction of personality trait definitions.

2.1 Myers–Briggs Type Indicator (MBTI)

MBTI provides users with a self-report questionnaire that indicates different psychological preferences for how one perceives the world and makes decisions. The test attempts to assign four categories: introversion or extroversion, sensing or intuition, thinking or feeling, judging or perceiving (thi, 2019). Each category has two letters, so there are a total of sixteen personality types in MBTI.

Introversion and Extroversion can also be described as attitudes. Extroverts recharge and get their energy from spending time with people, while introverts recharge and get their energy from spending time alone; they consume their energy through the opposite process (Tieger, 1990), (Carlson, 1985), (Carlyn, 1977).

Sensing or Intuition describes how one perceives the world and interprets new information. Those that were identified as Sensing tend to trust their five senses, which are factors that exist in reality and have data to support facts. Those who score higher rates of Intuition tend to seek external explanation from the world around them and have a greater tendency towards inquisitiveness (thi, 2019).

Thinking or Feeling relates with decision-making. Those that were identified as Feeling tend to empathize with situations, incorporating the needs of everyone who was involved. Those who score higher in Thinking tend to make decisions logically without incorporation or consideration of empathy or emotions (thi, 2019).

Judging and Perceiving indicates lifestyle preference. Those who were identified as Judging paid more attention to the result while those who were identified as Perceiving cared more about the process when they were approaching the result (thi, 2019).

MBTI's goal gives a personality type, instead of personality traits. Psychologists usually use traits versus types to talk about personality as person-to-person, the interactions between personality traits can vary wildly. Because computers have much less variation between interpretation of personality traits, we feel the MBTI is an apt assessment when attributing personality type to conversational AI. In this way we are simplifying the task of assigning personality to AI and generalizing personality types between chatbots to standardize user experience. Our work varies from previous work significantly by focusing on endowing the chatbot with a more implicit personality based on the MBTI.

There has been some controversy over the validity of the Myers Briggs test. Most articles addressing this validity are from 1960 to 1990. We have conducted a brief review of as many of these papers as we could find. Based of the work of Carlyn (1977) we have determined the MBTI test to be "reasonably valid" for testing individual personality traits. Based off the work of Carlson (1985) we have to determined the MBTI to be reliable as well. Combining reliability and validity, we conclude that the MBTI is a valid source material to draw upon assessing personality.

2.2 Big Five

The Big Five Theory identifies five personality traits and was developed in the 1980s (Rothmann S, 2003). It is one of the more recent and more commonly used assessments of personality traits and typically today, psychologists will use the traits and information from this theory to identify personality traits in patients. The five factors consist of openness to experience, conscientiousness, extroversion, agreeableness, neuroticism.

Openness to experience usually indicates how one reacts to a new environment. Low scorers of openness would stay traditional and cautious while high scorers will be curious and adapt to new things. Those that are curious tend to be more open-minded and listen to other's opinions (Rothmann S, 2003).

Conscientiousness involves planning and consideration. A highly conscientious person will stay organized and make plans ahead of time while lower scorers will stay relaxed and solve problems once he/she encounters them. High scorers exhibit a tendency to act dutifully and show deliberation as well (Rothmann S, 2003).

Extroversion relates to the "primary direction for mental functioning" (Carlson, 1985) where extroverts are oriented to attach themselves to the outside world. Lower scorers, or introverts, are oriented inwards, detaching themselves from the outside world (Carlyn, 1977). Extroverts are often seen as energetic, friendly, and outgoing. Introverts are commonly seen as quiet, contemplative, and shy.

Agreeableness are the traits exhibited when one comforts or supports another. Those that score highly have a tendency towards compassion and understanding and are much more likely to exhibit "blind support" or "unconditional love". In contrast, those that are low scorers will set aside empathy in an effort to prioritize honesty and logic (Rothmann S, 2003).

Lastly, Neuroticism is how one shows confidence in front of an audience. High levels of Neuroticism can lead to anxiety and poor performance, but lower levels of Neuroticism can equally lead to poor performance. A highly neurotic person will have a tendency to overthink, over-plan, and over-analyze performance-based situations where a lower scorer will under-think, under-plan, and under-analyze the same situations (Rothmann S, 2003).

These five personality traits reportedly change only moderately over a lifespan (Chopik and Kitayama, 2018). The Big Five traits have also been leveraged in HCI research to describe differences in how CAIs and robots express behavior (Völkel and Kaya, 2021).

2.3 Emoji Use

Marengo et al. (2017) has conducted a survey to determine a correlation between emoji use and a person's score on the Big Five Factor Test. To accomplish this, Marengo et al. (2017) isolated 91 emojis and gave each participant a questionnaire asking how the participant identified with the listed emoji and then administered the Ten-Item Personality Inventory which measures the Big-Five traits. They found that 36 out of the 91 emojis used were associated with one of 3 of the 5 traits they tested: Extroversion, Agreeableness, and Neuroticism. This study proved that emojis are sometimes correlated with personality traits and provides a basis for our research.

However, the research of Marengo et al. (2017) cannot be applied to our chatbot assessments because while they account for emoji identification, they fail to include research on emoji interpretation and use. As their research stands, it helps gain data on what personality traits have tendencies to see themselves in what emojis, but we are not able to analyze these results by comparing these traits with emotional intensity or interpreted emotion. So, we can use this information to assign emojis to chatbots for use, but we cannot help chatbots interpret emojis or develop their own understanding of emoji use through AI. While their research is helpful in encouraging ours, we are called to conduct our own survey analyzing emoji use in extroverts and introverts for the purpose of including emoji use in our chatbots.

Furthermore, Krekhov et al. (2022) has performed an intensive study on the interpretation of emojis. In this study through surveys, Krekhov et al. (2022) analyzed which emojis were associated with an emotion. They then conducted a second survey which analyzed out of all emojis associated with some emotion, how intense each of these emotions were represented as based on the emoji. They found that there were 46 emojis directly associated with one emotion and mapped to what intensity that emotion was felt by each emoji. We will use Krekhov et al. (2022)'s reported emotional intensity per emoji to graph the average emotional intensity of emojis used by participants with different scores of extroversion and introversion.

2.4 Chatbot Design

Previous work has defined both a set of verbal cues derived from psycho-linguistic literature to induce different levels of extroversion implemented in a chatbot app (FlowXo) and a systematic empirical analysis of N=34 participants evaluating these chatbot personalities after interacting with them for four days each (Sarah Theres Völkel and Mayer, 2020). The result indicates that agreeableness, extroversion, and artificiality are what users look for. However, the work on the speech-based CAI personalities only takes The Big Five traits as a consideration for conversational AI, which is insufficient in designing a well rounded CAI personality experiment (Völkel and Schödel, 2020). There is also no data given here on chatbot personality type or overarching personality interpretation.

2.5 Novelty and Design

Because there is no directly applicable research on emoji use and interpretation for CAI, we will conduct emoji use experiments based on the work in emoji interpretation by Krekhov et al. (2022). Because of a lack of time and resources and the work of Marengo et al. (2017), our emoji survey will focus on the use of emojis based on just introversion and extroversion and we hope more personality traits can be isolated and tested in the future. To the best of our knowledge, this paper will be the first to look into the interpretation of emojis based on the introversion and extroversion of the user and how that correlates with the reported emotional intensity of each emoji. We hope that our research will provide greater insight into the use of emojis and their relation to different personality traits in people.

Furthermore, once our emoji survey is completed and results analyzed, we will be able to incorporate our research into the creation of our 16 chatbots. The key component to consider while building a conversational AI includes Dialog Manager, Natural Language Understanding Module and Knowledge Module, Response Generation, and Conversational User Experience Handler (Sarah Theres Völkel and Mayer, 2020). All previous successful work that has won the Alexa Prize achieved at least one of the important factors that were mentioned above (Ram and King, 2018). Different from models that generate responses that are coherent to a pre-specified existing personality or profile (Qian and Zhu, 2018), our work focuses on finding the most optimal personalities with the attempts of all 16 personalities therefore diversifying the range of options for users and removing the one-size-fits-all personality model for AI. Our work also takes a major step towards creating more human-like AI in not only personalizing the CAI experience, but using emojis to represent a more human-like CAI.

3 Approach

We will design 16 chatbots that exhibit different personalities through targeted manipulation of language. We ask participants to interact with the 16 chatbots for 15 minutes each. The chatbots will prompt the user to respond to a certain tense topic, which pushes people outside of their comfort zones, encouraging a more complex conversation. In this section, we will identify conversation flow between user and chatbot and define a plan to manipulate the different personality perceptions. We hope that this research can aid in the process of implementing a chatbot in a FlowXo bot with dynamic personality traits. FlowXo is a free online software that allows users to create their own personalized chatbots. We will be using this software to create our chatbots quickly and conveniently. To ensure they are bug-free, we pilot-tested the chatbots in several trials.

3.1 Chatbot Emoji Use

In determining characteristics of different chatbots, we wanted to include emoji use in the expression of personality traits. Because there has been no defining research on the interpretation and use of emojis based on personality traits, we will conduct our own survey of cross referencing emoji use with percent extroversion in participants to analyze how extroverts and introverts use emojis. More information on our participant pool can be found in section 3.4 Participants and A Appendix.

We have determined the work of Krekhov et al. (2022) to be valid and will use it to reference the intensity of emojis' represented emotions to determine what to what intensity introverts and extroverts use emojis. In doing this, we will appropriately address our lack of resources and time constraints while still referencing credible research to conduct our survey. The 46 emojis we will use from Krekhov et al. (2022)'s work are shown in Figure 1. Each participant will be instructed to rate their likelihood of use of each emoji based on a 7-Point Likert Scale.

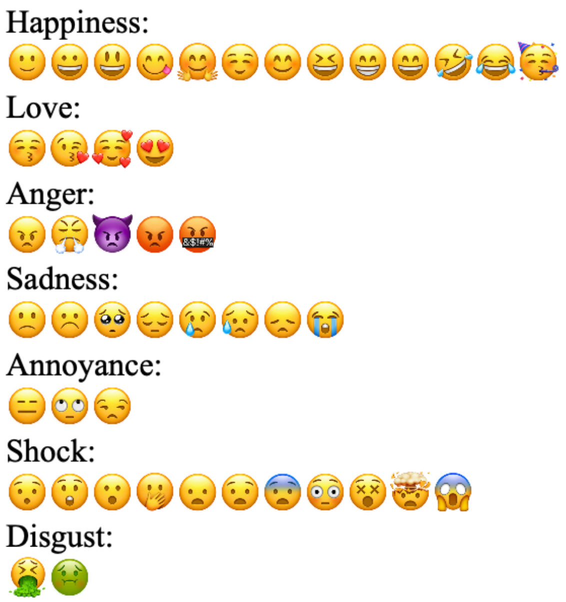

Figure 1: Emojis and their associated emotions

The end of the survey will contain a series of statements that users will agree or disagree with on a scale of 1 to 7 (referencing again the 7-Point Likert Scale). These statements have been pulled from the The Myers Briggs Foundation website which adapted them from Looking at Type: The Fundamentals by Charles R. Martin (CAPT 1997) (thi, 2019). The statements included are listed in Table 1.

After concluding the test, the participants are scored based on their extroversion and introversion in the form of a percentage (i.e. if you are 70% extroverted then you are likewise 30% intro-

Extroversion and Introversion Statements
I am seen as "outgoing" or as a "people person."
I feel comfortable in groups and like working in them.
I have a wide range of friends and know lots of people.
I sometimes jump too quickly into an activity and don't allow enough time to think it over.
Before I start a project, I sometimes forget to stop and get clear on what I want to do and why.
I am seen as "reflective" or "reserved."
I feel comfortable being alone and like things I can do on my own.
I sometimes forget to check with the outside world to see if my ideas really fit the experience.
I sometimes spend too much time reflecting and don't move into action quickly enough.
I prefer to know just a few people well.

Table 1: Statements taken from The Myers Briggs Foundation on determining extroversion and introversion. The first five statements are those extroverts tend to agree with. The second five statements are those introverts tend to agree with.

verted). In the following equation 1:

$$E_1 + E_2 + \ldots + E_5 + (\frac{7}{7} - I_1) + (\frac{7}{7} - I_2) + \ldots + (\frac{7}{7} - I_5) * 10 = P \tag{1}$$

E represents the amount one agrees with an extroverted statement (their self-reported agreement with the statement out of seven). The variables are labeled E_1 through E_5 to represent the five statements on extroversion. Similarly, I_1 through I_5 represents the amount one agrees with an introverted statement and are labeled to represent the five statements on introversion. P represents the percent of extroversion that the participant will be given. $100 - P = R$ would return the percent introversion the user scores, where R represents this percentage.

For each emoji, we will isolate the emojis that the participant identifies they have a higher tendency to use (emojis given a score of 5-7 inclusive) and average them. The following equation 2 represents this process.

$$\frac{(M_1 + M_2 + \ldots + M_N)}{N} = A \tag{2}$$

Here M represents the intensity of one emoji and A represents the average emotional intensity per user for their most frequently used emojis. So, within a single emotional category for one participant, we will consider only the emojis the participant ranked somewhere between 5 and 7 (inclusive). We will then average the intensity of these considered emojis and plot that point on a graph with the participant's percent extroversion.

3.2 Conversational Flow

At the start of the conversation, each chatbot will introduce itself, welcome the users and explain the procedure. Then the chatbot tackles the entry into a certain tense topic with the user. The tone and words used will be different based on the designated personality of the chatbot. Once the conversation reaches the end, the chatbot will ask the user for the feedback on the personality that the chatbot poses. To prevent some deviation from the ideal conversation flow, the chatbot will have some predefined responses to prompt the user to answer a question again such as: "Will you repeat that again?" or "I don't understand, just press the button." These responses should also be different based on the personality that the chatbot poses. These responses and questions will coincide with those listed in a database that stores numerous example questions of the 16 personality types, as mentioned earlier in section 2.4 Chatbot Design. All the questions and responses are designed for the chatbots before conducting the experiments.

3.3 Personality Design

To cultivate the impression of different personalities, we are manipulating the chatbots' language and tone. To this end, we first write the conversational flow for the extroverted and introverted bot as a baseline, and then adapt the responses and topics that certain MBTI personalities favor.

Each participant has to conduct the chatbot conversational flow 16 times (once per chatbot), to better distinguish the 16 chatbots for the final evaluation. Each chatbot is given a name to personalize it. Because users are most familiar with female characters for conversational agents, such as Alexa and Siri, these will be female. By standardizing this, we hope that the name of our agents will personalize our chatbots without influencing how gendered chatbots are communicated with and perceived (Holtgraves., 2011). The chatbots will be designed on FlowXo, a free online platform to create conversational flow for chatbots based on limited factors. All the conversation flow for different MBTI personalities was adapted from various databases, such as Quora.com and psychologyjunkie.com. We modified the conversation from the database and implemented emoji use into our conversation flow to design our 16 chatbots.

3.3.1 Extroverted versus Introverted Chatbot

One example of ways chatbots contrast in their personality traits through language and tone is with our extroverted and introverted chatbots. The extroverted chatbot should be enthusiastic and express its happiness with the users, writing re-

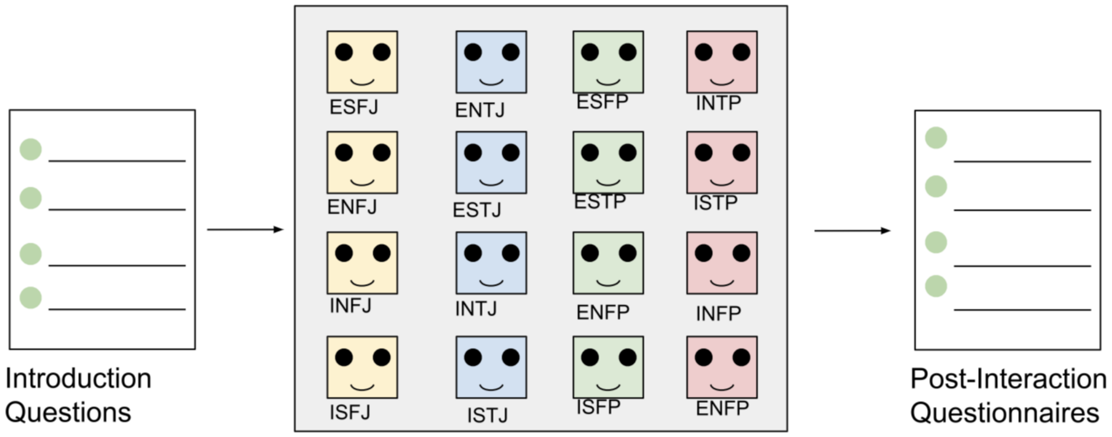

Figure 2: Conversation Flow of Experimental Design

sponses such as: "Oh my god!! I can't believe that it actually exists". Furthermore, an extroverted chatbot should also be assertive in its conversation. One thing that can also differ from an introverted chatbot is that an extroverted chatbot will use shorter text messages. Extroverted language is less formal and more loose. Extroverted people's predispositions also have a high total verbal output. Introverted people are the opposite. Some words that can be frequently used while designing an extroverted chatbot are "love", "happy", or "perfect" (Gill and Oberlander.). In this way, users will interpret our chatbot's speech as optimistic. Emojis will also show our chatbot's extroversion. Lastly the use of exclamation marks (!!!) and word expansion (writing "heeeeyyyy" instead of "hey") are more closely associated with extroverts and will aide in our extrovert representation (Dewaele and Furnham., 1999).

In contrast, introverted chatbots will have a tendency to write less emotional responses and share less personal information during a conversation (Holtgraves., 2011). The writing style is more formal (for example, writing "best regards" rather than "bye") (Ziang Xiao, 2010). We can also use more formal language throughout the conversation "such as 'transpire' or 'occur' instead of 'happen', 'encounter' instead of 'experience', or 'reflect' instead of 'think' (e.g. 'What transpired and why did you decide not to get into an argument about it?')" (Dewaele and Furnham., 1999). Also, we will use less emojis in texts since we are sharing less emotions than extroverts. When the introverted chatbot tries to answer any questions from the user, it will use goal-oriented and efficient language.

3.4 Procedure

The survey is going to be conducted through remote and in-person testing. We are only using participants who are English-first language speakers and our chatbots will only speak in English. Since our experiment is based on language interpretation, this will remove the issue of language barriers or misunderstandings. We will inform the participants about the procedure of the experiment. After signing the informed consent, the participants must fill out a demographics questionnaire stating their experience with previous voice assistants and their MBTI type. They will also fill out their expected desired MBTI of the voice agent before the experiment. Throughout the experiment, the experimenter will take notes of user feedback. After the experiment, the users will rate the chatbot's personality from 1 to 5 to indicate how much they would like to interact with the respective chatbot again in the future. Then the participants will be picked randomly and presented with all sixteen chatbots one by one.

3.5 Limitations

Due to the limitations of our work, we were only able to find participants, who are between 18 and 25 years old and pursuing a bachelor or master degree for both the emoji survey and the chatbot testing. Despite this, our emoji survey maintains validity due to the age range we are targeting according to Memon and Ansari (2021) and Weiß et al. (2020). For our survey, we hope to receive at least 50 (N=50) valid survey responses to collate and analyze. For the chatbot testing, our target is to find a population of 10 (N=10) to conduct the experiments with. All the experiment processes will be transparent to the participants.

4 Experiments

To measure the user's preference of the different chatbot personalities and influence of user personality on preference, we conducted the within-group field study. First, we conducted an experiment on the difference in emoji interpretation and use by college students based on extroversion and introversion with 33 participants (N=33). This was needed as there was no sufficient research clarifying the role personality traits play in emoji use. Then, with results from this survey, we designed the chatbots. We asked N = 10 participants to interact with 16 chatbots on their personal working station. I used FlowXo to create the conversational flow to design the sixteen chat bots.

4.1 Emoji Survey

Total Responses	Incomplete Responses	Usable Responses
47	13	34

Table 2: This table represents the responses we received after 10 days after releasing the survey link. Surveys listed as incomplete were left partially empty or were completed in under a minute.

Country of Origin	Participant Percent Distribution
America	78.72
Great Britain	14.89
Denmark	2.13
Spain	2.13
Puerto Rico	2.13

Table 3: This table represents the distribution of our participants by country of origin. We believe the country of origin should not largely affect our results as we are not looking into the difference in emoji use by country. More work on emoji interpretation by culture or country has been done by Krekhov et al. (2022).

We received 34 usable responses to collect data on for our survey of emoji use. We can see the distribution of country of origin in Table 3, but we are not separating or isolating responses based on this data. Within these responses the data we received was a number between 1 and 7 (inclusive) per emoji per participant. This number represented the likelihood of the participant using this emoji to express the emotion associated with that emoji by Krekhov et al. (2022). We also received scores between 1 and 7 that represented the participant's agreement with a given statement. There were 10 of these statements and they represented feelings of extroversion or introversion. After gathering this data, we looked at the emoji section.

By taking the average of every rank given to every emoji by every user, we found the number 3.44. This number represents the average likelihood of use of any given emoji by any given participant. Originally we had hoped that this number would be 4 as that would represent an evenly distributed emoji usage over all of our participants. We considered scaling this number and the attributed rankings so that our average would be 4, but decided against this as we did not want to misrepresent our data and agreed that 3.44 was close enough to our goal of 4. Because of this, our data is evenly distributed and represented truly as the users report. This number alludes to a possible overarching lack of use of emojis to the extent of the ranks 5 through 7 for all participants.

We also averaged the rank that each emoji was given over all participants. This number represents the average use of this emoji by our participants. In doing this, we found two emoji's had an average rank of less than two: 😲👿, with the former having a rank of 1.88 and the latter having a rank of 1.82. This told us that the average use of these emojis is between very rare and rare across all participants. Because of this we are excluding these two emojis from the results and data analysis leaving us with 44 emojis to analyze.

4.1.1 Data Organization

Emotion Expresseed	Average Likelihood of Emoji Use
Happiness	3.070
Love	4.544
Anger	3.041
Sadness	3.654
Annoyance	4.245
Shock	3.045
Disgust	4.735

Table 4: The right column is expressing the likelihood of any given participant to use an emoji within the emotional category listed from 1 to 7. This table shows us that people are more likely to use emojis to express love, annoyance, and disgust than any other emotion. They are least likely to use emojis to express happiness, anger, and shock

Next we isolated each section of the survey to calculate our values for graphing. First we looked at the results of the emoji section of the survey. For each participant we removed the emojis they rated less than 5. This left us with their frequently used emojis and we averaged the intensities of these emojis to find A from equation 2 in section 3.1 Chatbot Emoji Use. We also found the average intensity of the emojis we considered and the average of the average intensities per participant (P from section 3.1 chatbot Emoji Use). We will discuss these further in section 4.1.2 Data Representation for each emotion. The y-axis of our graph was adjusted to leave out excess white space but it did not remove any reported scores. If a participant did not rank any emojis between 5 and 7 (inclusive), then we left out their score from the graph.

Percent Extroversion	Number of Participants
0.000-25.000	0
25.001-50.000	17
50.001-75.000	17
75.001-100.000	0

Table 5: Number of participants per category of percent extroversion. All participants fell within the range of 25.001 to 75.000.

Next we looked at the section on introversion and extroversion. As explained through the equation in section 3.1 Chatbot Emoji Use, we compiled the data for each participant based on their self reported association with our 10 statements. After finding the percentage of extroversion for each participant, we calculated the average of these percentages. We found that the average participant's extroversion was 48.99. To maintain an evenly distributed network of data, we would hope for this average to be 50%. The closeness of our expected average to our calculated average shows us that our participants are evenly distributed between extroverts and introverts and adds validity to our results. We also found that none of our participants scored below 25% or over 75% for their level of extroversion, and were evenly distributed between the groups of 25.001 to 50.000 and 50.001 to 75.000. Because of this our graph's x-axis has been made to fit this range of 25 to 75.

4.1.2 Data Representation

Finally, we graphed the average intensities per participant on the y-axis and the participants matching percent extroversion on the x-axis. We used a scatter plot as the data collected per participant was not related to the other participants' collected data. We then found a trend line for our graph and it's associated equation. Our hypotheses remain that participants with a higher percent extroversion will have a higher tendency to use emojis with higher emotional intensities and that participants with a lower percent extroversion will have a higher tendency to use emojis of a lower intensity.

The slope of the equation of the trend line relates to our hypothesis and identifies if it is supported or not. A trend line with a positive slope supports our hypotheses and a trend line with a slope of 0 or a negative slope will support the opposite of our hypotheses: People with a higher percent extroversion will use emojis of a lower intensity than people with a higher percent introversion and/or people with a lower percent extroversion will use emojis of a higher intensity.

We also calculated the Pearson Correlation Coefficient (PCC) for our graph. The PCC calculates the correlation between two arrays of continuous variables. The coefficients can range from a scale of -1 to 1 and they tell us how correlated our data is. We added error bars to our graphs. These bars represent the standard deviation of our data set.

4.1.3 Putting Our Data Together

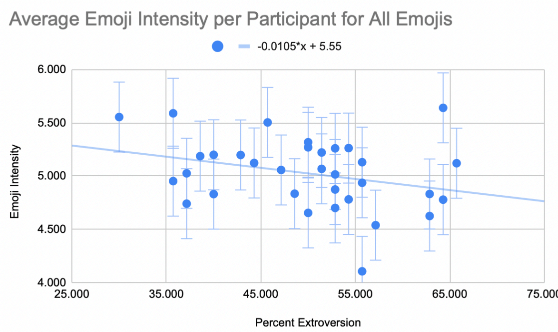

Figure 3: Graph mapping a participant's percent extroversion and the average emotional intensity of the emojis they most frequently use regardless of emotion.

Our standard deviation over all emojis is 0.328 which represents an average variation of roughly 5% between users in how they use emojis of different emotional intensities. Our Pearson Correlation Coefficent is -0.30 and therefore our data is moderately correlated. When graphing the average emotional intensity expressed by any given emoji to express any given emotion by any given participant, we see a direct downward trend. As illustrated in Figure 3, there is a clear negative slope to our graph. This leads us to conclude that the greater percent extroversion a person presents, the more likely they are to use emojis of a lower intensity. We can also conclude that the lower percent extroversion or higher percent introversion a person is, the more likely they are to use emojis of a higher emotional intensity. These conclusions are in direct contrast with our hypotheses.

In translating this to chatbot development, we have assigned chatbots who are identified as extroverted, emojis of a lower emotional intensity for use. Those chatbots who are defined as introverted have been assigned emojis of a higher rated emotional intensity, but less emojis overall. This way, extroverted chatbots may use more emojis of a lower intensity, and introverted chatbots will use much less emojis overall (in accordance with research showing that introverts tend to express less emotion over text (Holtgraves., 2011)) but those of a higher intensity when they are used.

4.2 Chatbot Experiments

4.2.1 Pre-screening and Participant Recruitment

MBTI Chabot	ESFJ	ESFP	ESTJ	ESTP	ENFJ	ENFP	ENTP	ENTJ	INFJ	INFP	INTJ	INTP	ISFJ	ISFP	ISTP	ISTJ
MAX	5	5	5	5	5	4	5	5	5	5	5	3	3	4	5	
75th Percentile	5	4.25	4.25	5	5	3.25	2.75	4.25	4.25	3.5	5	2.25	2	3.25	4.25	
MEDIAN	4.5	3	4	4	4.25	3	2	3	4	3.5	5	2	2	3	3	
25th Percentile	4	1.75	2.5	2.75	4.25	1.75	1	1	2.75	3.5	5	1	2	2	1.75	
MIN	4	1	1	2	2	1	1	1	2	3	4	1	2	2	1	

Table 6: Participants' evaluation of the sixteen chatbots.

We recruited participants from Emory University students and majors in Computer Science. To exclude some language barriers, we only distributed the survey among English-first language speakers. Before the experiment was conducted, each participant was asked to sign a consent for participation in the Interview Study. This ensured that our participants were able to remain anonymous. Since we must also consider the relationship between a user's personality and their preference for a chatbot's personality, we asked participants to take the MBTI test to see whether or not one's personality would affect his preference toward certain chatbots. Participants started the study by finishing an online questionnaire. Then, they were directed to the FlowXo chatbot we created.

When the chatbot introduced itself, participants were prompted with a few interactive questions. The conversation topic would be the same for all sixteen chatbots, but the tones and style would be different based on the conversation. After participants had interacted with all 16 chatbot versions, we asked participants to complete a final questionnaire. To refresh their memory about the chatbots they had interacted with earlier and allow comparison, we would show them a screenshot of the chatbots's welcome message, since all 16 chatbot's welcome messages are different. Once the participant finishes their survey with the chatbots, they are asked to describe their feedback of the chatbot's personality and rate their interaction desire as well as the chatbot's personality usability.

We were only able to conduct the survey with N=10 participants due to time restriction. Participants were between 17 and 23 years old (M = 22.4, $\sigma = 2.154$), and all were university students, with 60% completing their degree at Emory University, and 40% completing their degree at University of Illinois. All participants had experience with a chatbot before. On a scale from 1 = very unpleasant to 5 = very pleasant, we asked participants to rank their experience with chatbots overall. 20% of participants rated their experience as pleasant, 40% as neutral, and 40% as unpleasant.

4.2.2 Preference to Interact with the Chat-bots

Table 6 shows participants' evaluation of the sixteen chatbots. On a scale from 1 = very bad to 5 = very good. Participants expressed on average a greater desire to interact with the extroverted chatbots ($Median = 4.2, Mean = 3.71, SD = 0.96$) than with the introverted ones ($Median = 2.3$, $M = 3.15$, $SD = 1.24$). Because the data was not normally distributed ($W = 0.900, p < 0.001$), we conducted a FriedMan test, which helped to determined a significant effect of the chatbots on participants' interaction desire ($X = 4.64$, $p = 0.0331$). The Friedman test is a non-parametric test to find the difference in treatments across multiple attempts. The pairwise comparison is used to compare each of the mean differences to a critical value. Since our data is not normally distributed, we also need to apply the Pairwise Wilcoxon post-hoc test as well. The Pairwise Wilcoxon post-hoc test using Bonferroni correction did not yield a significant difference between the desire to interact with the extroverted and the introverted chatbots (extroverted vs introverted: $\rho = .413$, $r =$

0.35).

4.2.3 Interview Analysis

In general, all participants asked for more personalizing features for the designed chatbot. All ten participants were willing to spend extra money for a chatbot's personality that they favor. Three participants were in favor of chatbots adapting to the user's personality (they tended to give higher scores to the chatbots that have the same MBTI personality as themselves). One participant said, "I should adjust the information content based on my desire, not necessarily behave like me." Another one said, "People's MBTI do change as time goes by, so chatbots should keep up with the changes on my interest."

Aside from giving feedback on the chatbot, some participants resist the design of the chatbot. They knew that the chatbot is programmed, so they expressed a feeling that investing empathy in a non-human program was futile. These participants felt that the encouragement of personalized chatbots will make people rely on such devices and ultimately encourage less interaction with real humans. Those that gave higher scores on extroverted chatbots indicated an expression of desire to implement humor and humanness into chatbots. Those that gave low scores on introverted chatbots suggested that the chatbot should be formal and robotic, representing the group of participants advocating against personalized chatbots.

5 Analysis

5.1 Analysis of Emoji Use in AI

As mentioned in section 4.1 Emoji Survey, we found two emojis to be used the least: 😯👿. This tells us that these two emojis are less likely to be used by any given user. Alternatively, we found two emojis to be used far more than the others: 🙄🥰. Both emojis had an average ranking of above 5, with the former having a score of 5.382 and the latter having a score of 5.882. This tells us that these two emojis are more likely to be used by any given user.

Another fact we found is that personality has a relationship with the use of emoji. Our final analysis of our results conveys that extroverts tend to use emojis of a lesser intensity than introverts, and introverts tend to use emojis of a higher emotional intensity than extroverts. This information tells us a lot about human interpretation of emojis, but there is more work to be done in developing emojis for use of AI and chatbots.

We ultimately found that users prefer a more extroverted chatbot. These chatbots use emojis of a lower emotional intensity. This may have affected their preference of chatbot as the extroverted chatbot may have come off as calmer than an introvert a chatbot who uses less emojis overall, but those of a higher intensity when they do emote. Conducting the same experiment with and without emojis would give us more information as to which of these user preference was directly ascribed to.

5.2 Reflection Based on the Perceived Personalities

Based on the results we collected from the experiment, chatbots with a more cheerful, optimistic tone tend to be perceived as agreeable, extroverted, and open. These chatbots used emojis of a lower intensity which might have made them appear calmer. People tend to favor these type of chatbots, regardless of whether or not the user is extroverted. This experiment did demonstrate, however, that a chatbot's personality will affect user preferences and interaction behavior. We also found that most users enjoy interacting with human-like personalities.

Some feedback we received from participants highlighted the reasons for their ranking of the chatbots. By talking to ESFJ or ESTJ, most participants had a sense of talking to an actual human. Additionally, the introverted chatbots that focus on conciseness come off as robotic. One factor that also contributed to one's favor toward a certain chatbot is the word count. This also supports the idea that participant's favor anthropomorphic qualities in chatbots. Therefore, the chatbots' perceived humanness may determine not only subjective preferences for chatbot personality but also length of responses to chatbots. This confirmed and provided evidence for our previous hypothesis that people tend to favor extroverted chatbots.

5.3 Importance of Agreeableness in a Conversation AI

Our extroverted chatbots tended to focus more on agreeableness. Since there are studies that indicated that agreeableness is inter-correlated with extroversion in human personality, the high rankings for extroverted chatbots were what we expected (Ziang Xiao, 2010). Therefore, we conclude that controlling only one personality dimension is insufficient, and that personality manifestation must be considered as part of the multi-dimensional construct. When we are designing an actual conversational AI, we should take this in note to guarantee that operating only one dimension does have intended consequences on another one. Conclusively, agreeableness should be considered in interpersonal interaction while developing a voice assistant. Future experiments should focus on what is the actual effect of users' preference for the perceived extroversion, agreeableness, true extroversion or both.

One participant's feedback expressed a dislike of human-like behavior in CAI. This is due to the participant's view of the reliability of technology. This user felt that if the chatbot failed to interpret the user's emotion, it would affect the user negatively. Furthermore, the user considered all CAI as fake identifying that there should be a clear difference between humans and AI. In summary, we think there likely is no personality that universally fits all users as this one participant demonstrates. It would be quite complex for a computer to master all personalities, but creating customization features where users are able to manipulate the setting of when and how the AI should use different personality traits might be a possible and even more optimal plan.

5.4 Future Recommendations

The data we found on emoji use can be easily transcribed into chatbot production because of their numerical quality, however, we recommend that our emoji experiment be repeated on a larger scale and with a much larger participant pool to ensure validity. We would also encourage future researchers to develop data on emoji use and interpretation based on different personality traits as well. By encouraging this research in emojis, we can make data on emojis more readily available to AI developers for incorporation into text-based AI communication.

Although we are able to conduct a survey by creating a personality manipulation with 16 personality types, converting these verbal cues from actual human behavior was not enough to fully display an introverted chatbot. Our experiment expressed that it is hard for one to see the difference between some personality types that started with "E", such as ESFJ and ESFP. This can be proved by the identical tone and wordings from the ESFJ and ESFP chatbots. And even though there are some stereotypes for an introverted personality (formal wordings, conciseness, etc.), these stereotypes were not considered part of being introverted, in fact, it is part of chatbot's conscientiousness.

One reason that it is difficult for one to identify different types of extroversion could be the similar word use and interactive responses across the eight chatbots. For example, the introverted chatbot would not start a conversation with the user but only respond if the user messages first. Future work with CAI's can help explore and define these nuances especially if researched with voice agents rather than text-based agents. The incorporation of voice into CAI research would allow for clearer representation of personality traits.

6 Conclusion

Although there are many previous works demonstrating that conversational AI has a positive impact on some specific personalities with user experience, there has not been any text related to personality perception on MBTI of merely text-based chatbots. To make up for this gap, we designed a list of the most common verbal cues used by different MBTI personalities and conducted a survey on emoji use in extroverts and introverts. We used this data alongside the FlowXo program to induce different types of MBTI personalities in chatbots. Furthermore, we conducted an ethical experiment and empirical analysis of N=10 participants to evaluate their feedback regarding the chatbot experiments.

The work we have done in this paper gives us more insight into the difference in emoji interpretation between introverts and extroverts. Not only could we teach AI how to identify personality traits in emoji users by their communication, but we could also give technology more personality by programming it to use emojis in a way that conveys personable or more anthropomorphic personality traits. We also found that participants preferred chatbots that are extroverted rather than introverted. And furthermore, we found that one main reason that users favored extroverted chatbots was due to their agreeableness and engagement. However, the lack of non-verbal cues for introverted chatbots led to a lack of thoughtful responses on said introverted chatbots. This leads us to conclude that research on user preference of introverted conversational agents should be explored with non-text-based CAIs, but in a broader view, our experiment shed lights on 16 different human-like MBTI personalities that can be adapted to one's favor to further the overall interaction experience with chatbots.

This research could aide development of voice assistants and help designers to consider these personality traits in the future. While our experiment focused on text-based agents, future researchers should look into users' personality preference in voice-based agents. One can modify an agent's voice characteristics such as gender, dialects, or tones to take more attributes into consideration. We also found that a user's personality has an effect on their preference of a certain chatbot personality, and this could lead to development of a more dynamic CAI personality model based on user personality.

Users have been shown to be more open with a conversational AI than with a human listener in reporting mental health symptoms, and in some experiments these agents have been successfully used to treat persecutor delusions for people with psychosis (Craig TKRus-Calafell M, 2018). Encouraging openness about mental health issues is a major sector of psychology today and if users are more open with a computer than a human, they will be further encouraged to open up about their mental health when chatting with conversational AI. Here we can see that the development of conversational AI does not just have implications in a communicational sense, but also in the fields of psychology, psychiatry, and more.

It is important to note that we were not able to design a well-rounded experiment on the user's preference on conversational AI because of our small participant pool and lack of appropriate time and funding. Future researchers should expand and replicate our experiment accounting for and mitigating these limitations.

All our resources on our emoji survey, including the dataset, models, and survey link, are available through our open source project link: https://github.com/VelvetJumper/EmojiInterpretation.git

Acknowledgements

We would like to acknowledge Dr. Jinho Choi who helped guide this research.

References

2019. Myers-briggs type indicator® (mbti®) | official myers briggs personality test. *Journal of Personality*.

John G. Carlson. 1985. Recent Assessments of the Myers-Briggs Type Indicator. In *Journal of Personality Assessment*, volume 49, pages 356–365.

Marcia Carlyn. 1977. An assessment of the myers-briggs type indicator. *Journal of Personality Assessment*, 41(5):461–473.

William J. Chopik and Shinobu Kitayama. 2018. Personality change across the lifespan: Insights from a cross-cultural longitudinal study. *Journal of Personality*.

Jonathan Steuer Cliford Nass and Ellen R. Tauber. 1994. Computers are social actors. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Boston, Massachusetts, USA) (CHI '94))*.

Gerald Corr, Philip J.; Matthews. 2009. The cambridge handbook of personality psychology (1. publ. ed.). *Cambridge University Press*.

Ward T Craig TKRus-Calafell M. 2018. Avatar therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. *Lancet Psychiatry*.

Jean-Marc Dewaele and Adrian Furnham. 1999. Extraversion: The unloved variable in applied linguistic research. *Language Learning 49*, pages 509–544.

Alastair J. Gill and Jon Oberlander. Taking care of the linguistic features of extraversion, year = 2002. *Twenty-Fourth Annual Conference of the Cognitive Science Society*, pages 363–368.

Thomas Holtgraves. 2011. Text messaging, personality, and the social context. *Journal of Research in Personality*, pages 92–99.

Andrey Krekhov, Katharina Emmerich, Johannes Fuchs, and Jens Krüger. 2022. Interpolating Happiness: Understanding the Intensity Gradations of Face Emojis Across Cultures. New Orleans, LA, USA. ACM.

Davide Marengo, Fabrizia Giannotta, and Michele Settanni. 2017. Assessing personality using emoji: An exploratory study. In *Personality and Individual Differences*, volume 112, pages 74–78.

Maresa Biermann Martin Jentsch. 2019. Talking to stupid?!? improving voice user interfaces. *Mensch und Computer 2019 - Usabil- ity Professionals*.

Mahrukh Memon and Sanaullah Ansari. 2021. Investigating Different Age Groups' Preferences and Purposes of Emoji Use. In *Sukkur IBA Journal of Educational Sciences & Technologies*, volume 1, Jamshoro, Sindh, Pakistan.

Huang M. Zhao H. Xu J. Qian, Q. and X. Zhu. 2018. Assigning personality/profile to a chatting machine for coherent conversation generation. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, pages 4279–4285.

Prasad R. Khatri C. Venkatesh A. Gabriel R. Liu Q. Nunn J. Hedayatnia B. Cheng M. Nagar A. Ram, A. and King. 2018. Developing a personality model for speech-based conversational agents using the psycholexical approach. *CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20).*

Coetzer EP Rothmann S. 2003. The big five personality dimensions and job performance". *SA Journal of Industrial Psychology.*

Lale Kaya Sarah Theres Völkel and Sven Mayer. 2020. A user perceptions of extraversion in chatbots after repeated use. *CHI Conference on Human Factors in Computing Systems (CHI '22).*

Bilge Mutlu Sean Andrist and Adriana Tapus. 2015. Look like me: matching robot personality via gaze to increase motivation. *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, page 3603–3612.

J.L Szalma. 2009. Individual differences in human-technology interaction: incorporating variation in human characteristics into human factors and ergonomics research and design. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Boston, Massachusetts, USA) (CHI '94))*, page 381–397.

Paul D.; Barbara Barron-Tieger Tieger. 1990. The art of speedreading people. *New York: Little, Brown and Company*, page 66.

Sarah Theres Völkel and Lale Kaya. 2021. Examining user preference for agreeableness in chatbots. *3rd Conference on Conversational User Interfaces (Bilbao (online), Spain) (CUI '21).*

Sarah Theres Völkel and Ramona Schödel. 2020. Developing a personality model for speech-based conversational agents using the psyc- holexical approach. *CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20).*

Martin Weiß, Dariana Bille, Johannes Rodrigues, and Johannes Hewig. 2020. Age-related differences in emoji evaluation. *Experimental Aging Research*, 46(5):416–432.

Vera Liao Ziang Xiao, Michelle X. Zhou. 2010. Tell me about yourself: Using an ai-poweredchatbot to conduct conversational surveys with open-ended questions. *ACMTrans. Comput.-Hum.*

A Appendix

None of our participants were paid for their time. They were informed before taking the survey and chatbot experiment that their participation was fully voluntary and no compensation would be provided. By taking the survey, participants were agreeing that their participation was voluntary. All participants signed a consent form, which protects participant's privacy, before the chatbot experiments began.