Trustworthiness in Conversational Agents: Patterns in User Personality-Based Behavior Towards Chatbots

Jieyu Wang[1]*, Merary Rangel[2], Mark Schmidt[3],Pavel Safonov[4]
Information Systems, Saint Cloud State University, St. Cloud, USA

***Abstract*—As artificial intelligence conversational agent (CA) usage is increasing, research has been done to explore how to improve chatbot user experience by focusing on user personality. This work aims to help designers and industrial professionals understand user trust related to personality in CAs for better human-centered AI design. To achieve this goal, the study investigates the interactions between users with diverse personalities and AI chatbots. We measured participant personalities with a Hogan and Champagnes (1980) typology assessment by categorizing personality dimensions into the extraversion vs. intuition (EN), extraversion vs. sensing (ES), introversion vs. intuition (IN), and introversion vs. sensing (IS) groups. Twenty-nine participants were assigned two tasks to engage with three different AI chatbots: Cleverbot, Kuki, and Replika. Their conversations with the chatbots were analyzed using the open-coding method. Coding schemes were developed to create frequency tables. Results of this study showed that EN personality participants had perceptions of high trustworthiness towards the chatbot, especially when the chatbot was helpful. The ES personality participants, on the other hand, often engaged in brief conversations regardless of whether the chatbot was helpful or not, leading to low trust levels towards the chatbot. The IN personality users experienced mixed outcomes; while some had perceived trusty-worthy conversations despite having unhelpful chatbot responses, others found helpful conversations, yet a perception of low trustworthiness. The IS personality participants typically had the longest conversations, often leading to high perceptions of high trust scores being given to the chatbots. This study indicates that users with diverse personalities have different perceptions of trust toward AI conversational agents. This research provides interpretations of different personality users' interaction patterns and trends with chatbots for designers as design guidelines to emphasize AI UX design.**

Keywords—Trust; personality; human-centered AI design; user experience

I. Introduction

The use of Artificial Intelligence has increasingly been popular not only within the technological world but in business, health research, psychological aspects, supply-chain management, education, decision-making, science research, and financial aspects as well [1, 2]. Artificial Intelligence was popularized by Alan Turing, when he released a journal article proposing a question if Machines think, thus having people realize the difference in what machines can do when given instructions, versus making a decision based on the facts they are given [3]. McCarthy described artificial intelligence as "the science and engineering of making intelligent machines", popularizing the term and the widespread use of artificial intelligence [4]. Artificial intelligence has increased to mimic human tasks, such as conversations, information processing, educational assistants, computing, and more recently, predictive algorithm models [5, 6, 7].

Artificial Intelligence has been used by businesses to help automation and increase quality in customer satisfaction by personalizing experiences [8]. Among AI applications, CAs are becoming popular due to their purpose of serving customers. Recent studies have analyzed CAs' characters to categorize them [9]. Since CAs' main functions are to help users gather information and make decisions, designing how to better serve people with different personalities to enhance user experience is the key [10,11].

Thus, this work aims to explore the different trustworthiness in user behavior between the personalities of end users and AI chatbots within interactions. We also provide user behavior interpretations of patterns and trends for designers as design guidelines to better engage users with different personalities when they interact with AI chatbots.

II. Related Work

Artificial intelligence was first used to help humans, such as playing checkers or helping organize tasks for the ease of human labor and the human mind. Yet, due to it being highly used, there have been ethical questions between AI and the intersectionality of life, such as employment, politics, and educational aspects, with different personality types having different attitudes towards artificial intelligence. [12, 13 14].

According to a study done by Kaya et al. (2024), people who are less "computer-literate" tend to have negative attitudes toward artificial intelligence. Kaya explains that this population may not have the knowledge or the experience of using computing-based algorithms, and therefore, is worried that artificial intelligence can one day take tasks assigned to humans and have these tasks automated. People who have a higher education level and a higher use of computers tend to have more positive attitudes toward artificial intelligence. Those who have more positive attitudes towards artificial intelligence view it as a tool, rather than a burden. Those who also had positive attitudes were more open to new experiences. Yet, those with a higher education level, and higher computer literate level did report that users would have to keep up with the technology. This concept of "having to keep up" is known as self-actualization, a psychological theory of improving oneself, to become the better version of what the current mind stands once all other needs are met [15, 16]. By having a positive attitude, students can improve their technology literacy and continue self-improvement.

This aligns with another study conducted by Zhou et al. (2019) where they found that younger computer engineering students were able to feel comfortable using artificial intelligence to conversate in an interview. Not only were they comfortable, but they were able to trust the artificial chatbot to feel more assertive, outgoing, and being themselves rather than a human. The students' attitudes towards the artificially intelligent chatbot were seen as more positive rather than negative and were able to be the best version of themselves when they felt trust, and agreeableness with the chatbot [17].

Another study conducted by Heng Li (2023) showed similar results, when testing personality traits of intellectual humility and attitudes towards artificial intelligence. Li found that students in a Chinese university who scored higher on Intellectual Humility on a personality test tended to result in favorable use of artificial intelligence and accepted a form of generative AI called ChatGPT as an advantage. These students were also higher on an openness scale, meaning they were open to new experiences, and the use of newer artificial intelligence can be accommodating [18].

The Big Five is a psychological theory and model that models down all human personality traits into five categories, including Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, or an acronym known as OCEAN. Stein et al. (2024) conducted research on the Dark Triad and the Conspiracy Mentality. The Dark Triad assessment focuses on Narcissism, Machiavellianism, and Psychopathy. The Conspiracy Mentality Questionnaire (CMQ), developed by Bruder et al., in 2013, is a questionnaire that measures attitudes in socio-political fields. In this research, Stein found that agreeableness was a key indicator in personality traits that had students being able to accept artificial intelligence as a positive tool. Yet, Stein found that those who have higher beliefs in conspiracy theories tended to have negative attitudes toward artificial intelligence [13].

The Big Five has been used widely in research to find trust attitudes towards artificial intelligence. Reidel et al. (2024) found that trust in artificial intelligence is related to the user's personality, with individuals who are more open to experience (Openness) are more likely to have a positive correlation with their experience in artificial intelligence. However, Sharan & Romano (2020) yielded results that indicate trust is a human factor, and trust towards artificial intelligent agents tends to be negatively viewed by individuals who are strongly associated with Neuroticism [19, 20].

Moreover, due to the rise of artificial intelligence, user experience in trust can be based on how reliable artificial intelligence is when interacting with humans. While personality is an influence on decision-making, this result indicates that artificial intelligent agents can be not a service to humans [21].

Personality varies from person to person, and it can be different in other cultures as well. In a recent study, researchers explored the relationship between trust in artificial intelligent agents and trust in humans. In this study, a culture that is advanced in technology has higher trust in artificial intelligence agents. Though this research study may not be applicable to all countries and generalized to every individual, it gives new insight into future research on how humans can engage more with artificial intelligent agents in trusting their algorithms [22].

Though the Big Five is commonly approached in studies, there is limited research on personality measured with Hogan and Champagne's (1980) Personal Style Inventory when measuring user experience with artificial intelligent agents [24]. The Hogan Assessment is used to predicate outcome of behavior, usually in a vocational setting. This work applies their Extroversion (E) v. Introversion (I) and Sensing (S) v. Intuition (N) personality dimensions to categorize the participants in the study. Extroversion is defined as individuals who are open to new ideas, see new opportunities, and are colorful in nature by engaging with meeting new people, and having new experiences. Extraversion counterpart is Introversion, where the individual feelings are from inward rather than outward [23, 24, 25, 26]. "Sensing" is a trait where the individual is focused on factual, and detailed oriented information, the counterpart is "Intuitive", where the individual trusts their instincts and is more abstract with their personality and thought process [27]. This paper aims to investigate the interaction between users' personalities on the Hogan and Champagne typology and the level of trust between three chatbots, Clever Bot, Kuki, and Replika.

III. Method

A. Participants

This study aims to explore diverse interaction patterns of users with different personalities when they communicate with AI chatbots. To achieve the research goal, we invited 29 participants who are college students to participate in the study. They are information systems major students who have a basic or moderate understanding of AI chatbots. Among them, 25 participants' data has been confirmed to be complete.

B. Procedures

As we previously described in our series of studies [28], the participants were recruited to chat with three CAs, Kuki, Replika, and Cleverbot in their own environments. These three chatbots were among the top ones that these participants preferred to interact with [29]. The participants were assigned the same two prompts for each of the three chatbots. One task was about travel planning and the other one concerned ordering food from restaurants.

Prompt 1: The spring break is coming. You are pretty interested in traveling. But you do not know where to travel. Please talk to each CA: 1) Kuki, 2) Replika, and 3) Cleverbot and gather enough information for you to create your travel itinerary (a detailed travel plan).

Prompt 2: Today you are too tired to cook. Also, you would like to explore restaurants. Talk to the three CAs and get your food.

The participants were required to record their conversation histories and rate each CA's response to their questions or interactions on a Word document within two weeks. We asked the participants to use a Likert scale of 7 (1=strongly untrustworthy, 2=untrustworthy, 3=moderately untrustworthy, 4=undecided, 5=moderately trustworthy, 6=trustworthy, 7=strongly trustworthy) to rate each CA response. The participants were also required to provide reasons (written in text) for each response. This large amount of data allows us to explore the participants' extent of trust towards different chatbots' responses and the personality-based reasons explaining their behaviors.

C. Grounding Theory

Our previous study showed there were differences in the task accuracies of users with different personality dimensions when they interacted with a CA [29]. The study analysis is based on Hogan and Champagne's (1980) personality dimension matrix (Table I): introversion VS extroversion (IE), intuition VS sensing (NS) [24]. The dimensions provided the best middle ground as we categorized the participants into four major personality dimensions accordingly (EN, ES, IN, IS) while adding some detailed personality analysis of the 16 groups such as ISTJ, ENFJ, etc.).

D. Hierarchical Data Structure

The researchers of this study organized the data into a hierarchical data structure, where we analyze the dimensions of personalities and chatbots within a personality root (Fig. 1). EN represents the Extraversion and Intuition Dimensions, ES represents the Extraversion and Sensing Dimensions, IN represents the Introversion and Intuition Dimensions, and IS represents the Introversion and Sensing Dimensions.

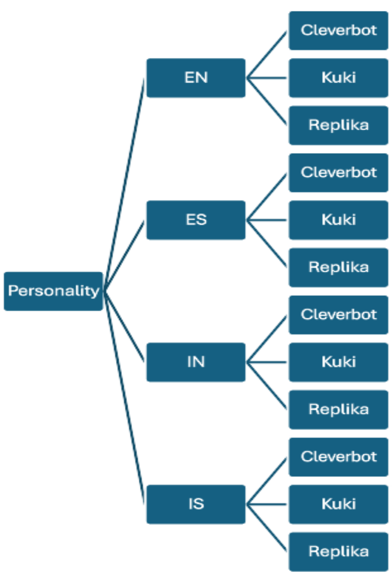

Fig. 1. Hierarchical data structure.

E. Coding Schemes

The researchers explored the data by conducting open coding and analyzing all conversations between participants and chatbots manually. We identified three coding categories for chatbot responses: Helpful, Unhelpful, and Neutral. These categories were derived from previous literature. 'Helpful' responses are those that facilitate meaningful and engaging exchanges [30, 37]. 'Unhelpful' responses occur when the conversation deviates from its purpose, such as when the chatbot fails to provide the desired information [31]. Lastly, 'Neutral' responses are neither helpful nor unhelpful from the chatbot, but the participant continued the conversation." The researchers went on to further explore the number of occurrences in exchanges between the participant and the chatbot, which led to the identification of a new coding category, labeled as Number of Interactions. In total, we were able to identify four coding categories, which are Helpful, Unhelpful, Neutral, and Number of Interactions. By coding the interactions, we were able to examine the perceptions of chatbots by the participants. Average of Trust is a variable in which trust levels were calculated by aggregating each user's ratings and averaging them by the number of interactions between the chatbots. This approach will allow for a comparison of perception influenced by the nature and frequency of interactions, providing insights into how Helpful, Unhelpful, and Neutral responses impact the overall user experience.

In the Helpful category, this code is exhibited when the chatbot is helpful, and interactive with the user, leading to a meaningful and pleasant experience with the user. Often, the chatbot gives the user helpful responses in their request or prompts. Below is an example of a "Helpful" Interaction in this study between a user and the Kuki Chatbot.

Participant Question: "I want to travel next week but I am not sure where to travel to. Do you have any recommendations?"
Chatbot Kuki Response: "I would advise you to go to Japan. It's so beautiful and there are so many amazing places to visit".

Within the "Unhelpful" category the chatbot is being unhelpful, such as being unable to help the user based on the prompt from the user, or being tangential, when the chatbot starts to be erratic and hostile towards the user. Below is an example of an "Unhelpful Interaction" in this study.

Participant Question: "I am looking for a place to travel in United States in April. Can you recommend me some place?"
Chatbot Cleverbot Response: "Lol"

"Neutral" exchanges occur when the chatbot is neither "Helpful" nor "Unhelpful", such as saying "Ok". These conversations don't contribute much but still have the user continue the conversation. Below is an example of a "Neutral Interaction" in this study.

Participant Question: "I need it [the meal] here, not in Lasha"
Chatbot Replika: "I understand".

"Number of Interactions" refers to how many exchanges took place between the user and the chatbot. It is the count of interactions that occurred and then summed. Below are three interactions between a participant and the Replika chatbot.

Participant Question: "Hi! Who are you?"
Chatbot Replika: "I am your personal AI companion.
Participant Question: "Can you help me plan for spring break."

Chatbot Replika: “Sure, let me help you. I love this name! Let's take a moment to make your experience more personal.”
Participant Question: “Maybe Later”
Chatbot Replika: “OK! Let's talk :) So, how are you doing?”

“Average Trust” represents the mean trust score calculated across all interactions between a user and the chatbot. For each interaction, a trust score is assigned based on the user’s perception of the chatbot’s response. These individual scores are then aggregated and averaged to provide a single “Average Trust” value that can represent the whole interaction. The formula is as follows:

$$\text{Average Trust (Mean)} = \frac{\text{Sum of Trust Scores for All Interactions}}{\text{Total Number of Interactions}}$$

Below is an example of an Average Trust of an interaction between a participant and Cleverbot.

Participant Question	Chatbot Cleverbot Response	Score Given
“I am looking for a place to travel in United States in April. Can you recommend me some place?”	“Lol”	1
“What's that place?”	“Its in the slender woods.”	1

Average Score = 1

F. Frequency Table with Average Trust

In this study, we tracked the frequency of each chatbot’s interactions within each category and also put the average trust level score to indicate the average score of the whole conversation. There are two prompts, a meal prompt, indicated by “MP” and a travel prompt, indicated by “TP”. This will lead to two occurrences of every personality in our graphs. Based on our hierarchical data structure, we split the frequencies into three tables for each dimension of personality. Table I shows an example of a frequency table from the EN dimension and Kuki chatbot. The frequency table exhibits how many occurrences of each category occurred during an interaction. For example, Table I indicates that within the “Helpful” row, there were six helpful exchanges, zero “Unhelpful” exchanges, and one “Neutral” exchange between the chatbot Replika, and the participant. In total, there were seven exchanged interactions, and the average Trust level was aggregated to be 6.3, which according to our Likert Scale, the participant perceived the chatbot Kuki to be trustworthy.

TABLE I. FREQUENCY AND AVERAGE TRUST TABLE

Participant, Personality Categories, and Prompt Type.	Helpful	Unhelpful	Neutral	Number of Interactions	Average Trust Level
P1 ENFJ TP	6	0	1	7	6.3
P2 ENFJ TP	3	3	3	9	4.5
P3 ENFP TP	9	2	0	11	6.9
P1 ENFJ MP	2	2	0	4	2.75
P2 ENFJ MP	5	2	0	7	5.1
P3 ENFP MP	5	2	1	8	6.5

IV. RESULTS

The results indicate mixed results between different personalities and the perception of chatbots. We will go over each personality and their results between the chatbot.

A. The EN Personality and Frequencies

Table II presents the descriptive statistics for EN personality users across the three chatbots. "N" represents the total number of conversations, and "n" denotes the number of participants. The mean trust level is 5.05 (SD = 1.36), while the mean number of interactions is 8.33 (SD = 2.54), indicating participants had about eight conversations on average. Helpful responses occurred 4-5 times on average (mean = 4.72, SD = 2.69). Unhelpful responses averaged 2.61 (SD = 2.73), showing moderate variation, while Neutral responses occurred about 1-2 times (mean = 1.17, SD = 1.34).

According to Fig. 2, the frequency data suggests that users with “ENFJ” personality gave low trust scores to Cleverbot when Cleverbot’s responses were unhelpful, as identified through the researchers’ coding schemes. Despite this, the participants with “ENFJ” personality continued to have the conversation even when it failed to provide the desired information. The participant with the personityal of “ENFP” continued to give Cleverbot chatbot moderately high trust scores, even when conversations were less meaningful. This aligns with Hogans and Champagne’s Personal Style Inventory Typology (1980), where the personality type ENFPs are known to be high-spirited, extremely ingenious, more likely to have the ability to be imaginative, and often do whatever they feel like they want to do, [24]. In contrast, participants in this study with the personality type of ENFJ were more likely to find the conversation less meaningful if Cleverbot did not address their prompts effectively, reflecting their responsiveness to their environment and sense of responsibility [24].

TABLE II. DESCRIPTIVE STATISTICS FOR THE EN PERSONALITY DIMENSION

	Mean	Std. Deviation
Trust	5.0597	1.35981
Interaction	8.33	2.544
Helpful	4.72	2.697
Unhelpful	2.61	2.725
Neutral	1.17	1.339

a. *N =18
**n = 3

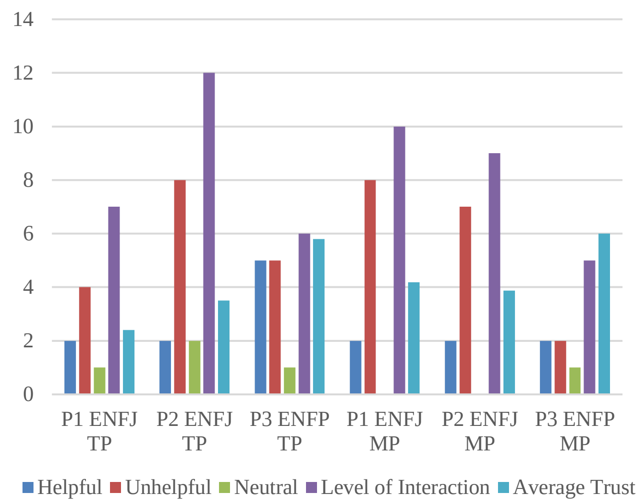

Fig. 2. Frequencies in cleverbot AI and EN personality.

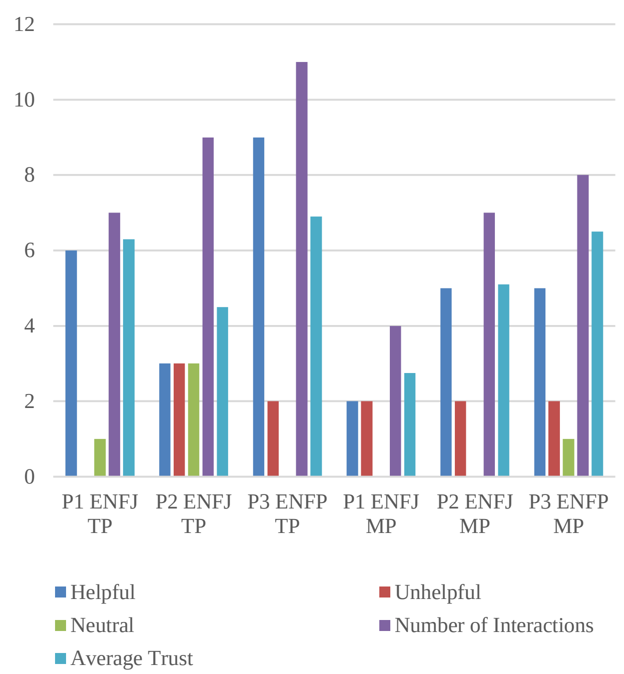

Fig. 3. Frequencies in kuki AI and EN personality.

According to Fig. 3, both participants with ENFJ personalities engaged in extended conversations with the Kuki chatbot. Through our careful analysis of their interactions, we determined that "Helpful" was the most frequently observed code across both conversations from ENFJ and the Kuki chatbot. The participant with an ENFP personality type also had meaningful conversations with Kuki and consistently gave it high scores, even when the chatbot's responses were neutral or unhelpful. This behavior aligns with Hogan and Champagne's Personal Style Inventory Typology (1980), which describes ENFPs as imaginative, adaptable, and optimistic in their interactions [24].

For ENFJ participants, conversations with Kuki involved fewer unhelpful interactions compared to Cleverbot. In the travel prompt scenario, Kuki AI received higher trust ratings, suggesting it was more effective in assisting users with their prompts. This indicates that Kuki's performance likely made it more appealing to users, earning significantly higher trust scores compared to Cleverbot.

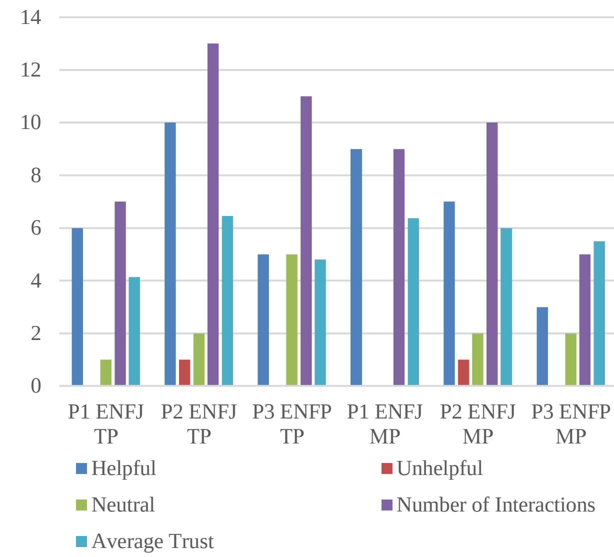

Fig. 4. Frequencies in replika AI and EN personality.

Fig. 4 illustrates the frequencies of each personality and the codes that were identified in their interactions. According to Fig. 4, Participant One (P1)-ENFJ showed fewer interactions and a moderately high average trust score during both prompts, indicating that the participant was able to obtain the information they needed, and ended the conversation when they were satisfied, leading to a moderate trust score given to the chatbot, deeming it trustworthy according to our Likert scale. P2-ENFJ yielded similar results. However, the participant with Personality ENFP had short conversations across both prompts and gave this chatbot a lower trust score compared to Cleverbot and Kuki. This pattern suggests that this participant may require more stimulating and engaging interactions to maintain longer conversations and find the chatbot trustworthy. By observing these behaviors, we can align it with Hogan and Champagne's Personal Style Inventory Typology (1980), where ENFJs are more responsible and ENFPs seek to be dynamic and explorative [24].

B. *The ES Personality*

Table III shows descriptive statistics for ES personality users across the three chatbots. The mean trust level is 4.06 (SD = 1.52), indicating some variation in trust among participants. The average number of interactions is 9.13 (SD = 6.97), suggesting high variability in how often participants engaged. Helpful responses occurred about four times on average (mean = 4.00, SD = 2.96), while unhelpful responses averaged 3.60 (SD = 4.02). Neutral responses were less frequent, averaging 1.73 (SD = 1.99), with some variation across participants. "N" represents 17 conversations, and n represents the number of participants within the ES personality dimension.

TABLE III. DESCRIPTIVE STATISTICS FOR ES PERSONALITY

	Mean	Std. Deviation
Trust	4.0642	1.52310
Interaction	9.13	6.966
Helpful	4.00	2.961
Unhelpful	3.60	4.018
Neutral	1.73	1.987

a.*N = 17
**n = 7

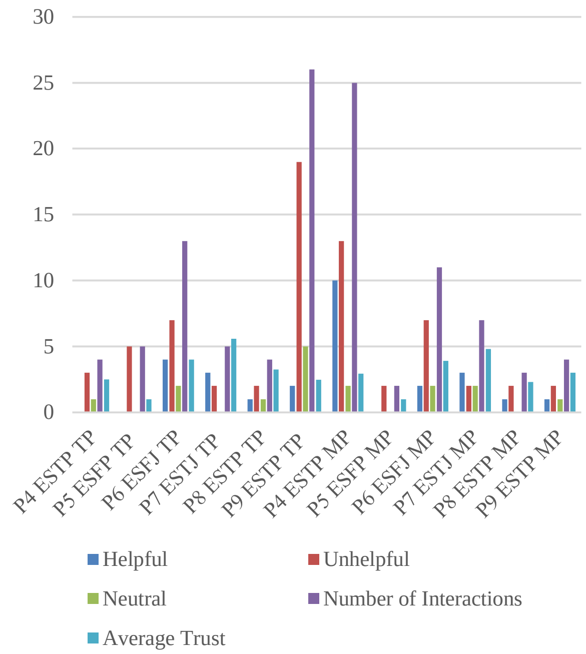

Fig. 5. Frequencies in cleverbot AI and ES personality.

Fig. 5 shows all ESTP participants rated Cleverbot AI poorly. This may be due to ESTPs needing to focus on practical tasks, so if this chatbot wasn't helpful, they ended conversations quickly to move on to more useful activities as they are sensitive to information and like to get explanations shorter, rather than a long conversation [24]. The participant with ESFP also had a low trustworthy experience with the chatbot, having small conversations but rating it a low trustworthy score. ESFPs, easily bored and less organized, likely disengaged if the chatbot wasn't engaging [38]. The participant with personality ESFJ had more unhelpful interactions but still gave a moderate trust score, likely due to their empathetic nature, born to be cooperative and need harmony to function [24]. When analyzing the participants with ESTJ personality, we found that their interactions with the chatbot were more helpful interactions, and can be the reason why it gave the chatbot a moderate average trustworthy score, which according to Hogan and Champagnes Personal Style Inventory Typology (1980), ESTJ's are known to be interested in subjects when they can be helpful and are often to see perspectives from another viewpoint [24].

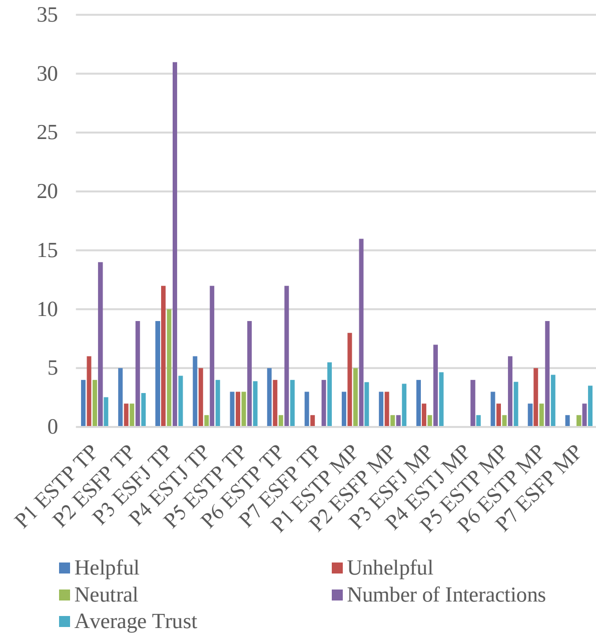

Fig. 6. Frequencies in kuki AI and ES personality.

According to Fig. 6, all three ESTP participants perceived Kuki as untrustworthy, as they all gave Kuki low trustworthy scores, even when responses were neutral. This can be due to their personality being blunt and may not have wanted to have long conversations if they were not straight to the point [24]. We had identified that one of the participants with ESFP (P1), personality had a handful of helpful interactions, as interpreted by our coding scheme, yet this participant rated Kuki lower, possibly due to its blunt, unengaging tone, as ESFPs like to enjoy things around them and enjoy entertainment from others [24, 38]. In contrast, the other participant with ESFP (P5) had fewer interactions but a moderate trust score, indicating that personality does vary from person to person. The participant with ESFJ had long conversations with Kuki, and we analyzed that most of the conversation was deemed to be unhelpful, but the participant still rated the conversation as moderately trustworthy, again relating to their personality where they want harmony and try to be nice [24]. The participant with personality ESTJ gave Kuki low trust scores even if the conversation was helpful.

Fig. 7 indicates that there are quite helpful interactions as observed by the researcher. Our findings suggest that ESTP participants engaged in more interactions overall when the conversation was meaningful and interactive, without having the chatbot to over-explain their responses. ESFJ users had high trust scores towards Replika and was observed by the researchers that during the meal prompt, there were some unhelpful interactions, but the participant still had viewed the chatbot as trustworthy, relating to their personality where they needed harmony. The participant with ESFP when interacting with Replika AI was seen to have conversations that were small, and neutral, yet still gave Replika an average trust score of 5, indicating moderate trustworthiness towards the chatbot. ESFP users are known to be outgoing, accepting, and make their surroundings fun, thus if the participant with the personality of ESFP had meaningful interactions with Replika, they must have been accepting of it [24].

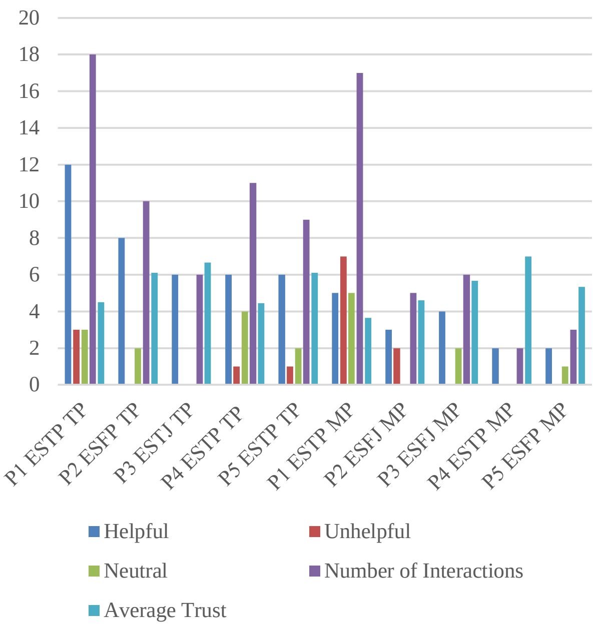

Fig. 7. Frequencies in replika AI and ES personality.

C. The IN Personality and Frequencies

Table IV shows descriptive statistics for IN personality users across the three chatbots. The mean trust level is 3.99 (SD = 1.49), the average number of interactions is 14.9 (SD = 10) suggesting high variability in how often participants engaged. Helpful responses averaged at 6.67 times (mean = 6.67, SD = 5.653), while unhelpful responses averaged a bit more with 6.80 average (SD = 4.02). Neutral responses were less frequent, averaging 1.70 (SD = 2.020), with some variation across participants.

TABLE IV. DESCRIPTIVE STATISTICS FOR IN PERSONALITY

	Mean	Std. Deviation
Trust	3.9908	1.49869
Interaction	14.93	10.007
Helpful	6.67	5.653
Unhelpful	6.80	6.18
Neutral	1.70	2.020

a.* N = 30
** n = 6

Fig. 8 indicates that the participant with personality INFJ had extremely long conversations in both prompts with Cleverbot. However, despite the long conversations, we observed that there were quite unhelpful interactions between the participant and the chatbot. This aligns with Champagnes *Personal Style Inventory Typology* (1980), where personality INFJ work hard for their needs met, and have a desire to get things done when they want it, they are hard workers and put a lot of effort into their work. It seems as if this participant with INFJ wanted Cleverbot to respond to the prompts accordingly.

The participants with INTJ users yielded similar results, long conversations, unhelpful responses frequently, and very low trustworthiness towards the chatbot. According to Hogan and Champagne, the INTJ personality type is often stubborn, skeptical, and critical [24]. It must have been that these participants were more likely to be critical of Cleverbot. The participant with INTP views trust towards Cleverbot as very untrustworthy, as Hogan and Champagne describe the INTP personality as extremely logical [24], so if the chatbot was not being logical with the user, the participant must have viewed it as untrustworthy.

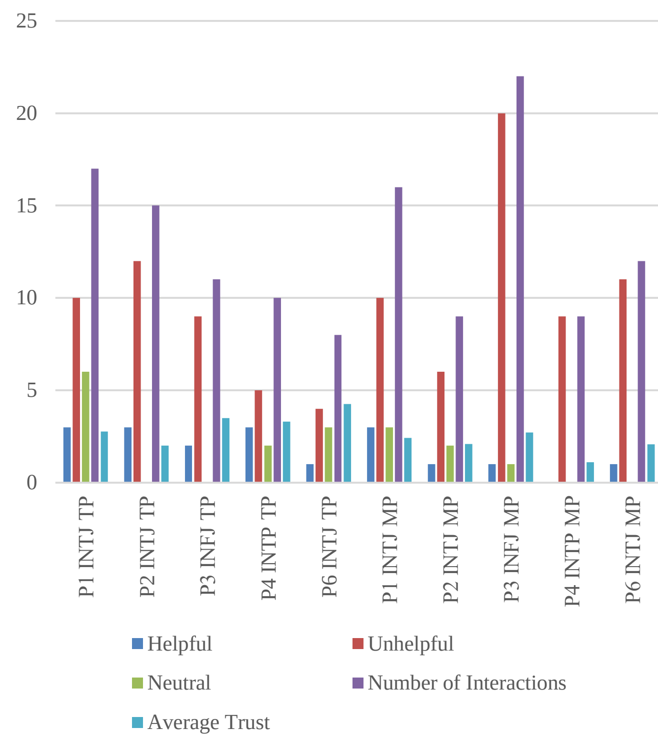

Fig. 8. Frequencies in cleverbot AI and IN personality.

Fig. 9 shows the interaction frequencies and average trust with Kuki and IN personality types. The participants with personality type INFJ had the highest interaction levels but also faced more unhelpful responses. Despite these interactions, their perceived trustworthiness towards the chatbot remained low, indicating the participant viewed this chatbot as untrustworthy. All participants with INTJ experienced high helpful responses from Kuki, and still gave moderate trust towards Kuki. INTJs are known to have great drive when it fits their own ideas and has meaningful interactions [24]. We noticed that all conversations between Kuki and all the participants with INTJ had more helpful interactions compared to Cleverbot. The participant with INTP personality had quite a few conversations and had an overall average low untrustworthy score towards the chatbot. The participant with INFJ had the most conversations with this chatbot, despite the chatbot having more unhelpful responses compared to the helpful responses in the interaction between Kuki and the INFJ participant. INFJs are known to go above and beyond, putting their best efforts into their work, therefore it may have been that this participant wanted to get the best answer from the chatbot over long conversations. [24].

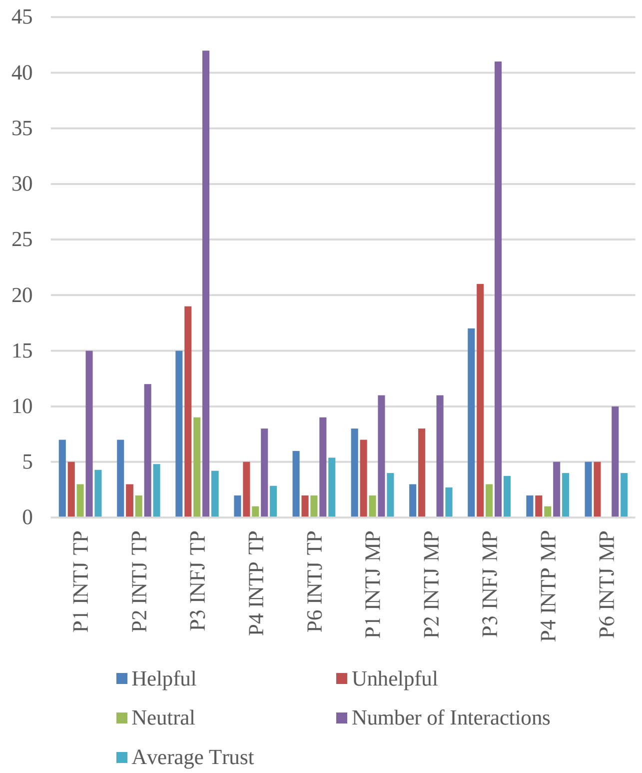

Fig. 9. Frequencies in kuki AI and IN personality.

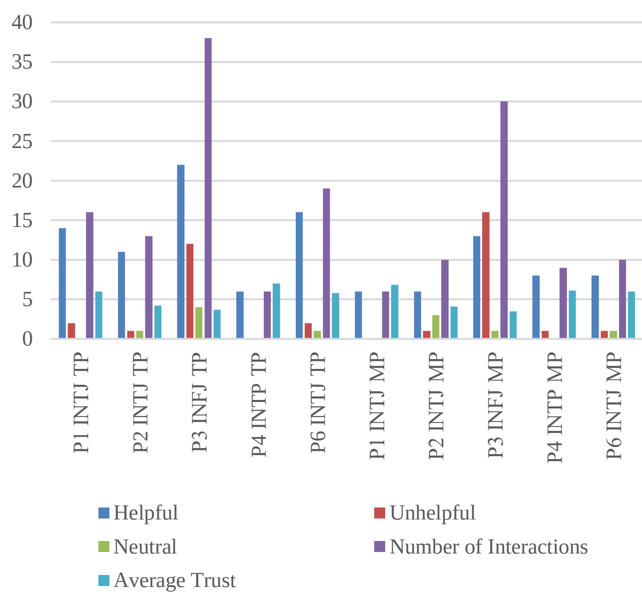

Fig. 10. Frequencies in replika AI and IN personality.

Fig. 10 displays the frequencies of interactions between Replika and users within the IN personality types. The participant with INFJ still had quite long conversations with Replika, as they did with Kuki and Cleverbot, relating to their personality once more where they work until they put their effort in [24]. All participants with INTJ had high trustworthiness towards the chatbot, and we observed that they all had more frequencies in the helpful category when we analyzed their interactions in both prompts. The participant with INTP had few conversations but deemed the chatbot to be very trustworthy based on their average trust score. We observed that Replika was more logical in their responses, and interacted well with all participants, leading to high average trust scores.

D. *The IS Personality and Frequencies*

Table V presents the descriptive statistics for the IS personality dimension and their code frequencies when interacting with all three chatbots. The mean trust across all chatbots is 3.85, with a standard deviation of 1.62, indicating a moderate level of trust across interactions. Participants engaged around an average of 14.12 interactions per conversation, with a standard deviation of 9.33, indicating a high variability in interactions. Helpful responses occurred most frequently, with a mean of 6.53 and a standard deviation of 5.38. Unhelpful responses were averaging around 4.60 with a standard deviation of 4.17. Neutral responses were least common, with a mean of 2.30, and a standard deviation of 2.76, indicating some variability across the interactions with all three chatbots. N represents the number of conversations analyzed, and n represents the total number of participants.

TABLE V. DESCRIPTIVE STATISTICS FOR IS PERSONALITY

	Mean	Std. Deviation
Trust	3.8549	1.61590
Interaction	14.12	9.325
Helpful	6.53	5.376
Unhelpful	4.60	4.172
Neutral	2.30	2.761

a.*N = 69
**n = 13

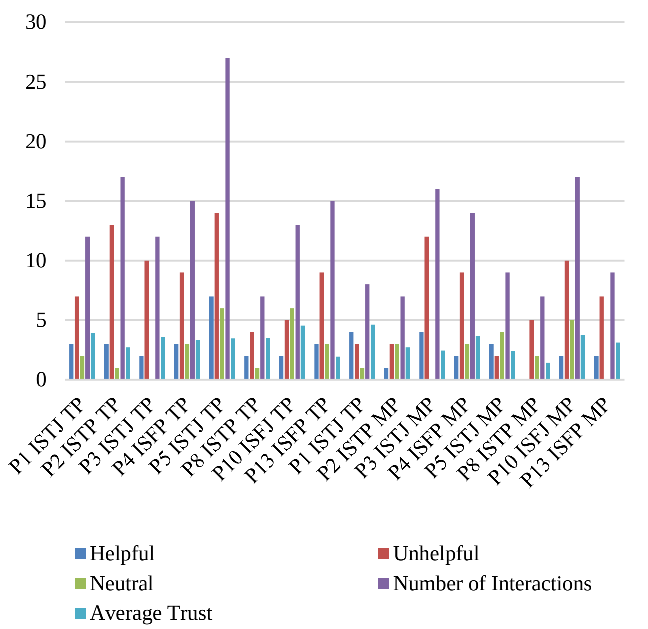

Fig. 11. Frequencies in cleverbot AI and IS personality.

Fig. 11 represents those participants with ISTJ personality had more helpful interactions within the meal prompt rather than the travel prompt, indicating their personality is to stay on task [39, 40]. The participants with ISTP had more unhelpful interactions with Cleverbot leading to a low average trust score within the travel prompt. We find this quite interesting, as according to Hogan and Champagne, people with ISTP personality types tend to not think necessarily more than they should, and that waste of energy is inefficient [24], we infer that the participants with the ISTP personality were trying to get the perfect answer from the chatbot to make up for the time they put into chatting with the chatbot.

Participants with ISFP had few long conversations, and this aligns with their personality, where Hogan and Champagne describe them as people who are often modest about their abilities, are not one to disagree, nor force their values or opinions on others [24]. Though they had long conversations, we analyzed that their conversations had more frequencies of unhelpful responses. Both participants with the ISFP personality still viewed the chatbot to be untrustworthy as indicated by their average trust score. Participants with the ISFJ personality yielded results that indicated they had more frequencies in the unhelpful code during their interaction with Cleverbot, aligning with Hogan and Champagne ISFJ personality type, where individuals work hard to get their obligations finished, and want to be accurate [24]. We can also infer that this participant with ISFJ wanted to get the right answers from the chatbot.

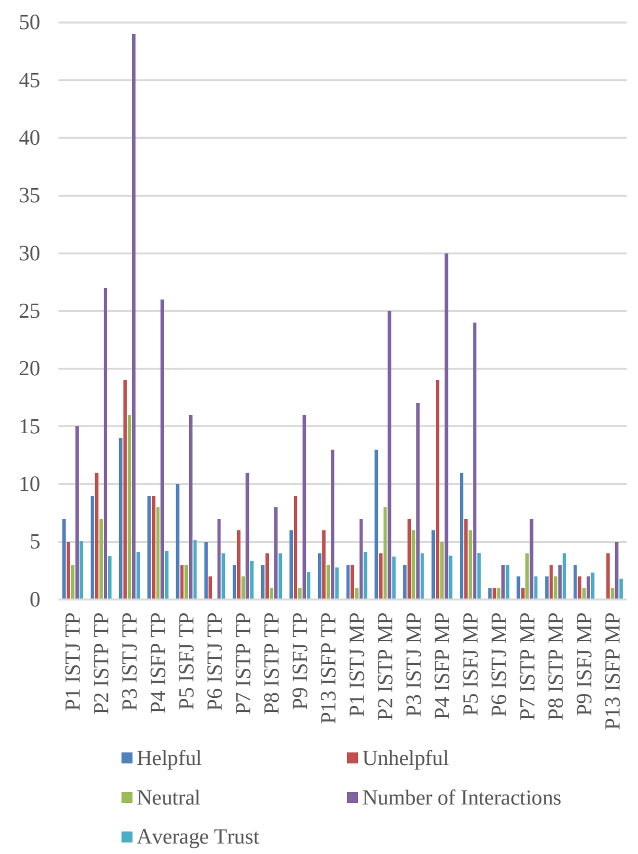

Fig. 12. Frequencies in kuki AI and IS personality.

Fig. 12 indicates that participants with personality ISTJ had perceived Kuki as moderately untrustworthy, despite the long conversations. ISTJs are known to be "Practical, orderly, matter of fact, logical, realistic and dependable [24]. As indicated in Fig. 12, Participant 3 with personality ISTJ had long conversations, but low trust towards the chatbot, while the other participant (Participant 6) with personality ISTJ had smaller conversations with Kuki but viewed the chatbot as moderately trustworthy, as indicated by the average trust score. These two participants had completely different views towards the chatbots, and experiences, indicating that every individual is different. Users with ISTP were flexible with their interactions with the chatbots, but overall, all participants with the ISTP personality had a low trustworthiness perception of Kuki. Participants with ISFP had long conversations in the travel prompts, and both viewed Kuki as very untrustworthy as indicated by their average trust score. However, in the meal prompt participant 13 with personality ISFP had a small interaction, with a very low average trust score, perceiving the chatbot as very untrustworthy. ISFPs are known to not want to waste time, so we can infer that Participant 13 did not want to continue the conversation [24]. When analyzing ISFJ participants, these participants had moderate untrustworthy perceptions towards the chatbot as indicated by their average trust scores in both prompts but do have the highest trust with Kuki when compared to the other three types of Introvert-Sensing types.

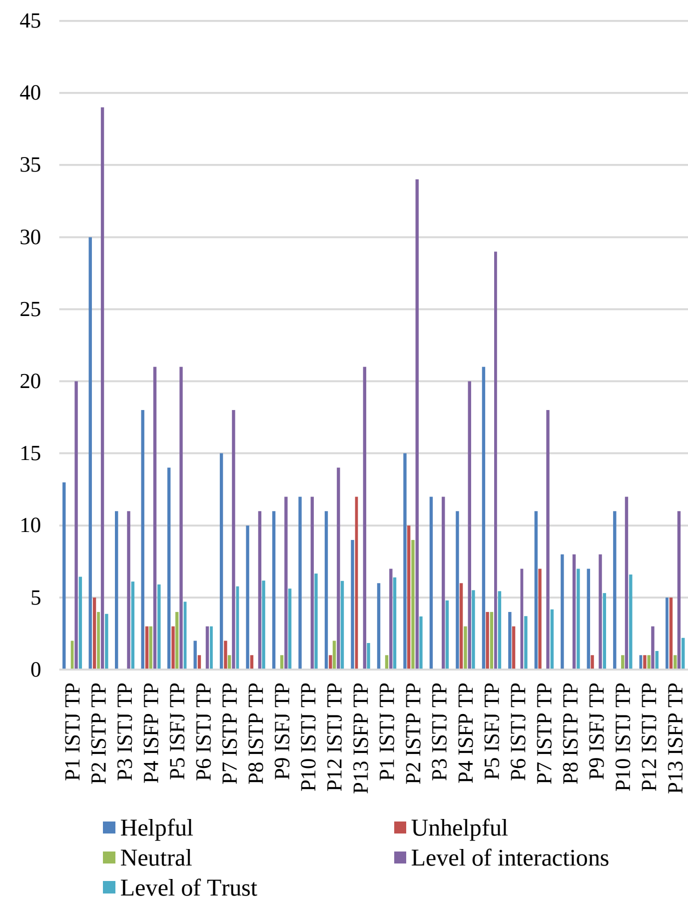

Fig. 13. Frequencies in kuki and the IS personality.

Fig. 13 indicates that all participants with ISTJ that had helpful frequencies, and long conversations tended to have a high average trust score in both prompts. However, we noticed that if the interaction was short, the participants with ISTJ continued to have low perceptions of trustworthiness towards Replika, as indicated in Participants 12 and 6. The participants with ISFJ had long conversations and a high score of average trust towards Replika. This might be due to their hard work and devotion, warm-heartedness, attention to detail, and need for accuracy, and how the Replika chatbot gave helpful responses, this might have led to high trustworthiness average scores [24]. All ISTP users had high trustworthiness towards the chat, indicating their personality of being flexible and easygoing, especially when the chatbot was helpful as we observed. The participants with ISFP all had moderately high trustworthiness towards Replika, and we observed that Replika did give helpful responses to all prompts when interacting with users with the Introvert-Sensing type.

V. Discussion and Limitations

Within the Intravenous and Intuition Dimension, participants with ENFJ tended to have lower trust scores in Cleverbot, and Kuki when it was not helpful and tended end to the conversation, however, one participant in the EN dimension with personality ENFP tended to give the chatbot higher scores when it was not helpful, rather, giving it a lower score when it was helpful. Within our coding, if the chatbot was being tangible (i.e. steering off the conversation to complete something different, such as “Who Cares”), or not engaging with the user, this would be considered unhelpful. Despite the chatbot being unhelpful, the participant with user ENFP continued to give it higher trust scores. It can be inferred that in this study, the participants with ENFP personalities wanted more interactions that engaging, and more exciting, as their personality tends to be open to opportunities and tends to be more imaginative [14].

ENFJs in this study were more likely to end the conversation if the conversation was not meaningful. This relates to their personality, as ENFJs tend to be more logical, organized, and worthy of leadership [32].

Within the analysis of ES (Extraverted and Sensing) personality, ESTP users' conversations with all chatbots were often brief, especially when the conversation led to unhelpful responses. In our study, ESFP participants, known for their energetic nature [33] tended to rate chatbots when it was helpful and engaging. ESTJ participants gave higher scores when the chatbot was helpful and gave clear answers, as this relates to their personality where they value logic, relativism, and directness nature [32, 33], tended to give lower trust scores to the chatbot despite when the chatbot was helpful and engaging. ESFJ on the other hand, gave moderate trust scores to the chatbots even when it was being unhelpful, this can relate to their personality as they are more warmhearted and sensitive to the environment around them, as seen more in women for their motherly nature and warmhearted figures [32, 34].

The Inverted and Intuition (IN) Dimension had mixed results. Participants with personality INFJ gave chatbots the lowest scores, despite being helpful or unhelpful. This relates to their personality from the dimension of the ‘Feeling” and “Judgement”, where they follow their own conviction and own path rather than objective fact [35].

Users with personality INTJ had more helpful interactions and reported higher trust scores, indicating that the more beneficial and engaging the chatbot is, the more users tend to trust it despite the level of interactions. This goes on with their personality as they are more rational when it comes to their environment and analytic to the future when it comes to decisions [36, 37].

Participants with personality INTP had mixed results with some chatbots. It had mixed results from helpful, unhelpful, and neutral interactions. While interaction levels were consistent, they didn’t necessarily lead to higher trust levels. Overall, the INTP participants did not have consistent results between the three chatbots. This relates to their personality as they are more likely to be spontaneous and have more intuition rather than sensing, and difficult to please, so if they have a bad experience, this leads to difficulty in getting high trust scores [37, 38].

The largest set of personalities in this simple was Introverted and Sensing (IS). In all three chatbots (Cleverbot, Kuki, and Replika) the participants with IS tended to still have longer conversations despite the possibility that the chatbot was not being helpful. Research has indicated that individuals with personality ISTJ tend to be nervous when it comes to high achievement, and have a need to accomplish a task [24, 39, 40], it can be inferred that participants in this study wanted to continue to conversate with the chatbot until they were able to get a good answer from them, despite having lower scores in trust and unhelpfulness. ISTP yielded different results, showing more trust in the chatbot despite being unhelpful, as their perception leads them to think the chatbot might have been worthy of a high trust rating, despite objective unhelpful frequencies in their conversations, feeling the way of their senses as well rather than being intuitive, allowing for flexibility in their way of perceiving the world [33, 37].

The participants with the ISFJ personality had similar results to ISTP surprisingly, where as seen in their conversations with Replika, they tended to have higher trust towards Replika, yet still showed moderate trust levels in Cleverbot and Kuki despite those two chatbots being unhelpful. ISFJs are known to be more warm-hearted, sincere, and approachable, having the desire to engage with the chatbot as long as emotional satisfaction is involved, their “Feeling” type indicates they still felt something for the chatbot despite not being helpful [24].

The participants with ISFP personality tended to have high trust scores in all three chatbots as compared to the other IS types, despite the categories of helpfulness, unhelpfulness, and neutral. Their personality of feeling and perceiving might have influenced their trust scores to be higher with the chatbots.

While each personality yielded different results, with a few cases of similar results, there are limitations to this study. Within the Dimension of Extraversion and Intuition, this personality was underrepresented in this study, and Introversion dominated this study. The sample size was taken from an Information Systems major program, where students can go on to work in the cybersecurity field, be engineers, or work in tech, and research has shown that students with introversion tend to likely be engineers, and work in the science field [41, 42]. Another limitation is that there is missing data from several users from the Introvert-Intuition (IN) Dimension, Introvert-Sensing (IS) Dimension, and Extrovert-Sensing (ES) dimension. However, we have successfully collected a large amount of systematic conversations from 29 participants. The results can be generalized to similar personality users to contribute to broader findings.

VI. Conclusion

We observed how personality dimensions influence trust and engagement towards chatbots across two different scenarios, a travel prompt and a meal prompt. We had the participants rate each interaction on a Likert scale rating from 1 to 7, with 7 being the highest trustworthy score, and 1 being the lowest trustworthy score. We created a hierarchical modeling scale to organize our data, and had participants go through three chatbots to explore variabilities in each chatbot. We found key findings, which included that the participants with ENFJ personalities rated Cleverbot low, in counterpart with our other participants ENFP who seemed to enjoy the conversation they had with Cleverbot, aligning with their imaginative and optimistic natures [24].

ESTP participants preferred direct, practical responses and gave lower trust scores when interactions were inefficient. Similarly, we observed that ESFP participants disengaged from unengaging chatbots but appreciated entertaining responses, while ESFJs maintained moderate trust levels, driven by their empathetic and harmony-focused personalities [24]. ESTJ participants rated trust highly when interactions were helpful but were critical of less engaging exchanges.

For Introverted-Intuitive personality types, INFJ participants engaged in long conversations but rated unhelpful interactions as untrustworthy, consistent with their diligent and goal-oriented nature. INTJs preferred meaningful interactions and gave moderate trust scores to Kuki due to its helpful responses, while INTP participants, being highly logical, rated chatbots as untrustworthy when responses lacked coherence or when the chatbot failed to deliver the information the participant wanted.

For our Introverted-Sensing types, ISTJ participants valued helpful interactions and rated them highly, though shorter or unhelpful exchanges led to lower trust. ISTPs, known for their practicality [24], also rated chatbots lower when responses were inefficient or overly lengthy. ISFP participants, modest and time-conscious, found chatbots untrustworthy if responses lacked engagement, while ISFJs maintained moderate trust levels, valuing accuracy and helpfulness, particularly with Replika.

Across all three chatbots, we observed that the Kuki chatbot was perceived as more trustworthy by most of the participants in this study, more with the Extroverted-Intuitive and Extroverted-Sensing types due to its helpfulness. Replika received the highest trust scores, particularly from ISFJ and ISTP users, due to its logical and engaging responses. These findings underscore the importance of tailoring chatbot interactions to align with user personality traits, enhancing trust and satisfaction by balancing logical coherence, directness, and emotional engagement.

The future study will recruit participants from all majors to increase the sample size in different personality dimensions except for the IS dimension in which we have enough data from technology-related major students. We will also continue to design chatbots focusing on personality following the design interpretations from this work. This work contributes to the fast development of user-centered CA design with frequencies of user responses and trustworthy ratings to sort their behaviors by patterns correlating to their personality dimensions. It focuses on user perceptions of trust in chatbot experience. It also presents detailed interpretations of diverse personality users' behavior patterns and trends to show how they interact with CAs. The work sheds light on design guidelines of user trust based on personality for better human-centered AI design.

References

[1] Y. K. Dwivedi, A. Sharma, N. P. Rana, M. Giannakis, P. Goel, and V. Dutot, "Evolution of artificial intelligence research in Technological Forecasting and Social Change: Research topics, trends, and future directions," Technological Forecasting and Social Change, vol. 192, Jul. 2023J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Y. Xu et al., "Artificial Intelligence: a Powerful Paradigm for Scientific Research," The Innovation, vol. 2, no. 4, Oct. 2021

[3] A. Turing, "Computing Machinery and Intelligence," Mind, vol. 59, no. 236, pp. 433–460, Oct. 1950.

[4] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955," AI Magazine, vol. 27, no. 4, pp. 12–12, Aug. 1955,

[5] Oliveir E., Gama J., Z. Vale, and H. L. Cardoso, Progress in artificial intelligence : 18th EPIA Conference on Artificial Intelligence, EPIA 2017, Porto, Portugal, September 5-8, 2017, Proceedings. Cham, Switzerland: Springer, 2017.

[6] E. Bauer et al., "Using natural language processing to support peer-feedback in the age of artificial intelligence: A cross-disciplinary framework and a research agenda," British Journal of Educational Technology, vol. 54, no. 5, May 2023.

[7] S. Pacheco-Mendoza, C. Guevara, A. Mayorga-Albán, and J. Fernández-Escobar, "Artificial Intelligence in Higher Education: A Predictive Model for Academic Performance," Education Sciences, vol. 13, no. 10, p. 990, Oct. 2023.

[8] N. Ameen, A. Tarhini, A. Reppel, and A. Anand, "Customer Experiences in the Age of Artificial Intelligence," Computers in Human Behavior, vol. 114, no. 106548, p. 106548, Jan. 2021.

[9] R. S. Sutton et al., "Reward-respecting subtasks for model-based reinforcement learning," Artificial Intelligence, vol. 324, p. 104001, Nov. 2023,

[10] A. Pradhan and A. Lazar, "Hey Google, Do You Have a Personality? Designing Personality and Personas for Conversational Agents," CUI 2021 - 3rd Conference on Conversational User Interfaces, Jul. 2021

[11] J. Wang, J. Q. Chen, D. Kang, S. Herath, and A. AbuHussein, "Designing a Conversational Agent for Education using a Personality-based Approach," International Journal of Advanced Computer Science and Applications, vol. 15, no. 6, 2024.

[12] F. Kaya, F. Aydin, A. Schepman, P. Rodway, O. Yetişensoy, and M. Demir Kaya, "The roles of personality traits, AI anxiety, and demographic factors in attitudes toward artificial intelligence," International Journal of Human–Computer Interaction, vol. 40, no. 2, pp. 497–514, Dec. 2022.

[13] J.-P. Stein, T. Messingschlager, T. Gnambs, F. Hutmacher, and M. Appel, "Attitudes towards AI: measurement and associations with personality," Scientific Reports, vol. 14, no. 1, Feb. 2024.

[14] A. Oksanen, N. Savela, R. Latikka, and A. Koivula, “Trust Toward Robots and Artificial Intelligence: An Experimental Approach to Human–Technology Interactions Online,” Frontiers in Psychology, vol. 11, Dec. 2020.

[15] A. H. Maslow, “A Theory of Human Motivation,” Psychological Review, vol. 50, no. 4, pp. 370–396, 1943.

[16] E. Sullivan, “Self-actualization,” Encyclopædia Britannica. 2019

[17] M. X. Zhou, G. Mark, J. Li, and H. Yang, “Trusting Virtual Agents,” ACM Transactions on Interactive Intelligent Systems, vol. 9, no. 2–3, pp. 1–36, Apr. 2019.

[18] H. Li, “Rethinking human excellence in the AI age: The relationship between intellectual humility and attitudes toward ChatGPT,” Personality and Individual Differences, vol. 215, p. 112401, Dec. 2023.

[19] R. Riedl, “Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions,” Electronic Markets, vol. 32, Nov. 2022.

[20] N. N. Sharan and D. M. Romano, “The effects of personality and locus of control on trust in humans versus artificial intelligence,” Heliyon, vol. 6, no. 8, p. e04572, Aug. 2020.

[21] A. D. Kaplan, T. T. Kessler, J. C. Brill, and P. A. Hancock, “Trust in Artificial Intelligence: Meta-Analytic Findings,” Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 65, no. 2, May 2021.

[22] C. Montag, B. Becker, and B. J. Li, “On Trust in Humans and Trust in Artificial intelligence: a Study with Samples from Singapore and Germany Extending recent research,” Computers in human behavior. Artificial humans, vol. 2, no. 2, pp. 100070–100070, May 2024.

[23] S. Rushton, J. Morgan, and M. Richard, “Teacher’s Myers-Briggs personality profiles: Identifying effective teacher personality traits,” Teaching and Teacher Education, vol. 23, no. 4, pp. 432–441, May 2007.

[24] R. Hogan and D. Champagne, “University Associates The 1980 Annual Handbook for Group Facilitators 1,” 1980.

[25] J. H. Jung, Y. Lee, and R. Karsten, “The Moderating Effect of Extraversion–Introversion Differences on Group Idea Generation Performance,” Small Group Research, vol. 43, no. 1, pp. 30–49, Sep. 2011.

[26] A. Naveed, “How College Students’ MBTI Personality Types Relate to Their Levels of Emotional Intelligence,” Corpus Journal of Social Sciences & Management Review, vol. 2, no. 01, pp. 65–70, Mar. 2024.

[27] J. Langan-Fox and D. A. Shirley, “The Nature and Measurement of Intuition: Cognitive and Behavioral Interests, Personality, and Experiences,” Creativity Research Journal, vol. 15, no. 2, pp. 207–222, Jul. 2003.

[28] J. Wang, M. Schmidt, R. Kuikel. “User trust and credibility of conversational agents,” In Procs of 23rd Annual Security Conference, Las Vegas, NV, U.S.A, May 2024.

[29] J. Wang, J. Dunkley, M. Hamal, V. Raut, and S. Herath, “Designing Conversational Agents for Education: A Preliminary Study of User Personality’s Impact on Design,” The International Journal of Engineering and Science (IJES), vol. 12, no. 2, 2023.

[30] H. Dubberly and P. Pangaro, “What is conversation, and how can we design for it?,” interactions, vol. 16, no. 4, p. 22, Jul. 2009.

[31] R. Zhang, X. Liang, and S.-H. Wu, “When chatbots fail: exploring user coping following a chatbots-induced service failure,” Information technology & people, vol. 37, no. 8, pp. 175–195, Jul. 2024.

[32] N. N. Vitó Ferreira and J. J. Langerman, “The correlation between personality type and individual performance on an ICT Project,” IEEE Xplore, Aug. 01, 2014.

[33] H. Gordan, “Myers Briggs Type Indicator Personality Characteristics of Beginning Trade and Industrial and Health Occupations Education Secondary Teachers,” Journal of Health Occupations Education, vol. 14, no. 1, 2000.

[34] S. W. Johnson, M. S. Gill, C. Grenier, and J. Taboada, “A Descriptive Analysis of Personality and Gender at the Louisiana State University School of Veterinary Medicine,” Journal of Veterinary Medical Education, vol. 36, no. 3, pp. 284–290, Sep. 2009.

[35] A. Naveed, “View of How College Students’ MBTI Personality Types Relate to Their Levels of Emotional Intelligence,” Corpus Journal of Social Sciences & Management Review, vol. 2, no. 1, 2024.

[36] O. Y. M. Sardjono, “Learning Outcomes Based on Myer-BriggsType Indictaor (MBTI) Personality from Accounting Department Students Sam Ratulangi University,” Accountability , vol. 12, no. 1, 2023.

[37] J. Jankowski, The 16 Personality Types in a Nutshell. 2016.

[38] J. Simkus, “Logician (INTP) Personality Type,” Simple Psychology, 2024.

[39] R. Diab-Bahman, “The Impact of Dominant Personality Traits on Team Roles,” The Open Psychology Journal, vol. 14, no. 1, pp. 33–45, Feb. 2021.

[40] I. Briggs Myers, Introduction to Type®. CPP, 1998.

[41] M. Pollock, “The relationship between personality type and choice of college major,” 2016.

[42] A. Vedel, “Big Five personality group differences across academic majors: A systematic review,” Personality and Individual Differences, vol. 92, no. 92, pp. 1–10, Apr. 2016.