

MASTER THESIS

MSc Management of Technology

Kevin Doan (4374053)

The Impact of Vocal Communication
and its Personalization on Intention to Use
of Chatbots Using Behavioral Activation
to Support Patients Experiencing Depression



Thesis committee:
1st Supervisor | Dr. Nadia Metoui
2nd Supervisor | Dr. Geescke Scholz
Chair | Dr. Aaron Ding

The Impact of Vocal Communication and its Personalization on Intention to Use of Chatbots Using Behavioral Activation to Support Patients Experiencing Depression

Thesis report

by

Kevin Quang Doan

to obtain the degree of Master of Science
at the Delft University of Technology
to be defended publicly on August 18, 2023 at 10:00

Thesis committee:

Chairman:	Dr. Y. Aaron Ding
First supervisor:	Dr. Nadia Metoui
Second supervisor:	Dr. Geeske Scholz
Faculty:	Technology, Policy and Management (TPM)
Student number:	4374053
Project Duration:	October 25, 2022 - August 18, 2023

An electronic version of this thesis is available at
<https://repository.tudelft.nl/>.

*you can only grow
when you admit to failure*

Preface

My journey at TU Delft started after I made the huge decision to switch from an academic direction. Initially, I obtained academic knowledge in Biomedical Engineering (2014-2018) and Technical Medicine (2017-2021) on the undergraduate level at the University of Twente. Not only was I interested in the inner workings of the human body, but I was passionate to save people's lives using innovation. Although it seemed like the ideal career path, soon I realized that I could not take all the emotional hardships that come with helping people in the clinical setting. Mistakes are guaranteed to happen in one's clinical career, regardless of how experienced one is, and that was a burden I did not want to carry. Despite my genuine interest in medicine and engineering, I had always felt the urge to explore business and (digital) technology on a deeper level. That is how I found the Master's in Management of Technology (MOT) at TU Delft in 2020. Although I might not be able to help people physically in a professional setting, I still could contribute to the impact of technology on society as a whole.

It caught my attention from the personal experiences of my social circle that TU Delft is famous for its prestige, high-quality scientific contribution, and quality of education. For many first-year students, this translated to high expectations of academic performance that needed to be met. When another university grades you with an 8.0, TU Delft would only give at most a 7.0 grade for the same energy put in your work. At least, that is how I felt when I finished my first semester at TU Delft. On top of that, it was already hard enough to find my way of studying in the field of social science after thinking as an medically trained engineer for 6 years. It was no surprise that occasionally I questioned why I ever put myself in this situation. When thinking about the pressure put on me, sometimes this pressure actually reached my eyes physically, to the point that my lacrimal glands started to secrete. It were these moments I was extra grateful for the trees that died for me so that I could cover up my tears. Although not meant cynically, it was nice to know that I was not alone. Classmates actually joked around with the "Tudelft.nl/uitschrijven" sticker on WhatsApp from time to time to point out to the fictitious page to quite, which ironically gave me ground to find light in the darkness. It actually triggered a series of mindsets that I have obtained throughout the years in my spare time. From my experience in entrepreneurship, I learned to say "YOLO" and "work hard, play hard". In the gym I was taught to say "fuck it, if you want it, you can do it" or "no pain, no gain". My academic experience in Twente also taught me to prepare just one course at a time for the exam when things get desperate, because passing one course is better than risking to have multiple failures. And even if I failed, you can only grow when you admit to failure. When all these motivational nuggets failed to serve me, there was no better story of redemption from the best friend of my father, who almost left TU Delft but got convinced to stay by another friend at an elevator. Now he keeps talking about buying houses or something - seems like a good life to me.

In the end, I am grateful that I stayed at TU Delft until the end and I am happy with the wisdom that I was only able to learn from studying at TU Delft, e.g. critical thinking and the persistence to keep asking why until the root cause surfaces. In fact, it was this curiosity that led me to simultaneously finish another Master's degree, but in Business Information Management (BIM) at RSM, Erasmus University (cum laude, 2021-2022). While it was at MOT where I developed an interest in human psychology and its influences on innovation, it was BIM where I explored its applications by combining digital nudging and healthcare technology. I was so passionate about its impact on society that it became my inspiration for my thesis topic at TU Delft: helping people who experience depression through innovation.

I want to express my gratitude to Dr. Nadia Metoui, my first supervisor, for guiding me in the process of finding a research topic that was not only interesting but also purposeful. Though I had experience with a master thesis before, Nadia always challenged me and provided valuable feedback to bring out the best in me. I would like to express my gratitude to my second supervisor, Dr. Geeske Scholz, for being there for me whenever I needed her for general feedback and for providing valuable insights on psychology in trustworthiness. Dr. Aaron Ding, as the chair, offered great feedback on the bigger picture of the study. Other TPM staff members at TU Delft were also of great help. I want to thank PhD candidate Wiebke Toussaint Hutiri for sharing her experience with designing chatbots and voice personalization. I also want to express my gratitude to Dr. Marcus Westberg for providing feedback on my survey design. Finally, I would like to give a shout out to my girlfriend, family, and friends for the unwavering support. Hopefully, all this has contributed to a thesis that is enjoyable to read. As I approach the end of my journey at TU Delft, there's only one thing on my mind:

[Tudelft.nl/uitschrijven](https://tudelft.nl/uitschrijven)

Kevin Quang Doan
Almere, July 2023

Locations - where I have been to work on this thesis

Almere, Noorderplassen Noord

Amsterdam, De Nederlandsche Bank

Amsterdam, Openbare Bibliotheek Amsterdam

Delft, Delft University of Technology

Dordrecht, Sterrenburg

Lelystad, Ziekenhuis St Jansdal

Rotterdam, Unilever

Executive Summary

In the 21st century, the global prevalence of mental health problems has significantly increased, affecting nearly a billion people. These conditions not only diminish the quality of life for individuals but also contribute to stigmatization, discrimination, and social isolation. Mental health problems can also negatively impact self-esteem, relationships, and work or educational performance. The COVID-19 pandemic has exacerbated these issues, with depression and anxiety becoming the most common mental disorders among young workers. Even after the pandemic, depression cases remain high, highlighting the need for improved access to mental health support for young adults. However, encouraging young adults to seek help for mental health is challenging.

To address this issue, digitalizing psychotherapies by means of mental health chatbot services could be a potential solution. To increase adoption, it might be effective to give the chatbot a voice to communicate, since other studies have suggested that it could increase perceived trust, and in return increase intention to use. As a result, in this study the following research question is answered following Wieringa's design science methodology adapted for knowledge problems: "*How does vocal communication and its personalization in chatbots affect the intention to use of young adults with depression?*". To answer the research question, the effects of vocal communication and its personalization in chatbots were tested on the two dimensions of trust: cognitive trust (trust based on rationale) and affective trust (trust based on emotions). Then the relationship between the two dimensions of trust and intention to use were measured. The study also analyzed the mediating role of trust, examined the influence of voice attributes (gender of voice and speech rate) on intention to use, and explored effects of demographic characteristics on intention to use.

This lab experimental post-test study was conducted on the online survey platform Qualtrics. Participants were randomly allocated to the control group (text-based communication), or one of the two treatment groups (voice-based communication or voice-based communication with personalization option), after which participants were redirected to a chatbot to interact according to their assigned study condition. More specifically, participants allocated to the first mentioned treatment group were able to interact with a chatbot that could communicate to the user using speech, while participants allocated to the second mentioned treatment group were able to select a voice for the chatbot before interaction. The chatbot itself followed a rigid script based on the existing psychotherapy protocol called behavioral activation. As data was collected only once and results from participants were compared within the study, it was a cross-sectional between-subjects study, with relatively high internal validity. The unit of analysis was individuals experiencing depression, that did not have hearing difficulties and were between the age of 18 and 30. A total of 601 participants joined the study by sourcing them from Prolific. As for data analysis, mostly the Mann-Whitney U test, and ordinal logistic regression were used. However, for randomization check, one-way ANOVA was used.

Findings from this study suggested that chatbots communicating through speech did not significantly influence affective and cognitive trust compared to text-based chatbots. It was speculated that participants in this study may not have had enough time to build trust with the chatbots using vocal communication. Furthermore, the nature of being depressed may also make it harder to increase affective trust. Personalizing the voice of the chatbot did not help increase trust based on emotions. In fact, personalization negatively affected cognitive trust, possibly because the chatbot was unable to respond in a personalized manner. Factors like prior experience with advanced chatbots like ChatGPT and gender bias may have influenced participants' expectations of chatbot performance and could raise ethical concerns, although more research is required to confirm. Also, The study revealed that both cognitive and affective trust were positively associated with the intention to use chatbots. In fact, cognitive trust indirectly mediated the relationship between personalized vocal communication and intention to use, although there might be other variables mediating this relationship as well. Finally, exploratory analysis indicated that users from developing countries were associated with higher intention to use of chatbots with vocal communication, while those from developed countries displayed an association of lower cognitive trust. For personalized vocal communication, users from developed countries were associated with lower cognitive, affective trust and intention to use. To understand the underlying mechanisms better, further research is required.

This study made significant contributions to existing literature in various research areas. It expanded knowledge about the adoption of conversational agents for mental health support, particularly focusing on increasing adoption among young adults who experience depression. The study explored the use of speech as a mode of communication in chatbots, hypothesizing that it could make it more emotionally and rationally comfortable for young adults to use. In fact, using voice to increase anthropomorphism levels in order to increase trust in chatbots was never studied before, which this study was able to provide more insight about this relationship. Likewise, providing personaliza-

tion options for the voice and its relationship with trust and chatbot adoption was also never tested within the context of supporting mental health chatbots. Finally, it provided insights into the adoption of chatbots for mental healthcare in developing countries, an area that had received limited research attention despite the potential lack of access to healthcare in these regions. The study also addressed the ethical dimension of gender-based biases that may arise from selecting gender-specific voices for chatbots, contributing to the knowledge around ethics in chatbot technology.

In terms of managerial and societal implications, the study's findings have relevance for companies developing chatbots for healthcare applications. The insights gained from the study can help improve the design of chatbots for mental healthcare like depression, potentially facilitating their adoption and relieving the pressure on mental health professionals. Furthermore, the introduction of mental health chatbots to the market could contribute to long-term societal well-being and potentially alleviate economic damages due to depression. Additionally, the study highlights the potential benefits of chatbot accessibility for users in developing countries who typically have limited access to healthcare. The study's findings can also guide the development of ethical and safe policies by governments and health authorities, especially when integrating chatbots with artificial intelligence and gender-specific voices.

Contents

List of Figures	viii
List of Tables	x
1 Introduction	1
1.1 Research Objective	2
1.2 Research Questions	3
1.3 Research Scope	3
1.4 Relevance of the Study	4
1.5 Thesis Outline	4
2 Literature Review	5
2.1 Methodology of Literature Review	5
2.2 Behavioral Activation	8
2.3 Conversational Agents for Depression	8
2.4 Trustworthiness & Intention to Use	9
2.5 Vocal Communication	10
2.6 Personalization of Vocal Communication	11
2.7 Conceptual Framework	12
3 Research Method	15
3.1 Research Approach	15
3.2 Experimental Design	17
3.2.1 Overview Experimental Design	17
3.2.2 Survey Design	19
3.2.3 Data collection	21
3.2.4 Data Analysis Methods	22
3.2.5 Data Analysis Strategy	24
4 Preliminary Analysis	28
4.1 Randomization Check	28
4.2 Attention Check	29
4.3 Validity Analysis	29
4.4 Reliability Analysis	29
5 Results	31
5.1 Data exclusion	31
5.2 Participant characteristics	31
5.3 Findings	34
6 Discussion	46
6.1 Reflections on the Research Questions	46
6.2 Reflections of Implications	49
6.2.1 Theoretical Implications	49
6.2.2 Managerial & Societal Implications	50
6.2.3 Ethical Implications	50
6.3 Limitations & Future Research	51
7 Conclusion	53
7.1 Key Takeaways	53
7.2 Recommendations	53
References	55
A The script of all the chatbots	64
B Questionnaires	65

C	Survey Design	66
D	Shapiro-Wilk test results	79
E	Spearman's rank correlation coefficient results	81
F	Demographics Characteristics per treatment/control group	83
G	Descriptive statistics of measurement scales	85
H	R-Code	85

List of Figures

1	The rising number of depression cases among young adults.	1
2	Decomposition of the main research question.	3
3	The scope of this research is delineated to the conjoint areas of cognitive science and the study of technology.	4
4	Literature review rounds.	5
5	PRISMA 2020 diagram: a summary of all the literature review rounds.	7
6	The conceptual framework that was derived from combining the knowledge gaps identified from the literature. Hypothesis are also displayed between the constructs.	13
7	Design science methodology adapted by Wieringa (2014). This particular framework is suitable for knowledge problems in information systems.	15
8	The research flow framework of this study. Each phase of the design science methodology produced an output that could be found in the designated chapter.	16
9	The between-subjects design used to collect data for sub-research question 1 and 3.	17
10	Factorial survey design for a 2 (gender: male, female) x 2 (speed: slow, fast) to address the influence of vocal attributes.	18
11	The between-subjects design used to collect data for sub-research question 2 and 3.	18
12	The overview of the entire research design.	20
13	The adjusted formula for Cohen's D, using Mann-Whitney U Test (U) and participant group size N1 and N2.	23
14	Comparison method 1: Data was compared between two groups (A and B) to observe the effect on the dependent variable (X).	24
15	Comparison method 2: Data was first stratified based on a socio-demographic characteristic (Z), after which data was compared between two groups (A and B) to observe the effect on the dependent variable (X).	25
16	Comparison method 3: Data from a research group (A) was first separated into two subgroups based on socio-demographic characteristics Z1 and Z2. Then, these subgroups were compared with each other to observe the effect on the dependent variable (X).	25
17	Comparison method 4: Data from a research group (A) was initially divided into four subgroups, each assigned a distinct combination of independent variables (a) and (b). The unique combinations available are a1b1, a1b2, a2b1, and a2b2. Subsequently, the collected data of the dependent variables (X) were compared with each other to explore the presence of an interaction effect.	26
18	Overview of absolute participant measurements for the control group.	33
19	Overview of absolute participant measurements for Treatment Group I.	33
20	Overview of absolute participant measurements for Treatment Group II.	33
21	The interaction plot between gender of voice and speech rate on cognitive trust.	35
22	The interaction plot between gender of voice and speech rate on affective trust.	37
23	The association of personalized vocal communication on cognitive trust.	38
24	The relationship between personalized vocal communication, cognitive trust and intention to use, in which cognitive trust mediates indirectly the effect of personalized vocal communication on intention to use.	41
25	The start of study, in which informed consent is asked to the participant. If not consent is given, the study will end immediately for the participant.	66
26	Since Prolific was used to find participants, their Prolific ID was asked in order to compensate the participant.	67
27	To align with Prolific's pre-screening requirements, these are also asked in this section to make sure only eligible participant can enter the study. If not eligible, the study will end immediately for the participant.	67
28	The study starts by introducing the participant with a fictitious situation. In addition, participant was asked to go to a silent area. If the participant was not able to be in a silent area, the study will end immediately for the participant.	68
29	This is the screen for Treatment Group II, in which the participant is able to listen and choose the preferred voice for the chatbot to have. The Control Group and Treatment Group I would not see this screen.	69
30	This is the screen in which all participants will be directed to the Botstar chatbot. Depending on the allocated treatment/control group (and if applicable, depending on the chosen voice), the participant is will be directed to the chatbot accordingly.	70

31	This is how the chatbot would look like for Treatment Group I and II, in which voice messages are used to communicate with the participant. The control group will see text messages instead of voice messages.	71
32	After interacting with the chatbot, the participant is made aware of the questionnaires that would come next.	71
33	Here participants are asked to answer the questions of the anthropomorphism scale.	72
34	Here participants are asked to answer the questions of the benevolence scale.	73
35	Here participants are asked to answer the questions of the ability scale.	74
36	Here participants are asked to answer the questions of the predictability scale. In addition, an attention check is applied to check whether the participant had paid attention to the messages of the chatbot.	75
37	Here participants are asked to answer the questions of the overall trust scale.	76
38	Here participants are asked to answer the questions of the intention to use scale.	76
39	Here participants are asked to answer the questions of the voice communication preference scale.	77
40	Here participants are asked about their demographic information.	78
41	Here participants are free to provide feedback/comment about their experience with the survey and chatbot.	79
42	The end of the study.	79

List of Tables

1	Search strings: First literature review.	5
2	Search strings: Second literature review.	6
3	Search strings: Third literature review.	6
4	Search strings: Fourth literature review.	6
5	Randomization check based on socio-demographics characteristics.	28
6	Randomization check for factorial design within Treatment Group II based on socio-demographics characteristics.	28
7	Descriptive statistics of socio-demographics.	29
8	Internal validity per measurement scale.	30
9	Socio-demographics characteristics of all the participants.	31
10	Frequency Table Countries of Residence of all the participants.	32
11	The influence of vocal communication on ability and predictability. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test. . . .	34
12	The influence of vocal communication on cognitive trust. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test.	34
13	The relationship between the voice attributes and cognitive trust. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test. . . .	35
14	Interaction effect of gender of voice and speech rate on cognitive trust. The results were derived by using data from Treatment Group I using ordinal logistic regression.	35
15	The relationship between vocal communication and cognitive trust, in which participants are stratified into country of residence or gender. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test.	36
16	The influence of vocal communication on affective trust. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test.	36
17	The relationship between the voice attributes and affective trust. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test. . . .	36
18	Interaction effect of gender of voice and speech rate on affective trust. The results were derived by using data from Treatment Group I using ordinal logistic regression.	37
19	The relationship between vocal communication and affective trust, in which participants are stratified into country of residence or gender. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test.	37
20	The influence of personalized vocal communication on ability and predictability. The results were derived by comparing data from the control group and Treatment Group II using Mann-Whitney U Test.	38
21	The influence of personalized vocal communication on cognitive trust. The results were derived by comparing data from the control group and Treatment Group II using Mann-Whitney U Test. . . .	38
22	The relationship between personalized vocal communication and cognitive trust, in which participants are stratified into country of residence or gender. The results were derived by comparing data from the control group and Treatment Group II using Mann-Whitney U Test.	39
23	The influence of personalized vocal communication on affective trust. The results were derived by comparing data from the control group and Treatment Group II using Mann-Whitney U Test. . . .	39
24	The relationship between personalized vocal communication and affective trust, in which participants are stratified into country of residence or gender. The results were derived by comparing data from the control group and Treatment Group II using Mann-Whitney U Test.	39
25	Mediation Analysis: Overview of the individual associations of voice or personalized voice (IV), cognitive trust (M) and intention to use (DV).	40
26	Mediation Analysis: Overview of the individual associations of voice or personalized voice (IV), affective trust (M) and intention to use (DV).	41
27	The influence of (personalized) vocal communication on overall trust, intention to use, anthropomorphism and vocal communication preference. The results were derived by comparing data from the control group and Treatment Group I, and derived by comparing data from the control group and Treatment Group II. Here, the Mann-Whitney U Test was used.	42
28	The difference between vocal communication and personalized vocal communication in chatbots, in terms of dimensions of trust, anthropomorphism, overall trust, intention to use and vocal communication preference.	43

29	The relationship between (personalized) vocal communication and intention to use, in which participants are stratified into country of residence or gender. The results were derived by comparing data from the control group and Treatment Group I, and derived by comparing data from the control group and Treatment Group II. The Mann-Whitney U Test was used.	43
30	The influence of the difference in country classification and gender on affective trust and cognitive trust within a research group.	44
31	The influence of the difference in country classification and gender on intention to use within a research group.	44
32	Shapiro-Wilk test on the Anthropomorphism scale.	79
33	Shapiro-Wilk test on the Benevolence scale.	80
34	Shapiro-Wilk test on the Ability scale.	80
35	Shapiro-Wilk test on the Predictability scale.	80
36	Shapiro-Wilk test on the Overall Trust scale.	80
37	Shapiro-Wilk test on the Intention to Use scale.	80
38	Shapiro-Wilk test on the Voice Preference scale.	81
40	Spearman's rank correlation coefficients of the Ability scale.	81
41	Spearman's rank correlation coefficients of the Predictability scale.	81
42	Spearman's rank correlation coefficients of the Overall Trust scale.	81
39	Spearman's rank correlation coefficients of the Benevolence scale.	82
43	Spearman's rank correlation coefficients of the Intention to Use scale.	83
44	Spearman's rank correlation coefficients of the Anthropomorphism scale.	83
45	Spearman's rank correlation coefficients of the speech delivery Preference scale.	83
46	Demographics characteristics - Chatbot communicates using text messages.	83
47	Demographics characteristics - Chatbot communicates using voice messages.	84
48	Demographics characteristics - Chatbot communicates using voice messages, in which voice attributes could be personalized.	84
49	Descriptive statistics of measurement scales.	85

1 Introduction

Since the start of the 21st century, there has been an increase in cases of mental health problems (Bor et al., 2014, Storrie et al., 2010), to the point that close to a billion people worldwide have experienced some form of such condition (The Lancet Global Health, 2020, WHO, 2022). It has been shown that these individuals have a lower quality of life (Rathnayaka et al., 2022). Not only are mental health problems associated with stigmas that could lead to discrimination, inclusion issues, and social isolation, but they could also negatively affect self-esteem, relationships, and the performance of work or education-related activities (Storrie et al., 2010). The latter could be clearly observed during the COVID-19 pandemic, which has resulted in an estimation of a total loss of \$1.6 trillion in terms of economic damages (Cutler and Summers, 2020, Hossain et al., 2020). Thus, reducing the prevalence of depression is something to be taken seriously.

As can be seen in Figure 1, the cases for depression have been increasing for young adults compared to the older age groups since the start of the 21st century (Twenge et al., 2018). The COVID-19 pandemic has also shown that depression and anxiety became the most common mental disorders among young workers (Campbell and Gavett, 2021, Hawes et al., 2022, Rathnayaka et al., 2022). In fact, compared to young adults who already had symptoms of depression before the pandemic, young adults without pre-existing depression had more severe symptoms of depression during the pandemic (Kleine et al., 2023). Even after the pandemic, depression cases remain high (Wang et al., 2022). Therefore, young adults should gain better access to mental health support.

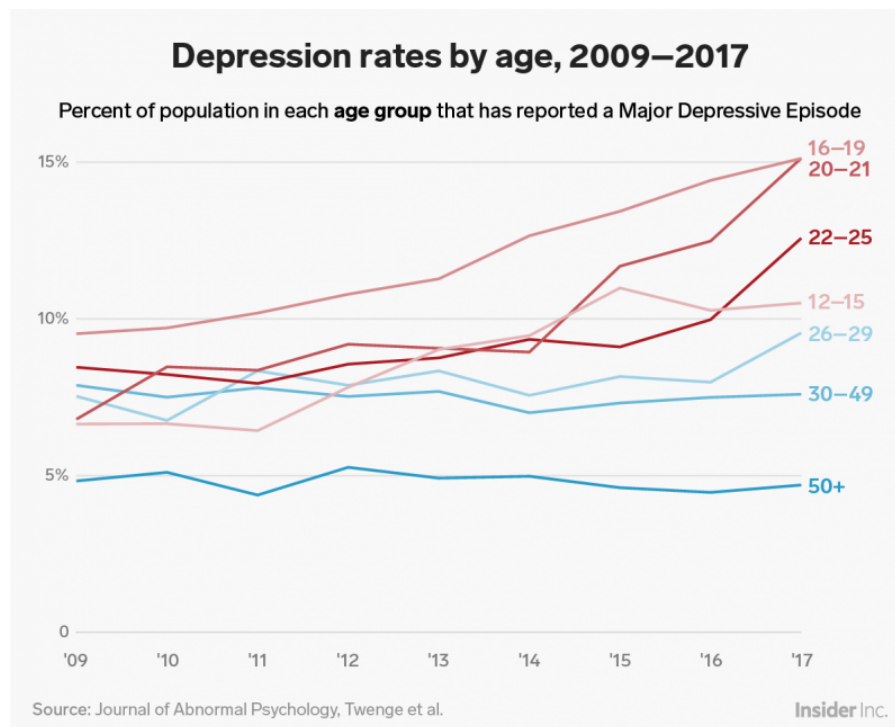


Figure 1: The rising number of depression cases among young adults.

However, it is a challenge to encourage young adults to seek help for mental health since they are less likely to do so (Bae Brandtzæg et al., 2021). One of the reasons is that young adults tend to believe they can sort out their own mental health problems, while in contrast, they are more willing to seek help when they feel comfortable enough to trust the potential help provider (Rickwood et al., 2007). Furthermore, it has been found that young adults face difficulties in gaining access to online social support, which discourages them from seeking mental help (Høiland et al., 2020). A general increase in the prevalence of depression comes with a rising demand for mental health professionals (Beck et al., 2018). However, the demand is outpacing supply, making it more challenging to provide support for individuals experiencing mental issues like depression (Rathnayaka et al., 2022).

When taking into account the problem of accessibility and the fact that younger people tend to adopt technology more easily (Hanson, 2010), one possible solution to encourage young adults to seek mental help is to digitalize psychotherapies. This service could be provided in the form of a mental health chatbot, or in other words, a chatbot that is able to provide mental health services.

1.1 Research Objective

The use of conversational agents (CA) seems promising as a respond to the rising demand for mental health professionals to treat depression and to make it easier for young adults to seek support for mental health. In fact, surveys have shown that people would be interested in using chatbots for mental health services if available (Dosovitsky and Bunge, 2022, Nadarzynski et al., 2019) and that people are more willing to disclose more information to a virtual agent than with a human being (Lucas et al., 2014). However, people tend to be skeptical of technology that is integrated with AI (Hengstler et al., 2016). Also, since chatbots usually are not able to show empathy, it is more likely that users do not feel heard, which could compromise engagement (Nadarzynski et al., 2019). Thus, there are factors that could make intention to use these chatbots more difficult (Müller et al., 2019).

It has been suggested that one of the factors contributing to adoption of CAs is perceived trust (Følstad et al., 2018). Since CAs are not humans, people tend to have lower perceived trust due to the lack of human-likeness (Følstad et al., 2018). Usually, with the perception of lower trust, it increases the belief that the product might fail or not work as intended (Nienaber and Schewe, 2014). As such, when applying these CAs in the context of healthcare, it is not surprising that people act more cautiously since it concerns one's health.

There are many factors that could influence one's perceived humanness (or anthropomorphism), and one of them is vocal communication (Schroeder and Epley, 2016, Wienrich et al., 2022). Recently, it has been shown that the use of AI to detect depression in the user's voice is becoming a reality (Rejaibi et al., 2022), which makes it imaginable that chatbots could use this technique to assess the progress of depression treatment by listening to the voice of users. To make the conversational interaction more user-friendly, it would make sense for chatbots in this context to gain speech-to-text and text-to-speech capabilities. In terms of perceived trust, vocal communication also plays a role. According to Leitman et al. (2010), cues contained in the speech of voice assistants could influence the attitude of individuals when interacting in the healthcare setting. In fact, voice pitch could change the perception of trustworthiness, where a higher-pitched voice increases trust (Leitman et al., 2010, O'Connor and Barclay, 2017). In addition, social interaction is preferred using voice assistants over chatbots, while chatbots were the preferred means for gathering information (Parnell et al., 2022). Thus, it seems that adding the ability for chatbots to speak out loud like humans could increase the trust towards chatbots as a tool to support people with depression, although this specific aspect has not been studied yet.

Another factor that can contribute to the adoption of CAs is the element of personalization. More specifically, personalization that can influence the emotional state of the user has been shown to affect users' feeling of comfort, which in turn influences trust towards IT (Komiak and Benbasat, 2006). Moreover, personalization that can influence cognitive trust could increase adoption in CAs (Chen and Park, 2021). In other words, personalization could increase user acceptance (Liu and Tao, 2022). The idea of combining anthropomorphism and personalization has not been extensively researched yet for chatbots, but it could be a promising contribution to increasing the adoption of chatbots for treating depression.

It is clear that there is room for improvement in the adoption of chatbots that could support people with depression. Therefore, the following main research question is derived:

How does vocal communication and its personalization in chatbots affect the intention to use of young adults with depression?

Here, vocal communication is defined as the chatbot using a voice to speak out loud its messages. Personalization is defined as the option to choose the voice of the chatbot based on different genders and speed rates.

1.2 Research Questions

In order to answer the main research question formulated in Chapter 1.1, the question is decomposed into three sub-questions. This is summarized in Figure 2, which is intended to give an overview of how all the sub-questions are related to the main research question.

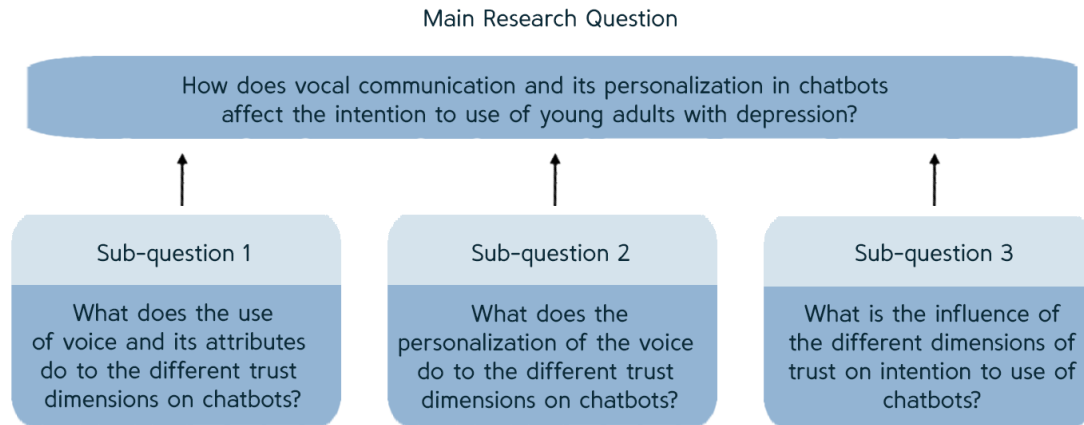


Figure 2: Decomposition of the main research question.

SQ1: What does the use of voice and its attributes do to the different trust dimensions on chatbots?

This sub-question aims to investigate how the voice of a chatbot could influence one's trustworthiness. More specifically, trust is divided into two types of trust: affective trust and cognitive trust. Furthermore, this sub-research question aims to identify the influence of the attributes of the voices of the chatbot. These attributes are the gender of the voice and speech rate.

SQ2: What does the personalization of the voice do to the different trust dimensions on chatbots?

With this sub-question, the aim is to understand what the impact could be on affective and cognitive trust when the user chooses the voice of the chatbot themselves.

SQ3: What is the influence of the different dimensions of trust on intention to use of chatbots?

From this sub-question, the aim is to gain insights about how affective and cognitive trust affect the intention to use the chatbot.

1.3 Research Scope

As understood from Section 1.1, the constructs vocal communication and its element of personalization are factors that could influence human behavior when applied in chatbot technology. As such, multiple angles could be taken to study their relationship within the fields of cognitive science and the study of technology. However, this study is specifically delineated to the conjoint areas of cognitive science and the study of technology. This focus is represented in Figure 3, where the dashed area represents the scope of study for this thesis. The study aims to provide advice on how to design chatbots more responsibly and safely for the end-user in the context of depression. In other words, the focus is on designing chatbots in a way that young adults with depression can trust chatbots enough to use them. This means that this study does not explore topics that are solely cognitive or technology-based. Thus, it neither discusses software architectural principles, nor examines the treatment effectiveness of depression, nor focuses on how people perceive constructs like trust.

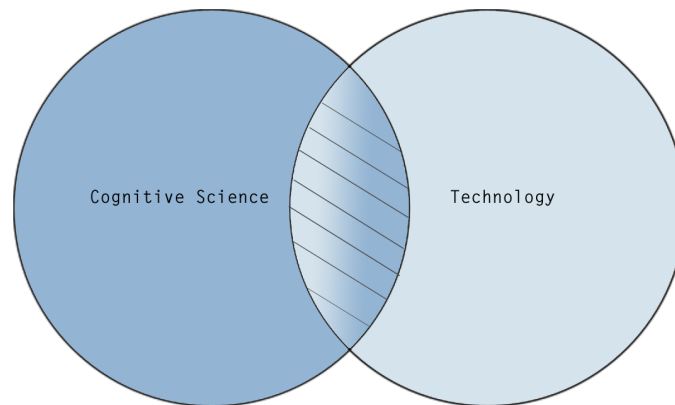


Figure 3: The scope of this research is delineated to the conjoint areas of cognitive science and the study of technology.

1.4 Relevance of the Study

The results from this thesis can contribute to the scientific field of CAs, specifically in understanding how adoption could be improved for chatbots that support young adults with mental health issues like depression. These findings may encourage further research, exploring other factors that could influence trustworthiness and intention to use these chatbots. Although the findings from this study are not directly implementable yet, with further research, healthcare technology companies would eventually be able to use the gained insights as guidelines for building such chatbots. Furthermore, since there is limited research done on voice assistance and treating depression ([Striegl et al., 2022](#)), this thesis could bring more insights about the impact of voice use on users with depression. Similarly, voice assistant builders could benefit from the findings. As mentioned before, the practical contribution of this thesis is that more people can be supported in the treatment progress of depression, considering the shortage of mental health professionals ([Darzi and Evans, 2016](#)). Additionally, it could help mitigate the workload of mental health professionals. Moreover, this would make mental healthcare more accessible, given the shortage of mental health providers in developed countries ([Butryn et al., 2017](#)), and benefit developing countries, where healthcare access tends to be more limited ([Peters et al., 2008](#)). The results could also contribute to increasing the long-term well-being of society, supporting governmental institutions in AI regulation in the healthcare sector, and sparking discussions on the ethical aspects of using chatbot technology to support people with depression.

The topic of this thesis also aligns with the Management of Technology (MOT) study program. Not only does it involve understanding the influence of chatbots on trust and intention to use, which falls within the purview of studying technology's impact on society, but the results from this thesis can also serve as a valuable corporate resource. Furthermore, this thesis topic is interdisciplinary in nature, as it requires insights from multiple perspectives, such as technology, cognitive science, and ethics.

1.5 Thesis Outline

This thesis comprises seven chapters, adhering to the design science research methodology adapted from [Wieringa \(2014\)](#). Chapter 1 serves as an introduction to the research topic, delineating the research questions addressed. Building upon this foundation, Chapter 2 offers an extensive literature review, providing a comprehensive understanding of the relevant aspects pertinent to designing and conducting the study. Chapter 3 expounds upon the employed research methodology, elucidating the intricacies of the chosen approach. Subsequently, Chapter 4 delves into the validation process of the experimental design, ensuring its appropriateness and reliability. In Chapter 5, the results of the study are presented, showcasing the outcomes following the execution of the research. In Chapter 6, a comprehensive analysis of the study's findings and trends is furnished. Additionally, this chapter encompasses discussions on the research's contributions and potential avenues for future exploration. Chapter 7 draws the thesis to a close, encapsulating the conclusions derived from the study and furnishing recommendations based on the outcomes. Supplementary information, including pertinent details and additional insights, can be found in the appendix located at the conclusion of this thesis.

2 Literature Review

This thesis extensively examines relevant constructs essential for addressing the main research question and its sub-questions, as outlined in Chapter 1. These valuable insights are acquired through a rigorous and comprehensive literature review. This chapter starts by elucidating the employed method for conducting the literature review. Then, the actual literature review is provided, which entails an in-depth discussions of the theoretical underpinnings of vocal communication, its personalization, trustworthiness, and intention to use, which play a crucial role in understanding the identified knowledge gaps. Additionally, the thesis explores behavioral activation and conversational agents for depression to provide contextual understanding and guide the research design. Subsequently, a conceptual framework, including the formulated hypotheses, is presented to structure the subsequent analysis.

2.1 Methodology of Literature Review

To identify knowledge gaps, a literature review is required. To understand the state of the knowledge systematically, this study used the 2020 guidelines of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) (Page et al., 2021). The primary purpose of PRISMA is to enhance the reliability and reproducibility of systematic reviews and meta-analyses by promoting complete and transparent reporting. In this study, four literature review rounds were conducted, each with its own purpose and results. This is demonstrated in Figure 4.

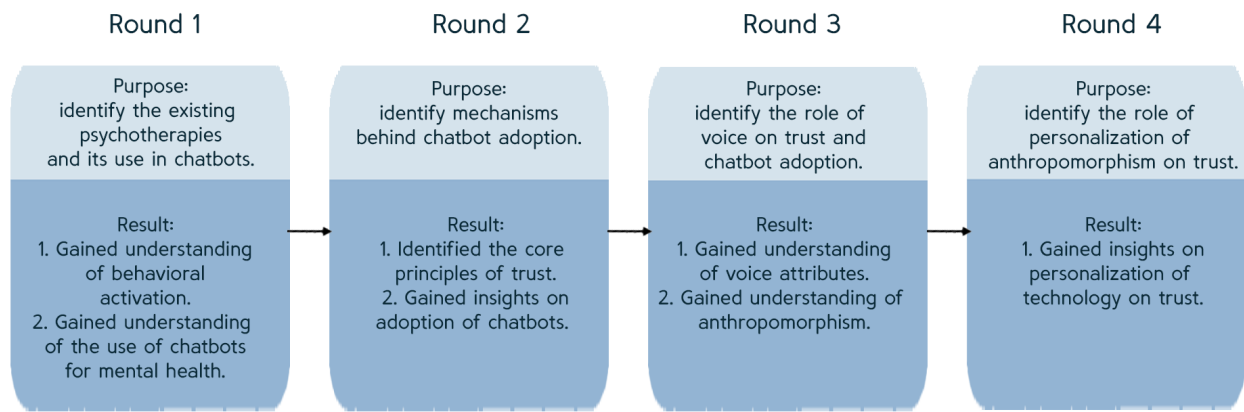


Figure 4: Literature review rounds.

The first literature review aims to identify the existing psychotherapies and their use in chatbots. Table 1 contains the search strings, the number of hits, and selected sources. Since Google Scholar was used as the database, every search resulted in an extensive number of hits, most of which were irrelevant (Gusenbauer, 2019). This explains the high number of hits, which was also the case with all the other literature review rounds. After checking the title and abstract of the relevant articles, a total of 30 articles were selected based on their relevance to core concepts. Exclusion criteria included articles describing hardware robots and early detection of depression. As a result, this round of literature review provided a better understanding of behavioral activation and the use of chatbots for mental health.

Table 1: Search strings: First literature review.

Search string	Hits	Selected
(behavioral OR behavioural) activation literature review	19400	3
cognitive (behavioral OR behavioural) therapy literature review	16700	3
("artificial intelligence" OR ai) chatbot	17600	4
("conversational agents" OR chatbot) "mental health"	8270	8
("conversational agents" OR chatbot) AND depression AND ("behavioral activation" OR "behavioural activation" OR "cognitive behavioral therapy" OR "cognitive behavioural therapy")	2170	12

The second literature review aims to identify the mechanisms behind chatbot adoption. For this purpose, the accumulated search strings from Table 2 were used. Exclusion criteria included articles using the Technology Acceptance Model (TAM) to describe chatbot adoption, as TAM does not focus on different dimensions of trust. From all the hits, a total of 40 articles were selected from Google Scholar. As a result, this study was able to identify the core principles of the concept of trust, and broadens the understanding of chatbot adoption.

Table 2: Search strings: Second literature review.

Search string	Hits	Selected
trust (technology OR "information system") literature review	1840	3
trust ("artificial intelligence" OR ai) in healthcare	18200	5
(cognitive OR affective) trust	265000	3
(affective OR cognitive) trust AND (chatbot OR "conversational agent")	12500	3
(affective OR cognitive) trust AND (adoption OR "intention to use")	84300	4
(chatbot OR "conversational agent") AND health AND (adoption OR "intention to use")	13600	12
(chatbot OR "conversational agent") AND mental health AND (adoption OR "intention to use")	6610	10

The third literature review aims to identify how voice could play a role in trust and adoption when given to a chatbot as a means of communication with its user. Table 3 shows the used search strings, the number of hits, and selected articles. A total of 19 articles were identified, which increased the understanding of the role of voice attributes in influencing trust and amplified the understanding of anthropomorphism.

Table 3: Search strings: Third literature review.

Search string	Hits	Selected
trust AND voice AND (chatbot OR "voice assistant")	15000	3
anthropomorphism chatbot (trust OR adoption)	5300	4
anthropomorphism chatbot (trust OR adoption) AND health	4150	6
anthropomorphism	98300	3
voice attributes trust	18700	3

The fourth literature review aims to identify the core principles of the personalization of anthropomorphism and its relationship with trust. Table 4 shows the used search strings, the number of hits, and the selected articles. A total of six articles were identified that increased the understanding of the role of personalization of technology on trust.

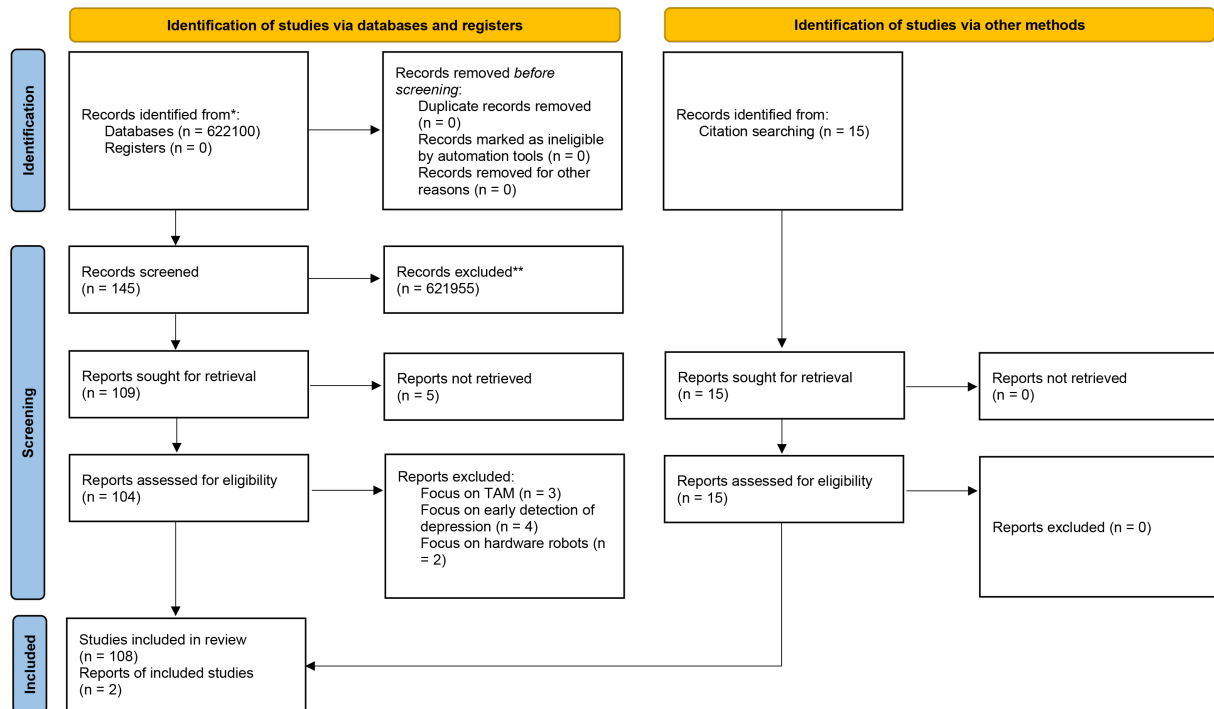
Table 4: Search strings: Fourth literature review.

Search string	Hits	Selected
(personalization OR personalisation) AND anthropomorphism AND trust	8870	3
(personalization OR personalisation) AND technology AND trust	5740	2
health AND anthropomorphism AND (personalization OR personalisation) trust	3950	1

Finally, additional sources were found using the backward snowballing method, which involves finding more articles by looking through the references of the papers that were already included. This has resulted in selecting 15 articles.

To summarize all the rounds of literature reviews, as well as the flows due to the backward snowballing method, the entire flow can be found in 5. In the following Chapters 2.2, 2.3, 2.4, 2.5 and 2.6, all the results from conducting the literature review are summarized.

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources



*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers).

**If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: <http://www.prisma-statement.org/>

Figure 5: PRISMA 2020 diagram: a summary of all the literature review rounds.

2.2 Behavioral Activation

The idea of finding ways to treat mental health issues started in the 1970s ([Lewinsohn, 1974](#)), in which cognitive behavioral therapy (CBT) has been thus far the most researched and effective method ([Beck, 1979](#), [Dimidjian and Davis, 2009](#)). However, a newer approach emerged that was considered to be more effective to treat depression: behavioral activation (BA) ([Martell et al., 2001](#), [Richards et al., 2016](#)). Since people with depression have a reduction of reinforcement for non-depressive behavior, BA effectively addresses this by essentially nudging people to participate in pleasant activities ([Hopko et al., 2003](#)). How this behavior is encouraged depends on the situation of the patient and the chosen methodology of the mental health clinician.

The basic unanimous protocol of BA begins with assessment techniques to understand the current behavioral situation of the patient so that the right activation assignments could be designed, and to have a baseline activity measurement to compare with as treatment progresses ([Kanter et al., 2010](#)). One of these assessment techniques is called activity monitoring. In this assessment, the activities that the patient is currently doing on a daily basis are explored, and the mood related to each activity is then mapped. This assessment is then concluded by explaining the treatment rationale, that there is a meaningful relationship between mood and activity. Then the assessment of goals and values of the patient is inventoried to guide activation assignments. More specifically, the patient thinks about a possible list of valuable life domains and is assisted in identifying and articulating important values, leading to specific statements of values. These values are then used to recommend particular activation assignments that correspond to the value. For example, the value statement "I value being a good parent" for instance offers activation tasks like spending more time with one's kids or fixing a kid's damaged toy.

After concluding the two assessment techniques, activation techniques follows to improve interaction with environmental resources that can provide positive reinforcement. One of them is called activity scheduling, with the idea that the clinician will assign homework in one way or another (for example, written on activity charts, in day planners, other forms made for this purpose, or just verbally specified) to the patient in order to encourage them to engage in certain behaviors that will increase contact with positive non-depressive reinforcers. More specifically, the instructions of the assigned homework usually specify the behavior of interest in detail, which includes the what, where, when and how of the behavior (for example, play tennis at 8AM on Sunday with friends on the playground for 1 hour). Sometimes additional tasks are assigned that could support the patient to execute the desired behavior (for example, to prepare a gym bag with appropriate clothing the night before) ([Martell et al., 2001](#)). In the case that the patient is lacking the ability to emit the desired behavior part of the activity scheduling or does not know how to engage in effective behavior, the patient is suggested to receive additional skills training (usually social skill related training). Sometimes patients also receive relaxation training to target sleep difficulties, and contingency management to increase the likelihood of successful completion of activation.

2.3 Conversational Agents for Depression

The concept of intelligent machines that are able to support humans with decision-making and problem-solving came together in 1950 by Alan Turing ([Caldarini et al., 2022](#)). Artificial intelligence (AI) has evolved to come the closest to what Alan had in mind ([Poola, 2017](#)). One of its applications is known as Natural Language Processing (NLP), which is able to automatically analyze and represent human language ([Cambria and White, 2014](#)). This made it possible for the creation of dialogue systems that are able to respond back in a personalized matter ([Klopfenstein et al., 2017](#)), or as how [Araujo \(2018\)](#) defines it "intelligent software that uses natural language as input and generates natural language as output to engage in a conversation with the user". Although the definition and taxonomy of these dialogue systems are inconsistent in the literature ([Maedche et al., 2019](#), [Stieglitz et al., 2018](#)), in this thesis these dialogue systems are defined as conversational agents (CA). Depending on its application, these CAs could be addressed differently. Conversational agents that are visually text-based are called chatbots (CB), whereas it would be referred to as voice assistants (VA) when it is voice-based ([Diederich et al., 2022](#)).

With the recent advancement of AI, NLP, computing power, and available data, more interest have emerged in diverse industries ([Adamopoulou and Moussiades, 2020](#)). This has resulted in CAs being used as task-oriented CAs and social CAs ([Mnasri, 2019](#)), of which the latter is most promising to support people with mental health problems like depression ([Rathnayaka et al., 2022](#)). As such, attempts have been made to design CAs to help people with depression.

For example, Woebot is a chatbot that is created to provide personal support for users with mental problems. It uses the psychotherapy technique called cognitive behavioral therapy, which is also used in physical setting

by medical practitioners. As such, Woebot has shown to be able to reduce symptoms of depression significantly ([Bae Brandtzæg et al., 2021](#), [Fitzpatrick et al., 2017](#)). The chatbot Bunji has a similar purpose as Woebot but uses behavioral activation as its psychotherapy method ([Rathnayaka et al., 2022](#)). More specifically, the psychotherapy protocol used includes personalized conversation (which is mostly equivalent to BA's activity monitoring), emotional support (which is mostly equivalent to BA's activity scheduling), and remote health monitoring (a method to keep track of the progress of one's mood). Wysa is another mental health chatbot, of which was shown to be more effective when users are more engaged ([Inkster et al., 2018](#)). Another attempt to help people with mental health issues was made with the introduction of chatbot ChatPal, but is targeted to people in rural areas ([Boyd et al., 2022](#)). Chatbot iHelp provides self-assessment and guidelines for depression, sleep, anxiety, self-esteem and stress ([Cameron et al., 2019](#)). With chatbot CARO, users are able to have empathetic conversations and receive medical advice for depression ([Harilal et al., 2020](#)).

As for voice bots, less research has been done on its application for treatment of mental health issues. For example, one attempt was made of combining voice assistance with CBT to explore its usability ([Striegl et al., 2022](#)). Although it was shown to be usable, more interestingly was that the use of voice assistants was preferred over chatbots among elderly participants. However, no significant difference was found among young people. Furthermore, adding the ability of being emphatic did not contribute to the perceived usability and acceptance ([Gotthardt et al., 2022](#)).

The results from existing chatbots to treat mental health problems like depression seem promising, although there are still factors hindering the widespread adoption ([Hengstler et al., 2016](#), [Huckvale et al., 2020](#), [Müller et al., 2019](#), [Nadarzynski et al., 2019](#)). According to [Liu and Tao \(2022\)](#), concerns around topics related to anthropomorphism and personalization are contributors that slow down the acceptance of chatbots. This is in line with other studies that suggest that the aspect of personalization seems to contribute to willingness to adopt ([Lewis et al., 2022](#), [Lipschitz et al., 2019](#)). In addition, it was also suggested that engagement is important to elicit behavioral change ([Perski et al., 2017](#)), something that the anthropomorphism of chatbots could potentially influence ([Moriuchi, 2021](#)). Other factors that could influence the adoption mental health mobile apps are affordability, informational content, explanations of mental health conditions in plain language, cutting-edge security measures like encryption and de-identification, and the inclusion of emergency contacts ([Alqahtani and Orji, 2020](#)). According to [Melcher et al. \(2022\)](#), students who use mental health apps have concerns related to privacy and the accuracy of recommendations.

2.4 Trustworthiness & Intention to Use

The notion of trust has been historically investigated in the context of relationships between people, organizations, and society ([Rousseau et al., 1998](#)). This has shown that trust induces a sense of belonging ([Schoorman et al., 2007](#)) and promotes human contact and cooperation without friction ([Botsman, 2017](#)). In addition, the trustee's perceptions of expertise, the quality of being well-meaning, and integrity in the trustor are shown to be key determinants of trust ([Mayer et al., 1995](#)).

When applying the concept of trust in the context of information systems, research has traditionally addressed the relationship among people in which information systems are the mediators ([Lee and See, 2004](#)). With the increase in automation and complexity of information systems, the information system itself has become the trustee in the relationship with its users ([Söllner et al., 2016](#)). This has led to the idea that perceived trust is important for technology acceptance and adoption ([Bahmanziari et al., 2003](#)).

According to [Rousseau et al. \(1998\)](#), trust is defined as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another". Trust is seen to be especially relevant in risky circumstances when the trustor depends on the trustee's conduct ([Mayer et al., 1995](#)). One of the most important aspect on which trust depends is the judgment of the user on the competency in terms of system quality and information quality ([Nasirian et al., 2017](#)). Users cognitively process information and evidence that is accessible, and they anticipate that technology (i.e. CAs) will perform as promised. This is also known as cognitive trust, and this type of trust is more rational in nature. Although much research tends to mostly focus on the dependence of trust on cognitive assessment ([Chen and Park, 2021](#)), trust also depends on affective assessment ([Lewis and Weigert, 1985](#), [Schoorman et al., 2007](#)) and to some extent on the social aspect ([Botsman, 2017](#)). This aspect of affective trust reflects the faith of having emotional security and the confidence one places in another based on feelings ([Rempel et al., 1985](#)). In the context of CAs, users feel affective trust only when CAs are responsive, show consideration and care, create emotional connections, are socially attractive, and make the

user feel comfortable and secure relying on the CA ([Chen and Park, 2021](#), [Komiak and Benbasat, 2006](#), [Nasirian et al., 2017](#), [Troshani et al., 2021](#)). Moreover, in the context of neurology, trust is a changing affective state that involves both emotional and cognitive components ([Adolphs, 2002](#)), which is also shown in the human–intelligent agent interaction context ([Komiak and Benbasat, 2006](#)) and in the financial context ([Lappeman et al., 2022](#)). In other words, trust develops through logical and careful reasoning (cognition-based), paired with a consideration of feelings and intuition (affect-based). It has also been shown that positive emotions, logical justifications, or a mix of both can serve as the driving forces behind trust ([Chen and Park, 2021](#)).

The concepts of cognitive trust and affective trust could be described further using antecedents, individual factors that contribute to these two dimensions of trust ([Calefato et al., 2015](#)). Affective trust could be assessed by looking at integrity and benevolence of the trustee. Cognitive trust could be assessed by looking at ability to perform and predictability of the trustee. While personal information may increase trust in terms of perceived benevolence, professional information may be utilized to assess the trustee's ability to perform and predictability ([Redmiles and Schumann, 2011](#)). Moreover, the antecedents of affective trust and (partially) cognitive trust are in line with the Tripod model, which states that integrity, benevolence, and ability to perform are required to assess the trustworthiness of a person or organization ([Mayer et al., 1995](#)).

With the introduction of conversational agents, it has become clear that the degree of its intelligence and the way they interact with users could fundamentally affect the perception of trust ([Maedche et al., 2016](#)). As a result, chatbots exhibit traits of both interactive and intelligent IS system classes, which are two separate but interrelated system classes. On the one hand, encounters with chatbots are gradually advancing towards the level of interpersonal exchanges due to anthropomorphic traits ([Pfeuffer et al., 2019](#)), including the formation of emotional attachments ([Purington et al., 2017](#)). In fact, it has been shown that interpersonal communicational competence (ICC) could increase social presence ([Skjuve and Brandzaeg, 2019](#)), and in return, perception of trust ([Wald et al., 2021](#)). On the other hand, accountability and privacy issues are brought up by the systems' autonomy as well as the pervasiveness of the intrusive technologies incorporated in them ([Cowan et al., 2017](#)). This is in line with the fundamental idea that perceived trust of chatbots relies on a coherent mix of cognitive and emotional trust. Understanding this relationship in the context of treating depression using chatbots could be of value since treating people with depression using BA depends on the exchange of information. For example, some users distrust chatbots to be able to provide them with the same level of service as humans could ([De Cicco et al., 2022](#)), and that trust in chatbots is potentially lower when they must share private or sensitive information ([Raunio, 2021](#)). As such, increasing cognitive and affective trust could help alleviate these problems. Finally, there are others factors that relate to trust in chatbots. One study suggested that increasing initial trust could also contribute to enhancing engagement and the intention to use chatbots ([Mostafa and Kasamani, 2021](#)). Another study suggested that the context in which the use of chatbot is applied could influence trust ([Aoki, 2020](#)).

Intention to use is related to one's behavior and is defined as the "intended use", which, in turn, could determine the actual usage ([Pousttchi and Wiedemann, 2007](#)). As suggested earlier, the adoption and acceptance of technology depend on the perception of trust ([Bahmanziari et al., 2003](#)). This is also applicable for the intention to use chatbots ([De Cicco et al., 2022](#)), which has been studied in several contexts, such as chatbots in insurance ([Rodríguez Cardona et al., 2021](#)), chatbots in the financial context ([Ng et al., 2020](#)), chatbots in the context of shopping ([Kasilingam, 2020](#)), and chatbots in the mental health context ([Kasilingam, 2020](#), [Sitthipon et al., 2022](#), [Zhu et al., 2022](#)). More specifically, studies have shown that both dimensions of trust (affective and cognitive) are required to increase intention to use healthcare-related digital applications ([Meng et al., 2022](#)). Finally, it has been suggested that age could be a moderating factor in the intention to use chatbots, meaning older people tend to put less effort in using them ([Terblanche and Kidd, 2022](#)).

2.5 Vocal Communication

In this thesis, vocal communication is defined as the ability of chatbots to communicate by speaking its messages out loud in audio format. By providing chatbots with a human voice, the chatbot could become a bit more human-like. At least, this has been shown with voice assistants like Siri and Alexa, in which their vocal communication anthropomorphizes the voice assistant by increasing the perceived humanness ([Schroeder and Epley, 2016](#), [Wienrich et al., 2022](#)), which in turn could influence perceived trust ([Følstad et al., 2018](#), [Hsiao and Chen, 2021](#)) and also intention to use ([Melián-González et al., 2021](#), [Ng et al., 2020](#)).

According to [Hirsch et al. \(2002\)](#), the term anthropomorphism is defined as "the attributing of human characteristics to inanimate objects, animals, plants, or other natural phenomena, or to God". This concept came

into existence since social interactions with a robot would become more meaningful if it possessed some sort of human-like characteristics (Duffy, 2003) and invited the user to interact in a conversation (Laurel, 1997). The mechanism behind anthropomorphism could be explained with the social presence theory, which states that when more social cues and human-like features are incorporated into human-computer interactions, individuals are more inclined to exhibit more social behaviors (Nass and Moon, 2000) and can assist the progress of an individual's social understanding (Duffy et al., 2002). Another research also supported the association between human-likeness and social presence, showing that anthropomorphic signals influence users' perceptions of social presence (Araujo, 2018). Moreover, according to the Media Equation Theory, in some situations people tend to treat technical devices like real social actors, even when they are aware of their technical nature (Nass and Moon, 2000). The Uncanny Valley theory, on the other hand, contends that anthropomorphism may have a negative impact since users may not feel comfortable dealing with AI that does not perfectly mirror human-to-human interaction (Mori et al., 2012). In this case, enhancing the human-likeness of a chatbot could significantly increase the users' feeling of eeriness, and the feeling of eeriness negatively influences users' trust in the chatbot, resulting in lowered willingness to reuse the chatbot (Song and Shin, 2022). However, other studies have suggested that with the right degree of anthropomorphism, users develop perceptions of a social presence that increases their trust in the anthropomorphic AI application (Qiu and Benbasat, 2009).

At this point, several ways have been tried to alter the human-likeness of chatbots. It was shown that the presence of visual representation of avatars with chatbots led to more social responses (Toader et al., 2019). Racial appearance seems to contribute to this relationship as well. For example, although black people, in general, are more likely to be stereotyped as less competent than white or Asian people, with Black AI bots the opposite is true. In other words, Black AI bots are rated as having higher levels of competence and humanness than Asian and White AI bots, which has a positive impact on overall satisfaction and may even result in more favorable negotiation behaviors (Davis et al., 2023). Furthermore, names given to a chatbot influence the perception of humanness, where human-like names lead to higher perceived anthropomorphism (Cowell and Stanney, 2005). In addition, communication style, linguistic cues, and personalization are able to affect social perception (Sah and Peng, 2015, van Hooijdonk and Liebrecht, 2018). Finally, in the commercial context, it has been shown that a higher level of perceived humaneness allows more effective conversations and an increase in brand perception. This leads to an increase of the intention of purchase (Roy and Naidoo, 2021). In the context of healthcare, recent research found a positive impact of anthropomorphism on trust in smart healthcare services (Liu and Tao, 2022).

As briefly mentioned before, anthropomorphized CAs could influence trust. More specifically, since anthropomorphized CAs should give users the sense of social attraction and presence (Nowak, 2004, Xu, 2020), this could potentially influence affective trust. As with cognitive trust, this could also potentially be affected by anthropomorphized CAs since the use of an anthropomorphic CA directs the user's attention to the four key natural indications of the CA: competence, responsiveness, actionability, and accessibility (Laurel, 1997).

In this thesis, vocal communication is the anthropomorphized attribute of focus. In fact, it has been shown that the relationship with a voice assistant was correlated with increased trust and anthropomorphism (Seymour and Van Kleek, 2021). The human voice consists of pitch, intensity, and voice quality, which uses these components to communicate emotions vocally (Leitman et al., 2010). It has been suggested that the pitch of the voice could influence perceived levels of trust (Elkins and Derrick, 2013), although the reasons why this relationship exists is not clear. Furthermore, this study also showed that vocal pitch is inversely related to trust at the beginning of the interaction but attenuates over time due to increase familiarity and decreased novelty. Moreover, it was shown that although content is computer-generated, adding a real human voice to communicate the content could increase human-likeness (Schroeder and Epley, 2016). Then, from the literature, it is clear that female voices are preferred in voice assistants (Tolmeijer et al., 2021). This has been shown to be true when using voice assistants for receiving medication advice (Goodman and Mayhorn, 2023). However, according to Tolmeijer et al. (2021), gender-ambiguous voices are as trusted when compared with gendered voices. It has also been shown that the outcome of trust on vocal pitch could be different across general, economic, and mating-related context (O'Connor and Barclay, 2017). Since the outcomes were different in each context, it was inferred that the preference of voice could be context-dependent. Finally, it has also been shown that the rate of speech could also affect perception of trust (Schirmer et al., 2019).

2.6 Personalization of Vocal Communication

Providing the user the ability to choose the voice of a chatbot could be promising for increasing trust, especially when considering that the effects of personalization and anthropomorphism on user acceptance of smart healthcare

services are mediated by trust (Liu and Tao, 2022). This requires understanding the concept of personalization, which is context-dependent (Blom and Monk, 2003). According to Vesanen (2007), personalization within the context of technology is defined as "personalization is the use of technology and customer information to tailor electronic commerce interactions between a business and each individual customer. Using information either previously obtained or provided in real-time about the customer, the exchange between the parties is altered to fit that customer's stated needs as well as needs perceived by the business based on the available customer information". It has been suggested that the process of personalization could be iterative (Murthi and Sarkar, 2003) and could happen within a continuous loop, in which operations describe what is done at the different stages of the process, and the objects are defined as the elements that are needed to perform the operations (Vesanen and Raulas, 2006). For example, in the context of marketing, one first learns about the customers' preferences, then matches the offerings to the customer, after which this process of learning and matching is evaluated and applied in the next cycle.

Personalization within the context of chatbots to treat depression could be regarded in two ways: personalization of appearance/presentation and personalization in terms of machine intelligence. When it comes to appearance personalization of technology (e.g. mobile phones and personal computers), several factors could induce the urge to personalize (Blom and Monk, 2003). For example, factors such as ownership and ease of personalization could bring about the desire to change the appearance of technology. As a result, the effects on the user of these factors could touch upon the social and cognitive emotions of the user, respectively, which by itself could reinforce the urge to personalize further. Furthermore, the concept of personalization of technology has been implicitly linked in the literature with the urge of finding familiarity (Blom and Monk, 2003, Komiak and Benbasat, 2006), suggesting that it is potentially able to influence affective trust. In other words, changing human-like attributes of a chatbot could change the emotional state of the user. In fact, the relationship between eeriness and humanization is moderated by the familiarity of the chatbot avatar (Song and Shin, 2022). Also, people instinctively look for peripheral cues that increase their sense of familiarity when they are in a new place to reduce social uncertainty (Liew et al., 2017).

When it comes to personalization in terms of machine intelligence, users desire to interact with chatbots that makes the user feel heard and understood (Lipschitz et al., 2019). In fact, when using chatbots to treat depression, one of the problems for widespread engagement and adoption is the lack of personalization in terms of intelligence (Graham et al., 2019, Lewis et al., 2022, Nadarzynski et al., 2019), resulting in lower cognitive trust (Lipschitz et al., 2019).

2.7 Conceptual Framework

From the literature review (Chapters 2.2, 2.3, 2.4, 2.5 and 2.6) knowledge gaps about the adoption of chatbots that could support people who experience depression could be derived. Therefore, in this chapter, relevant knowledge gaps are identified that this study aims to investigate. To guide the execution of this study, hypotheses were also identified from common trends within literature findings. These hypotheses guide answering the sub-research questions of this study.

SQ1: What does the use of voice and its attributes do to the different trust dimensions on chatbots?

It has been derived from literature review that anthropomorphized CAs could influence trust. More specifically, since anthropomorphized CAs are expected to give users the sense of social attraction and presence (Nowak, 2004, Xu, 2020), this could potentially influence affective trust. However, this has never been tested before using vocal communication as the vehicle for anthropomorphism. Although, it has been suggested that the pitch of voice, and thus voice itself, could influence perceived levels of trust (Elkins and Derrick, 2013). Similarly, the rate of speech is also able to affect the perception of trust (Schirmer et al., 2019). Regarding cognitive trust, this could also potentially be affected by anthropomorphized CAs, as the use of an anthropomorphic CA directs the user's attention to the four key natural indications of the CA: competence, responsiveness, actionability, and accessibility (Laurel, 1997). Given that cognitive trust is about perceived competency, it is possible that anthropomorphism and cognitive trust could have a relationship. However, this has never been tested before, particularly when using vocal communication as the vehicle for anthropomorphism.

Hypothesis 1: Users with depression who interact with a chatbot that has vocal communication capabilities are more likely to have a higher perceived cognitive trust.

Hypothesis 2: Users with depression who interact with a chatbot that has vocal communication capabilities are more likely to have a higher perceived affective trust.

SQ2: What does the personalization of the voice do to the different trust dimensions on chatbots?

It has been suggested that the two essential characteristics of smart technology are anthropomorphism and machine intelligence (expressed mainly by personalization) (Liu and Tao, 2022). In fact, within the context of the latter, it has been shown that low retention rates of mental health apps could be explained by the lack of personalization, and that the element of personalization could increase user engagement with more personalized feedback (Lewis et al., 2022). This also affect the cognitive trust of the user (Lipschitz et al., 2019). However, personalization within this context is not further explored in this thesis. Instead, this thesis focuses on personalization of appearance within the context of mental healthcare.

The concept of personalization of technology has been implicitly linked in the literature with the urge of finding familiarity (Blom and Monk, 2003, Komiak and Benbasat, 2006). As such, it is possible that users could choose a voice for the chatbot that feels more familiar to feel more comfortable. This could increase affective trust, something that has not been explored yet within this context. Although, since it has been suggested that personalization of appearance could influence both the cognitive and emotional response of users under the right conditions (Blom and Monk, 2003), choosing a suitable voice for the user could potentially increase both cognitive and affective trust, with the latter having a stronger relationship. These relationships have not been explored yet.

Hypothesis 3: Users with depression who interact with a chatbot that has personalized vocal communication capabilities are more likely to have a higher perceived cognitive trust.

Hypothesis 4: Users with depression who interact with a chatbot that has personalized vocal communication capabilities are more likely to have a higher perceived affective trust.

SQ3: What is the influence of the different dimensions of trust on intention to use of chatbots?

It has been shown that perceived trust could influence intention to use CAs (De Ciccio et al., 2022). However, trust is arguably two-dimensional: cognitive trust and affective trust. Although one study has been conducted to show the relationship between these two dimensions of trust and intention to use (Chen and Park, 2021), it has never been tested in the context of supporting people with depression using chatbots that are able to speak.

Hypothesis 5: Depressed users with high perceived cognitive trust in the chatbot are more likely to have intention to use the chatbot.

Hypothesis 6: Depressed users with high perceived affective trust in the chatbot are more likely to have intention to use the chatbot.

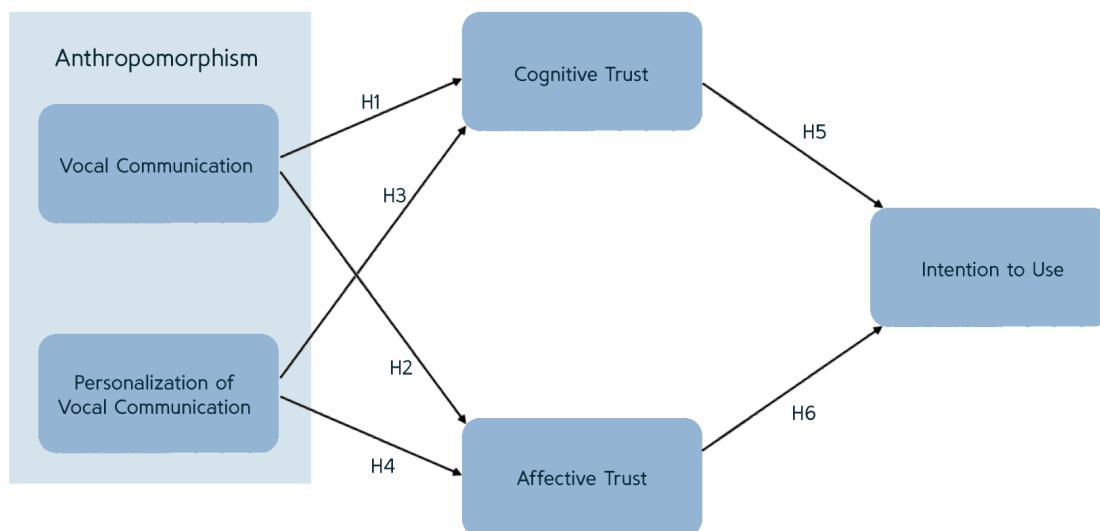


Figure 6: The conceptual framework that was derived from combining the knowledge gaps identified from the literature. Hypothesis are also displayed between the constructs.

From all six hypotheses, it is clear that investigating the idea of enabling chatbots to communicate out loud (vocal communication) and allowing the user to choose their preferred voice (personalized vocal communication) could be promising for chatbot adoption. To elucidate the interrelationships between these constructs and the alignment of the hypotheses within these studied relationships, a conceptual framework was formulated (see Figure 6). In this framework, each hypothesis is presented by its corresponding abbreviation. For instance, H1 corresponds to Hypothesis 1, H2 to Hypothesis 2, and so forth. To be precise, the conceptual framework was synthesized in this study through combining different literature findings and suggestions.

3 Research Method

This chapter describes the testing of the conceptual framework and its hypotheses through well-defined research methodologies. Firstly, the research approach and framework are expounded upon. Subsequently, a comprehensive exposition of the experimental design is provided, offering detailed insights into the study's methodological intricacies.

3.1 Research Approach

To address the main research question and its decomposed sub-questions, the design science methodology proposed by [Wieringa \(2014\)](#) was adopted. This research framework is tailored for studying knowledge questions or resolving design problems within the context of information systems. Knowledge questions are about obtaining knowledge about how the world works, in which the knowledge is obtained by investigating the relationship of the study object (also known as artifact) in the real-world. This research starts with a proposition claimed to be true, which after conducting the study, the results are evaluated on truth, degree of certainty, evidence, replicability, alternative explanations and generalizability. On the other hand, design problems call for a change in the world and are referred to (re)designing an artifact so that it better contributes to the achievement of some goal of stakeholders. The results are then evaluated on utility, goal contribution, limitations and reusability. Given that the main research question of this study *"How does vocal communication and its personalization in chatbots affect the intention to use of young adults with depression?"* pertains to understanding the artifact's behavior in real-world scenarios, this study employed the design science methodology to investigate knowledge questions. Since it is common for these types of problem to be investigated using quantitative methods, this study did the same with an experimental design (more in Chapter 3.2). However, the qualitative methodology was only used to investigate the open feedback from participants, which represents only a small part of the study (more in Chapter 3.2.3).

The design science methodology adapted for answering knowledge problems consists of five stages, which is depicted in Figure 7. First, the research design is selected in the research & inference design phase. Then, in the second phase, the validity and reliability of the research design are checked in the validation phase to ensure the production of robust and credible data. Only when this is true, the study could then be conducted in the research execution phase. In the final phase (data analysis), the collected data is then subjected to data analysis to identify trends and facilitate meaningful reflections. Ultimately, the goal is to address the knowledge problems and provide relevant answers. However, usually new knowledge problems could arise from the data analysis phase, which could bring the research back to phase 1 (research problem analysis) to start a new study. This continuous research cycle is also known as the empirical cycle.

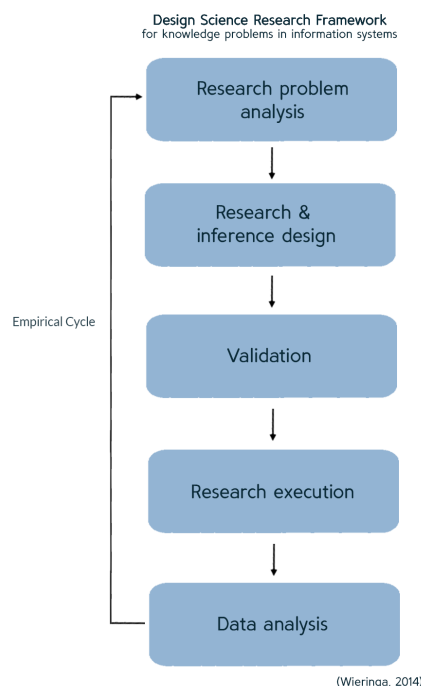


Figure 7: Design science methodology adapted by [Wieringa \(2014\)](#). This particular framework is suitable for knowledge problems in information systems.

In order to understand how the design science methodology was implemented in this study, a comprehensive overview is provided in Figure 8. The research flow of this study commences with the research problem analysis phase, served to identify the knowledge problem. This phase leads to the definition of the main research question, its decomposed sub-questions and the corresponding hypotheses, which are considered as the output of this phase (found in Chapter 1 and Chapter 2). Based on the sub-questions, the research approach was determined during the research & inference design phase (presented in this chapter). The validation phase was utilized to ascertain the collection of valid and reliable data through the experimental design (discussed in Chapter 4). Subsequently, the collected data is presented in Chapter 5 as the output of the research execution phase. Finally, relationships were explored from the findings in the data analysis phase (elaborated in Chapter 6), which made it possible to answer the main research question.

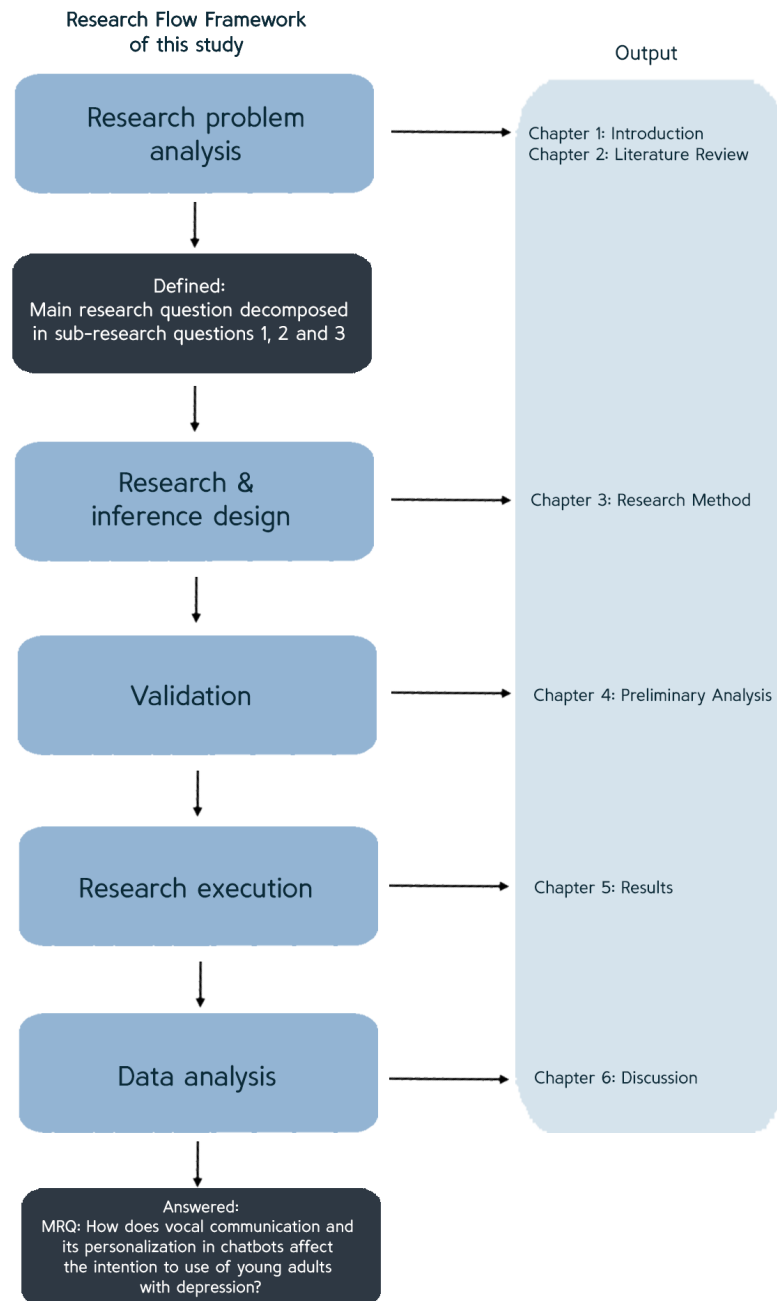


Figure 8: The research flow framework of this study. Each phase of the design science methodology produced an output that could be found in the designated chapter.

3.2 Experimental Design

Chapter 3.1 describes how the design science methodology as the chosen research approach was adopted, tailored for addressing knowledge problems. Considering the prevalent use of a quantitative approach in resolving such problems, this chapter meticulously examines the experimental design employed in this study, providing a comprehensive explanation of its implementation.

3.2.1 Overview Experimental Design

To address sub-question 1, *"What does the use of voice and its attributes do to the different trust dimensions on chatbots?"*, a between-subjects design was selected, which is an experimental study design that involves manipulating an independent variable and randomly assigning participants to different groups, each experiencing a distinct condition of the independent variable (Charness et al., 2012). Subsequently, the responses of these groups are compared to assess the effect of the independent variable. To address sub-question 1, the group involved with the manipulation of an independent variable (treatment group) was allowed to interact with a chatbot that was only able to communicate through speech (vocal communication), while the group with no manipulation involved (control group) was allowed to interact with a chatbot that was only able to communicate through text (no vocal communication). Thus, the independent variable was vocal communication. Given that the objective of sub-question 1 was to observe the influence of vocal communication on different dimensions of trust, the dependent variables were measured as affective trust and cognitive trust, assessed through questionnaires (see Chapter 3.2.3). Additionally, to address the objective of sub-question 3, *"What is the influence of the different dimensions of trust on the intention to use of chatbots?"*, intention to use was also measured in this between-subjects design using questionnaires. To provide a clearer illustration, this between-subjects design is depicted in Figure 9, where the measurement stage represents assessing affective trust, cognitive trust, and intention to use to address both sub-question 1 and 3.

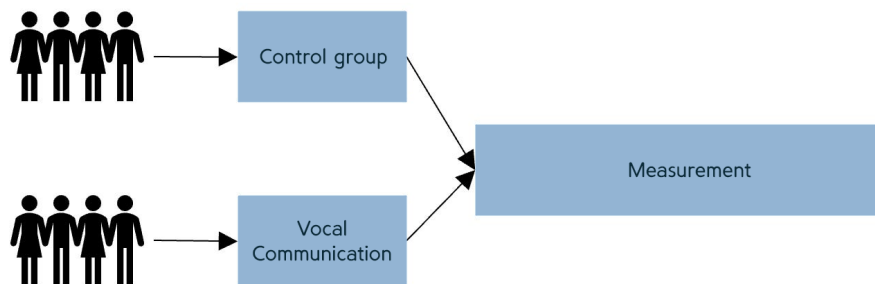


Figure 9: The between-subjects design used to collect data for sub-research question 1 and 3.

Thus far, sub-question 1 was addressed by designing a between-subjects design. However, this was only the first half of the design. To address sub-question 1 in its entirety, a between-subjects factorial design (a combination of between-subjects and factorial design) was actually adopted to address sub-question 1.

In factorial design, participants are randomly assigned to one of the vignettes (Steiner et al., 2016, Wallander, 2009). Typically, a vignette is a systematically varied descriptions of situations or persons to elicit the beliefs, attitudes, or behaviors of respondents with respect to the presented scenarios. In this study, the vignettes contain a combination of the variables of interest, which were gender of voice and speech rate. Both are characteristics of the voices used by the chatbots. Consequently, the factorial design followed a 2 (gender: male, female) x 2 (speech rate: slow, fast) configuration, as depicted in Figure 10. Hence, participants of this treatment group could be randomly allocated to one of the four available chatbots, each equipped with a distinct combination of gender of voice and speech rate.

Thus, the initial between-subjects design enabled the examination of the influence of voice (vocal communication) on different dimensions of trust. However, to further investigate the impact of voice attributes (which are the characteristics of a voice), a factorial design was incorporated within the existing between-subjects design.

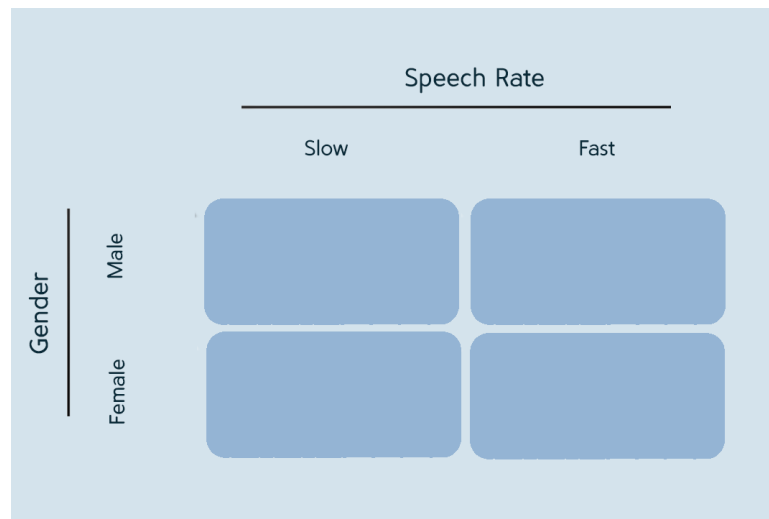


Figure 10: Factorial survey design for a 2 (gender: male, female) x 2 (speed: slow, fast) to address the influence of vocal attributes.

To address sub-question 2 (*What does the personalization of the voice do to the different trust dimensions on chatbots?*), a similar between-subjects study design was employed, akin to the approach used for sub-question 1. The difference is that the group involved with the manipulation of an independent variable (treatment group) was allowed to interact with a chatbot that was only able to communicate through speech after the user has chosen a voice (personalized vocal communication), while the group with no manipulation involved (control group) was allowed to interact with a chatbot that was only able to communicate through text (no vocal communication). In other words, to address sub-question 2, the independent variable under investigation was the personalization of vocal communication. However, unlike the factorial design utilized in sub-question 1, the current study did not employ a factorial approach. Consequently, participants were granted the autonomy to choose one of the four available voices, as opposed to being randomly assigned. To address sub-question 3 (*What is the influence of the different dimensions of trust on intention to use of chatbots?*), in this between-subjects design the intention to use was also measured using questionnaires. To have a better understanding, this between-subjects design is demonstrated in Figure 11, in which measurement represents measuring affective trust, cognitive trust and intention to use to address both sub-question 2 and 3.

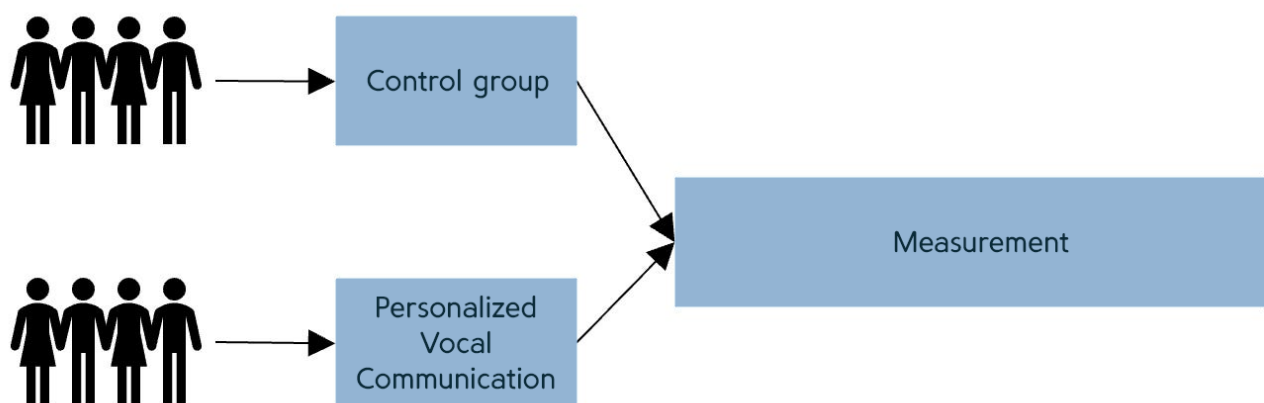


Figure 11: The between-subjects design used to collect data for sub-research question 2 and 3.

In the end, to address sub-questions 1, 2 and 3, two very similar between-subjects designs were required. The biggest difference was the independent variable that was involved. Therefore, both the between-subjects designs were incorporated in one cross-sectional survey study, in which the participants were equally randomized and assigned to either the control group, or one of the treatment groups that represented one of the two independent variables. More specifically, Treatment Group I introduced independent variable vocal communication, while Treatment Group II introduced independent variable personalized vocal communication. As mentioned before, to address

sub-question 1 in its entirety, factorial design was also implemented in Treatment Group I.

It was chosen to use the online survey platform Qualtrics to conduct the survey study. Not only does Qualtrics make it easier to share the survey with many potential participants, but it was also able to randomly allocate participants to either the control group or the treatment group. Given the focus on young adults experiencing depression as the unit of analysis, identifying such individuals could prove challenging. To expedite participant recruitment, the online marketplace Prolific was employed. Prolific, renowned for its academic research participant recruitment services ([Palan and Schitter, 2018](#)), attracts individuals who are less prone to dishonesty and more naive, contributing to a relatively higher quality of participants for lab experiments ([Peer et al., 2017](#)). Since Prolific has the option to delineate the target group to specific attributes (pre-screening requirements), this was utilized to find the right candidate for this study. More specifically, only participants who experience depression were allowed to be included in the study. In addition, only participants between the age of 18 and 30 were allowed, since the unit of analysis were young adults. Participants were also filtered on hearing ability: participants should not have hearing difficulties, since participants were required to listen to messages using the voice of the chatbots. Moreover, a similar number of male and female participants were recruited, not only to enhance randomization but also to serve as a baseline for analyzing the influence of voice attributes, as gender may impact voice preferences ([Jones et al., 2010](#)). Also, participants from diverse geographical locations were allowed to participate in the study, ensuring a broad and representative sample. A monetary reward of £0.80 was established for participants who successfully completed the entire study. Participants who did not fulfill this criterion or who did not complete the research from start to finish were excluded from the analysis. Finally, participant size required for this particular study was estimated to be around 600. This was estimated using the software G*Power ($f = 0.21$, uses ANOVA one-way between groups analysis, $\alpha = 0.05$, power = 0.8, number of groups = 3) based on average results from the effect size of cues to action from previous studies ([Al Shamsi et al., 2022](#), [Lee et al., 2021](#), [Schreuter et al., 2021](#)).

3.2.2 Survey Design

In this section, the survey design is explained, presenting the sequential experience that participants would encounter during the survey. To guide the explanation, a visual representation of the participant survey flow is provided in Figure 12.

Informed Consent & Pre-screening

As participants entered the survey, they were requested to provide consent for the study's data collection. Once participants agreed to the provided consent, their eligibility was verified based on the pre-screening requirements of Prolific, ensuring that the correct participants were selected by the platform.

Introduction Context

Subsequently, the study commenced with a brief introduction, prompting participants to envision themselves experiencing depression with a lack of interest in engaging in daily activities. Notably, participants were informed about the existence of a chatbot capable of offering long-term mood elevation assistance. Subsequently, participants were invited to partake in an intake session with the chatbot to familiarize themselves with its capabilities. Once the participant confirmed that they were situated in a silent environment, they were randomly allocated to one of the research groups.

Between-subjects designs

As previously mentioned, in order to comprehensively address all the sub-questions, a unified survey design was devised by incorporating all the between-subjects designs. This integrated design is also depicted in Figure 12 under the "Between-subjects designs" section, where three distinct lanes are presented for the control group, the treatment group for investigating sub-questions 1 and 3, and the treatment group for addressing sub-questions 2 and 3.

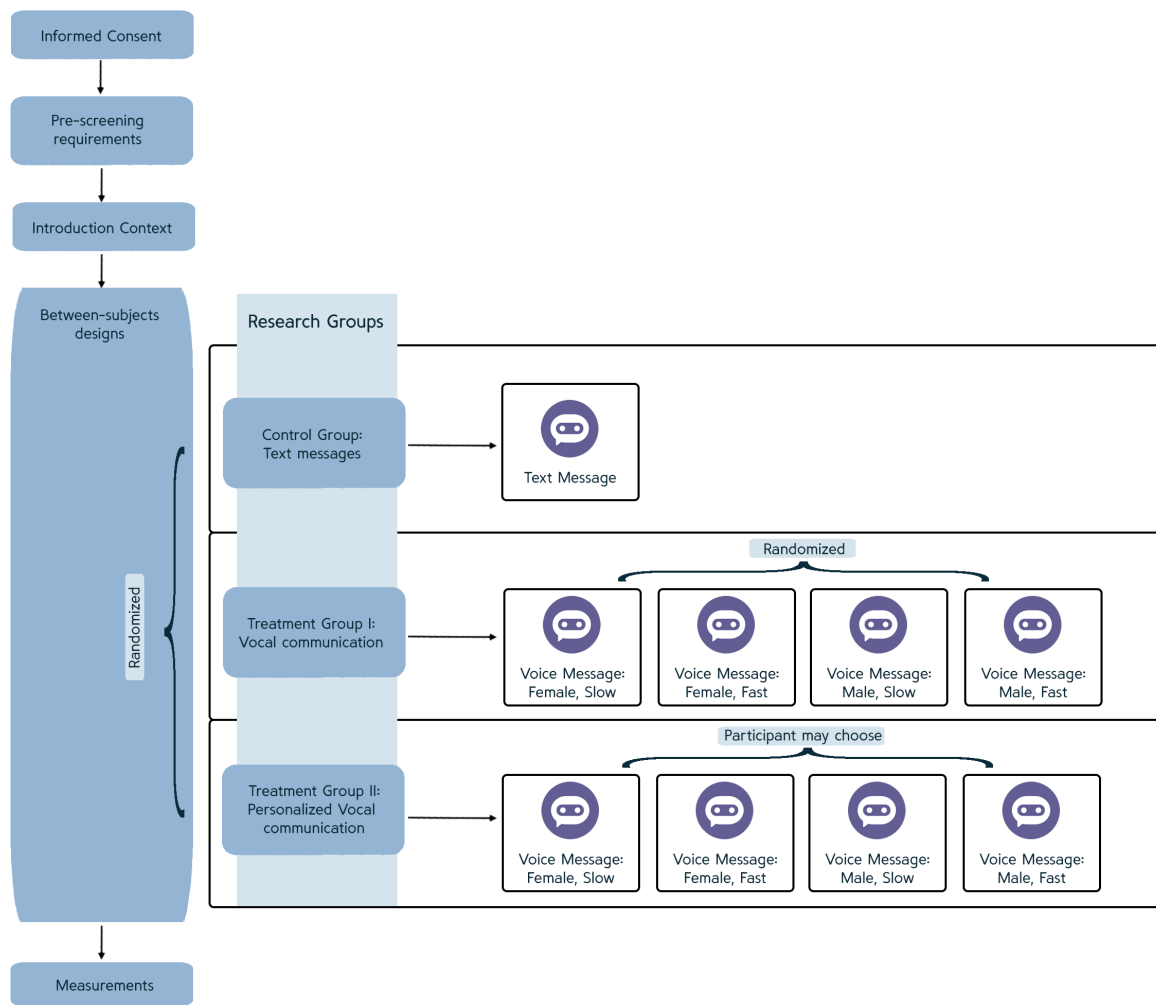


Figure 12: The overview of the entire research design.

Research group: Control group

The results obtained from the control group were utilized for comparative analysis with those from each of the treatment groups, aiming to assess the levels of affective trust, cognitive trust, and intention to use.

Participants assigned to the control group engaged with a chatbot capable of text-based communication. The chatbot was built using the online platform Botstar (pro version). Botstar employs decision-tree principles to create chatbots, wherein all possible content provided to users is predetermined based on limited input options. Consequently, the chatbot used in this study did not incorporate artificial intelligence technology for generating personalized responses. Given the study's focus on improving the mood of participants experiencing depression, the chatbot's entire dialogue system was structured and scripted following the behavioral activation protocol (see Appendix A for the script). The script commenced with activity monitoring, prompting the participant to specify one activity. Subsequently, the participant was asked to identify a personally meaningful value or goal, followed by an explanation of its significance in relation to activities. Finally, participants were encouraged to consider planning to engage in an activity that contributes to what they find personally meaningful, akin to activity monitoring within the behavioral activation protocol.

Research group: Treatment Group I

Participants assigned to Treatment Group I, intended to address sub-question 1, were further randomized to be allocated to one of the four available chatbots, each equipped with a voice that varied in terms of gender and speech rate. The outcomes from this treatment group aimed to investigate whether chatbots utilizing voice messages could impact trustworthiness and intention to use, in comparison to the control group. It was also explored if voice attributes could potentially play a role in this relationship.

Once participants were assigned to Treatment Group I, an invitation was given to interact with a similar Botstar chatbot as used in the control group. However, the key distinction lay in the mode of communication, which involved the delivery of script content through voice messages (vocal communication). The voices employed by the chatbots in this treatment group were prerecorded by professional voice actors, sourced from the online freelance marketplace Fiverr. Specifically, one male and one female actor were enlisted to read the script, with instructions to adopt a neutral speed and an emphatic tone of voice. These voice attributes were selected as they are deemed suitable and well-accepted means of communication for individuals experiencing depression (Gotthardt et al., 2022). It should be noted that gender attributes were inherent to the voice actors, whereas speech rates of the voices were digitally manipulated in post-production. Specifically, the speech rates were adjusted to be slightly slower and slightly faster, by a margin of 20%, to explore their influence on user perception and interaction.

Research group: Treatment Group II

Participants in Treatment Group II, intended to address sub-question 2, underwent a process similar to that of Treatment Group I, wherein the chatbot communicated through voice messages. However, a key distinction lay in the participant allocation procedure. Instead of further randomization, participants in this treatment group were given the autonomy to select one of the four available voices after listening to a preview of each voice. The outcomes from this treatment group were aimed at investigating whether chatbots employing user-selected voice messages could influence trustworthiness and intention to use, in comparison to the control group. Given that participants were permitted to choose the voice of the chatbot themselves, it should be noted that the results from this treatment group could not be employed to assess the effect of voice gender and speech rate on all the dependent variables of this study. This was due to the variability introduced by individual voice preferences and selections, rendering it challenging to isolate the specific influence of voice attributes on the measured dimensions. Nonetheless, the insights gained from this treatment group provided valuable information on user preferences and the potential impact of user-selected voices on trust and intention to use.

Measurements

Upon completing their interaction with the chatbot, participants were presented with a comprehensive list of questionnaires designed to measure various constructs, including affective trust, cognitive trust, intention to use, and other relevant variables (refer to Chapter 3.2.3 for a comprehensive overview of all measures). These questionnaires were thoughtfully selected to capture crucial aspects of the participants' perceptions and experiences, contributing to a more holistic understanding of the research question. A detailed representation of the survey layout for participants can be found in Appendix C, offering a visual depiction of the survey's structure and question sequence.

3.2.3 Data collection

In Chapter 3.2.2, it was mentioned that data was collected at the end of the survey design (Measurements). In this section, the measurements used to collect the data is discussed.

Upon completing their interaction with the assigned chatbot, participants' responses to the dependent variables were measured using sets of questionnaires, also known as scales. Each set of questionnaires typically consisted of at least two questions, or items, designed to assess the specific dependent variable of interest. To ensure uniformity and consistency, a 7-point Likert scale was employed throughout this study to collect participants' responses. This section provides a detailed explanation of the design and structure of the questionnaires used to measure the two dimensions of trust (cognitive trust and affective trust), intention to use, anthropomorphism, overall trust, and vocal communication preference. Although not explicitly covered in Chapter 3.2, the inclusion of measures for overall trust, anthropomorphism, and vocal communication preference was deemed necessary to enhance the comprehensiveness of the findings and contribute to a more comprehensive narrative. Additionally, since participants were provided with the opportunity to offer open feedback after their interaction with the assigned chatbot, the measurement of the feedback is also addressed in this section.

Two dimensions of trust

The questionnaires used to measure cognitive trust and affective trust were adapted from a study conducted by Calefato et al. (2015), tailored to suit the specific context of this research (refer to Appendix B for the detailed questionnaires). As previously highlighted in the literature review, cognitive trust encompasses the dimensions of ability and predictability, which were measured individually and their responses subsequently aggregated. In contrast, affective trust involves integrity and benevolence. However, considering that the development of one's moral

and ethical norms requires time and multiple interactions [Rusman et al. \(2010\)](#), the measurement of integrity was deemed challenging in this study, as participants were exposed to the chatbot for the first time. Consequently, affective trust items related to integrity were excluded from the assessment, aligning with prior research findings ([Calefato et al., 2015](#), [Redmiles and Schumann, 2011](#)). Thus, the measurement of benevolence effectively represented affective trust. Additionally, the measurement of overall trust was directly inquired from participants towards the end of the survey for indicative purposes, enabling comparison with the outcomes obtained from measuring the two dimensions of trust. As a result, the questionnaires pertaining to overall trust were administered subsequent to the assessment of the two dimensions of trust, ensuring that they did not influence the responses to the trust dimensions. The obtained data required quantitative analysis for further exploration and interpretation.

Intention to use

The questionnaire designed to measure the intention to use the chatbot was adapted from a study conducted by ([Melían-González et al., 2021](#)), customized to suit the specific context of the present research (see Appendix B for the detailed questionnaire). Furthermore, participants were directly queried at the conclusion of the survey regarding their preference for voice messages over text messages in the chatbot, representing their vocal communication preference. As previously mentioned, assessing vocal communication preference aimed to reinforce the primary findings and the narrative of this study. To analyze the obtained data effectively, quantitative analysis was deemed necessary.

Anthropomorphism

In order to further augment the understanding of the relationship between chatbots using voice and text messages as modes of communication, participants were directly questioned about their perceived anthropomorphism of the chatbot at the end of the survey. The questionnaire used to measure perceived anthropomorphism was derived from the work of [Bartneck et al. \(2009\)](#), and subsequently adjusted to suit the specific context of this research (see Appendix B for the questionnaires). To effectively analyze the collected data and draw meaningful insights, a quantitative analysis was deemed essential.

Demographic information

In order to check whether the collected data were randomized sufficiently, demographic data (age, gender and educational level) was collected from the participants. More specifically, it was also important to check whether a similar number of male and female participants finished the study, since one's gender could influence one's voice preference in certain contexts ([Jones et al., 2010](#)). Furthermore, country of residence was automatically collected by Prolific. For exploratory purposes, country classification (developed versus developing) was added using sources from [International Monetary Fund \(IMF\) \(2022\)](#). In order to analyze the results, quantitative analysis was required.

Open feedback

In the end of the survey study, the option was given to participants to provide textual comments about their experience with the survey and with the chatbot. The purpose of asking open feedback is to potentially understand why certain trends were observed from the questionnaires. In order to analyze the results, qualitative analysis was required.

3.2.4 Data Analysis Methods

From Chapter 3.2.3, it was explained how data was collected. However, the collected data itself in its raw form cannot be used directly to answer the sub-questions. Instead, data processing and analysis were required to find the relevant results that could contribute to answering the sub-questions.

In this section, all the relevant data analysis methods are discussed, in which general information is given to understand the concepts behind the data analysis techniques. First, two statistical tools (Mann-Whitney U Test and Ordinal logistic regression) are explained that was used in this study. Then, two types of statistical relationships (mediation and interaction effects) are described, which were important to use ordinal logistic regression properly. Finally, the qualitative approach is discussed to address analyzing the open feedback data. All data processing for analysis were done with the statistics software R (see Appendix H) and in Microsoft Excel.

Mann-Whitney U Test

Since all the dependent variables in this study were measured using a Likert-scale, the collected data were nonparametric in nature. Thus, to quantitatively compare the results from all the dependent variables, the Mann-Whitney U Test was used. This test, also known as the Wilcoxon rank-sum test, is a non-parametric statistical test used to compare two independent groups and determine whether there are significant differences between their distributions

(McKnight and Najab, 2010). It is an alternative to the independent two-sample t-test, but it does not assume that the data follows a normal distribution. Therefore, the Mann-Whitney U test is often used when the data is ordinal (such as the Likert-scale) or when it fails to meet the assumptions of the t-test, such as when the data is highly skewed or has outliers. As a result, Since the output is a sum of ranks, differences in sample sizes do not affect the output. However, the bigger the sample size difference, the smaller the effect size becomes. The output of this test would be the sum of the ranks of the first sample (W-value).

In order to determine the effect sizes of a Mann-Whitney U Test, Cohen's D is calculated that is adapted to data that follows a nonparametric distribution (see Figure 13 for the entire formula) (Ivarsson et al., 2013).

$$Cohen's D = \sqrt{\frac{Eta}{(1 - Eta)}} * 2 \quad Eta = |(Z * Z)/(N1 + N2)| \quad Z = \left| \frac{U - \left(\frac{N1 * N2}{2}\right)}{\sqrt{\frac{N1 * N2 * (N1 + N2 + 1)}{12}}} \right|$$

Figure 13: The adjusted formula for Cohen's D, using Mann-Whitney U Test (U) and participant group size N1 and N2.

Ordinal Logistic Regression

When looking for association relationships like investigating mediation relationships or interaction effects, Ordinal Logistic Regression (OLR) was the most suitable method to use in this study. Ordinal logistic regression, also known as proportional odds logistic regression, is a type of regression analysis used to model relationships between one or more independent variables and an ordinal dependent variable. An ordinal variable is a categorical variable with a meaningful order (e.g. Likert scale), but the differences between the categories are not necessarily equal. In other words, OLR assumes that one or more independent variables need to be categorical, ordinal or continuous (Kleinbaum and Klein, 2010). Another important assumption that holds with OLR is the proportional odds assumption. This assumption states that the relationship between each pair of outcome groups has a similar slope. This makes it also suitable for factorial experiments (South et al., 2021).

Mediation

A mediation relationship refers to the process through which the mediator (M) explains or accounts for the relationship between the independent variable (IV) and the dependent variable (DV) (BarNir et al., 2011). The mediator acts as an intermediate variable that helps to elucidate the underlying mechanism or pathway by which the independent variable influences the dependent variable. The mediation relationship is often depicted using a simple causal diagram:

$$IV \rightarrow M \rightarrow DV$$

The arrow from the independent variable to the mediator (IV → M) represents the effect of the independent variable on the mediator. The arrow from the mediator to the dependent variable (M → DV) represents the effect of the mediator on the dependent variable. The mediation relationship suggests that the influence of the independent variable on the dependent variable is at least partially transmitted through the mediator.

Mediation can be understood in terms of direct effects, indirect effects and total effect. The direct effect refers to the influence of the independent variable (IV) on the dependent variable (DV) without considering the mediator (M). In other words, it is the effect of the IV on the DV that is not explained by the mediator. The indirect effect refers to the influence of the independent variable (IV) on the dependent variable (DV) that is transmitted through the mediator (M). It represents the effect of the IV on the DV that operates through the mediator variable. The indirect effect provides insights into the underlying mechanism through which the independent variable influences the dependent variable. If the indirect effect is statistically significant, it indicates that the mediator plays a role in transmitting the effect of the independent variable to the dependent variable (Hayes, 2009). The total effect of the independent variable (IV) on the dependent variable (DV) is the sum of the direct effect and the indirect effect. It represents the overall influence of the independent variable on the dependent variable, considering both the direct pathway and the pathway mediated through the mediator.

Interaction Effect

An interaction effect is a fundamental concept in statistical analysis, particularly in the context of factorial designs and regression models (Jaccard and Turrisi, 2003). It arises when the effect of one independent variable on the dependent variable is modified by the presence or level of another independent variable.

Mathematically, an interaction effect can be represented through an interaction term in a regression model. Suppose we have a study with two independent variables, X and Y, and a dependent variable, Z. The regression model can be written as:

$$Z = \beta_0 + \beta_1 X + \beta_2 Y + \beta_3 * (X * Y) + \epsilon$$

In this equation: Z represents the dependent variable. X and Y are the two independent variables. β_0 , β_1 , β_2 , and β_3 are the coefficients to be estimated through the regression analysis. $X*Y$ is the interaction term that captures the combined effect of X and Y. ϵ represents the error term. The coefficient β_3 (the interaction coefficient) determines the strength and direction of the interaction effect. If β_3 is statistically significant, it indicates the presence of an interaction effect. The interpretation of the interaction effect depends on the context of the study and the scaling of the variables involved.

Thematic analysis

To address analyzing the open feedback data, thematic analysis was used in this study. Thematic analysis is a qualitative research method used to identify, analyze, and report patterns (themes) within a data set (Braun and Clarke, 2012). The process of thematic analysis typically involves coding the data, which in this study was chosen for the inductive approach. Then, codes were categorized into themes.

3.2.5 Data Analysis Strategy

In the previous Chapter 3.2.4, the data analysis methods used in this study were explained. However, to extract valuable information and trends from the collected quantitative data, a data analysis strategy is required to extract the relevant information to answer the sub-questions. Typically, this involves splitting the data set into groups so that comparisons can be made between these groups to identify trends. Therefore, in this section, the data analysis strategy is explained, along with how the data analysis methods were applied in this study based on the chosen data analysis strategy. More specifically, the section begins with an overview of the general methods of data comparison used throughout this study. Subsequently, an overview is provided of the specific relationships investigated by following one of the general methods of data comparison, while also highlighting the data analysis methods employed.

Comparison method 1

Once data was collected after conducting the experiment described in Chapter 3.2, relationships in the collected data were studied in order to answer the sub-questions. There were three ways in which data was compared. Figure 14 demonstrates the first method, in which collected data was compared between two research groups to observe the effect on the dependent variable. For example, collected data from the control group was compared with Treatment Group I, so that the effect of vocal communication on affective trust could be studied.

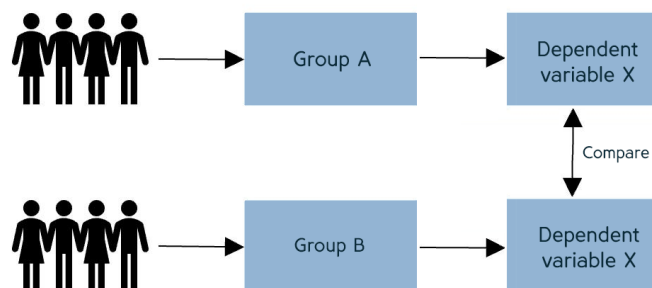


Figure 14: Comparison method 1: Data was compared between two groups (A and B) to observe the effect on the dependent variable (X).

Comparison method 2

The second method was similar to that of the first method. However, it differed by first stratifying participants based on one socio-demographic characteristic, as can be seen in Figure 15. In other words, the context of the research groups were changed by selecting only participants who possess the socio-demographic characteristic of interest. For example, collected data from participants in the control group were compared with collected data from participants in Treatment Group I, in which all the participants were living in developing countries. Then, the effect of vocal communication on affective trust could be studied, while taking into account of that specific socio-demographic characteristic being part of the research context.

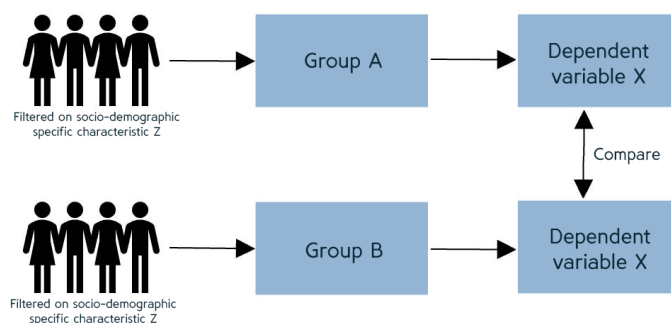


Figure 15: Comparison method 2: Data was first stratified based on a socio-demographic characteristic (Z), after which data was compared between two groups (A and B) to observe the effect on the dependent variable (X).

Comparison method 3

The third method of comparison is visualized in Figure 16. Here, data from one research group was first stratified into two subgroups based on one of the categories of socio-demographic characteristics. For example, participants from Treatment Group I were divided into a subgroup containing only participants being male and into a subgroup containing only participants being female. In this method, the influence of the socio-demographic characteristic could be observed on the dependent variable. In other words, different from the other two comparison methods described above, vocal communication and personalized vocal communication were part of all things being equal when comparing the subgroups (*ceteris paribus*).

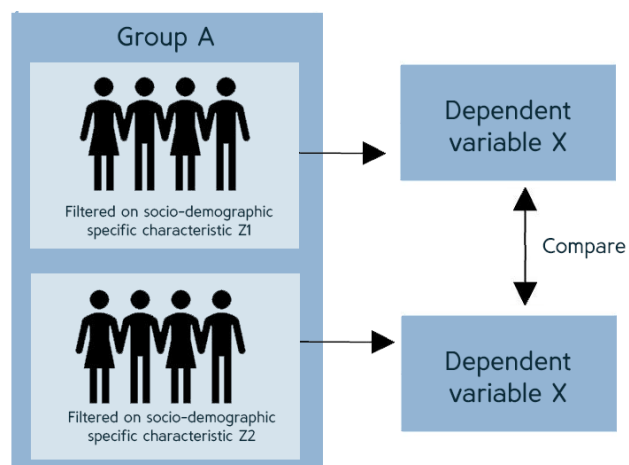


Figure 16: Comparison method 3: Data from a research group (A) was first separated into two subgroups based on socio-demographic characteristics Z1 and Z2. Then, these subgroups were compared with each other to observe the effect on the dependent variable (X).

Comparison method 4

The last method of comparison is illustrated in Figure 17, which depicts the approach used to study interaction effects. Data from a research group (A) was initially divided into four subgroups, each assigned a distinct combination of independent variables (a) and (b). The unique combinations available are a1b1, a1b2, a2b1, and a2b2. Subsequently, the collected data of the dependent variables (X) were compared with each other to explore the presence of an interaction effect.

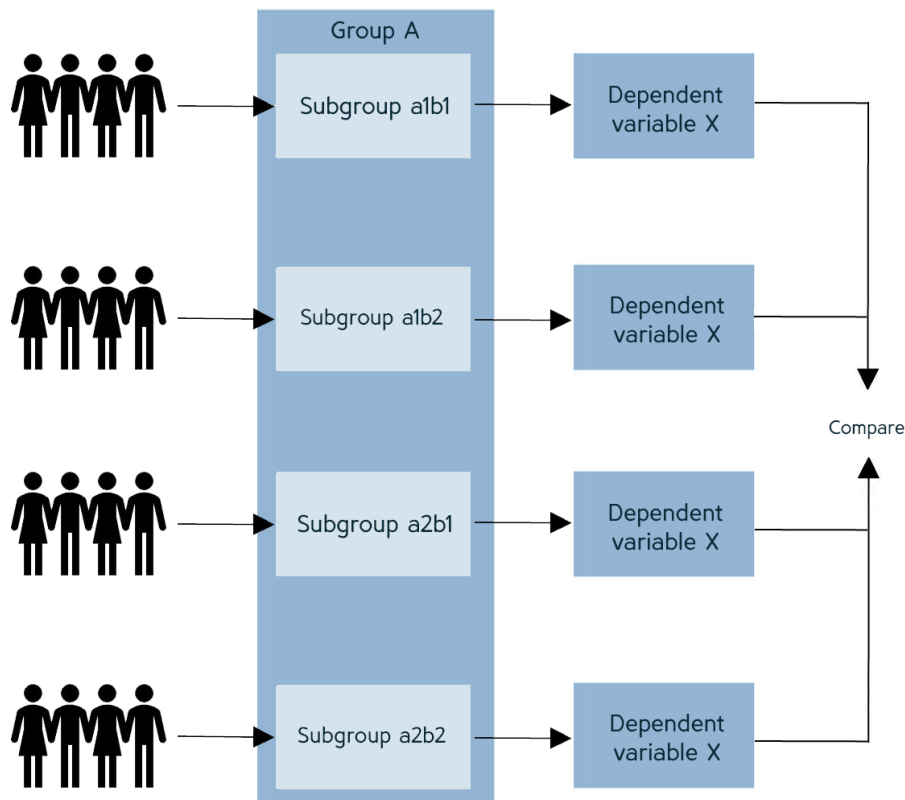


Figure 17: Comparison method 4: Data from a research group (A) was initially divided into four subgroups, each assigned a distinct combination of independent variables (a) and (b). The unique combinations available are a1b1, a1b2, a2b1, and a2b2. Subsequently, the collected data of the dependent variables (X) were compared with each other to explore the presence of an interaction effect.

The investigated relationships

Now that the general methods of comparing data have been explained, the following section provides an overview of all the relationships that were investigated.

The experimental design used in this study provided data, from which many relationships could be identified. To begin with, data from the control group was compared with data from Treatment Group I to identify the causal relationship between vocal communication and the two dimensions of trust (see Figure 14: Comparison method 1). Similarly, data from the control group was compared with data from Treatment Group II to identify the causal relationship between personalized vocal communication and the two dimensions of trust. For both relationships, the Mann-Whitney U Test was used. Since results from the Mann-Whitney U Test do not provide insights on the direction of the relationship (positive or negative), ordinal logistic regression was used to address this problem. However, it was only useful to look for the direction of the relationship if significant results were found using the Mann-Whitney U Test.

Then, the causal relationship between the two dimensions of trust and intention to use was identified by comparing the control group with both Treatment Group I and Treatment Group II apart (see Figure 14: Comparison method 1). For this relationship, the Mann-Whitney U Test was used. Finally, when significant results were found using Mann-Whitney U Test, ordinal logistic regression was also used to determine the direction of the relationship.

In order to explore the role of the two dimensions of trust as a mediator between vocal communication and intention to use, data from the control group was compared with Treatment Group I (see Figure 14: Comparison method 1). More specifically, ordinal logistic regression was used to find the indirect effect and the direct effect. Data from the control group was also compared with Treatment Group II to look for the same type of relationship, but the focus was then on personalized vocal communication instead of vocal communication.

To observe whether there was a difference in results between participants who experienced vocal communication and personalized vocal communication, the results from Treatment Group I and Treatment Group II were compared with each other in regards to the dimensions of trust, intention to use, anthropomorphism, overall trust,

and vocal communication preference (see Figure 14: Comparison method 1). During this investigation, the causal relationship was investigated using the Mann-Whitney U Test. Finally, when significant results were found using Mann-Whitney U Test, ordinal logistic regression was also used to find the direction of the relationship.

In this study, additional dependent variables were studied to support the main findings. As such, the relationship between vocal communication and that of overall trust, anthropomorphism, and vocal communication preference were also studied individually by comparing data between the control group and Treatment Group I (see Figure 14: Comparison method 1). Similarly, this was also done with personalized vocal communication instead of vocal communication by comparing data between the control group and Treatment Group II. For all of these causal relationships, the Mann-Whitney U Test was used. Finally, when significant results were found using Mann-Whitney U Test, ordinal logistic regression was also used to find the direction of the relationship.

From factorial design it was possible to explore the interaction effect of voice attributes, gender of voice, and speech rate on the two dimensions of trust (see Figure 17: Comparison method 4). This was done by using data from only Treatment Group I and utilizing ordinal logistic regression as the data analysis method. Also, the individual effects of gender of voice and speech rate on the two dimensions of trust were also studied. To investigate this type of relationship, data from the control group was compared with data from Treatment Group I (see Figure 14: Comparison method 1). To find this type of relationship, the Mann-Whitney U Test was used. Finally, when significant results were found using Mann-Whitney U Test, ordinal logistic regression was also used to find the direction of the relationship.

Thus far, all the mentioned relationships were based on all the participants without taking into account their socio-demographic characteristics. However, it was possible to stratify the participants in terms of country classification (residence of developed or developing country). Therefore, data from the control group was compared with Treatment Group I, selecting only data from participants from developing countries to investigate the influence of vocal communication on the two dimensions of trust and intention to use (see Figure 15: Comparison method 2). This was also done with only participants who were living in developed countries. Similarly, data from the control group was compared with Treatment Group II, which means that the same type of relationship was studied but focused on personalized vocal communication instead of vocal communication. Since it was also possible to stratify the participants based on gender (participant is male or female), the exact same analysis was done for gender. For all of these causal relationships, the Mann-Whitney U Test was used. Finally, when significant results were found using Mann-Whitney U Test, ordinal logistic regression was also used to find the direction of the relationship.

Moreover, the influence of socio-demographic characteristics were further explored. Data collected from participants in Treatment Group I were separated into participants who lived in developing countries and participants who lived in developed countries. Then, the levels of cognitive trust, affective trust, and intention to use were assessed by comparing the results of these two subgroups with each other to find any significant differences using the Mann-Whitney U Test (see Figure 16: Comparison method 3). This comparison strategy was also done by separating participants solely based on being male or female. Finally, this entire process was also repeated for participants in the control group and Treatment Group II. When significant results were found using Mann-Whitney U Test, ordinal logistic regression was also used to find the direction of the relationship.

4 Preliminary Analysis

Following the design science methodology of [Wieringa \(2014\)](#), in this chapter the quality of the collected data was assessed using the experimental design described in Chapter 3. Before results could be interpreted, preliminary tests were conducted to assess reliability and validity of the scales used to measure the constructs. In addition, the collected data needs to be checked on randomization.

4.1 Randomization Check

Although Prolific was used to invite participants, of which the pool consisted of a similar number of female and male participants, and Qualtrics was used to assign participants randomly to control/treatment groups, it was still important to quantitatively check whether the participant groups were actually randomized. More specifically, there should not be a significant difference between participant groups in terms of socio-demographics characteristics such as age, gender and education. This is essential to enable comparisons between participant groups, while eliminating the undesired influence of socio-demographics characteristics of participants.

Since collected socio-demographics data in this study followed a parametric distribution, one-way ANOVA was used to determine quantitatively whether mean differences were present. When comparing participants group based on treatment/control groups, no significant differences were found ($p > 0.05$). This can be seen in Table 5, which shows that there were no differences in age, gender and education among participants. Similarly, when comparing participant groups within Treatment Group II based on the four possible vignettes, no significant differences were found either ($p > 0.05$). This can be seen in Table 6, which also confirms that there were no differences in age, gender and education among participants. Finally, it is important to check if all participants have provided their gender, age, and education, to ensure no disproportion was present in terms of socio-demographic characteristics. An overview of this could be seen in Table 7, which shows that 100% of all the participants have provided their gender, age and education.

Table 5: Randomization check based on socio-demographics characteristics.

Variables	Control Group	Treatment Group I: Vocal communication	Treatment Group II: Personalized vocal communication	<i>P-value</i>
	<i>Mean ± SD</i>	<i>Mean ± SD</i>	<i>Mean ± SD</i>	
Age	2.59 ± 0.49	2.62 ± 0.49	2.51 ± 0.50	0.11
Gender	1.57 ± 0.59	1.59 ± 0.59	1.55 ± 0.57	0.80
Education	2.65 ± 0.77	2.71 ± 0.86	2.76 ± 0.81	0.16

Table 6: Randomization check for factorial design within Treatment Group II based on socio-demographics characteristics.

Variables	Vignette				<i>P-value</i>
	Female Slow	Female Fast	Male Slow	Male Fast	
	<i>Mean ± SD</i>	<i>Mean ± SD</i>	<i>Mean ± SD</i>	<i>Mean ± SD</i>	
Age	2.61 ± 0.50	2.55 ± 0.50	2.71 ± 0.46	2.63 ± 0.49	0.54
Gender	1.65 ± 0.61	1.47 ± 0.58	1.65 ± 0.56	1.61 ± 0.60	0.51
Education	2.88 ± 0.91	2.69 ± 0.84	2.69 ± 0.88	2.63 ± 0.83	0.53

Table 7: Descriptive statistics of socio-demographics.

Statistic	N	%	Mean	SD	Min	Median	Max
Gender	601	100.0%	1.57	0.58	1	3	3
Age	601	100.0%	2.58	0.50	1	2	4
Education	601	100.0%	2.71	0.82	1	3	5

Note: % indicates the percentage of participants that have disclosed information concerning the variable.

4.2 Attention Check

Survey participants are generally prone to respond without considering the content, and therefore, they may have a tendency to respond more positively. This is known as acquiescent response bias (Watson, 1992). To reduce the possibility of this bias, an attention check was included in the survey. Nearing the end of the survey, participants were asked about the core message of the chatbot (Shamon and Berning, 2019). If participants failed to respond correctly, their data were excluded from the study.

4.3 Validity Analysis

Validity refers to the accuracy of the measurement (Fitzner, 2007). In this study, a sub-type of construct validity, known as convergent validity, was measured. Internal and external validity were also discussed.

Convergent validity was used to check the correlation between items measuring the same construct. Given the ordinal nature of the scale used to measure the constructs in this study, the non normality was first checked and confirmed with the Shapiro-Wilk test (Royston, 1992) (see Appendix D). As such, the nonparametric Spearman's rank correlation coefficient was used for measuring convergent validity (Kruskal, 1958). The guidelines suggested the correlation coefficient to be acceptable above 0.5 with $p < 0.05$ (Carlson and Herdman, 2012). In this study, the correlation between the items of scales anthropomorphism, predictability, overall trust, intention to use, and vocal communication preference were significant ($p < 0.05$) with all of the coefficients above 0.5 (see Appendix E). However, with the scale measuring benevolence, item 10 ("This chatbot keeps me updated about what is in its plans.") had most of its coefficients not meeting the acceptability criteria ($r_{Q1Q10} = 0.50$, $r_{Q2Q10} = 0.38$, $r_{Q3Q10} = 0.39$, $r_{Q5Q10} = 0.28$, $r_{Q6Q10} = 0.035$, $r_{Q7Q10} = 0.43$, $r_{Q8Q10} = 0.42$, $r_{Q10Q11} = 0.43$). Similarly, item 6 ("This chatbot does not provides its services with skills.") and 7 ("This chatbot does things competently/in a capable manner.") of the Ability scale consistently did not meet the acceptability criteria as well. Therefore, results from these three items were removed from data analysis.

Given the nature of lab experiments, the internal validity of this study was relatively high (Gutfleisch et al., 2021). However, external validity had been criticized for lab experiments, as the designed treatment allocation do not always represent real-life situations (Gutfleisch et al., 2021, Simpson and Piquero, 2002, Steiner et al., 2016). A certain level of external validity was present given the randomized controlled assignment of participants into experimental groups. In addition, low ecological validity could be reduced by asking whether participants judged the designed scenario realistic or not (Simpson and Piquero, 2002). During the pre-test of this study, a small number of potential participants were asked to provide feedback. Furthermore, entire survey design and chatbots were evaluated by an expert (Dr. Marcus Westberg from the Technical University of Delft) qualitatively as sufficiently valid. Although this increased the face validity of the study design, the external validity was therefore not necessary more valid.

4.4 Reliability Analysis

Reliability is the ability to measure consistently (Fitzner, 2007). In this study, the internal consistency reliability of measuring the two dimensions of trust and intention to use were determined using Cronbach's Alpha (α) (Santos, 1999). This was necessary to ensure that the set of questions used to measure a construct could produce similar responses when given to a particular individual. A Cronbach's Alpha (α) between 0.70 and 0.95 indicates that the set of questions was measuring the same variable (Santos, 1999, Ursachi et al., 2015). In this study, a few scales had to be adjusted to improve the internal consistency. For example, the predictability scale had insufficient internal consistency ($\alpha = 0.64$). This can be seen in Table 8, where the internal consistency of all the scales is listed. Taking into account the results from using Spearman's rank correlation for measuring convergent validity,

data from item 10 of the benevolence scale, items 6 and 7 of the ability scale, item 1 of the predictability scale and item 4 of the intention to use scale were removed from data analysis to achieve improved internal consistency for all the used scales (see Appendix B for all the questionnaires). As a result, the collected data from the experimental design were valid and reliable enough to be analyzed to answer the main research question and its sub-questions.

Table 8: Internal validity per measurement scale.

Scale	Original N	Original Cronbach's Alpha	Final N	Final Cronbach's Alpha
Anthropomorphism	3	0.93	3	0.93
Benevolence	11	0.94	10	0.94
Ability	7	0.92	5	0.92
Predictability	3	0.64	2	0.70
Overall Trust	2	0.89	2	0.89
Intention to Use	4	0.94	3	0.95
Vocal communication preference	2	0.84	2	0.84

5 Results

This chapter presents the comprehensive findings obtained through the experiment. It commences by providing an account of the participant count. Subsequently, a concise overview of the participant characteristics is provided. Finally, the chapter presents all pertinent findings necessary to address the sub-questions.

5.1 Data exclusion

The methodology outlined in Chapter 3.2.1 established specific inclusion criteria for participant selection. Initially, a total of 931 individuals commenced the survey during the period from June 15 to June 22, 2023. Subsequently, after screening out participants who did not meet the pre-defined requirements (requirements were experiencing depression, absence of hearing difficulties, and age falling within the range of 18 to 30), the remaining pool consisted of 638 participants. Further filtering, involving the exclusion of participants who failed the attention check, resulted in a final sample size of 601 participants. Among these, 201 participants were assigned to the control group, 199 participants to Treatment Group I, and 201 participants to Treatment Group II (see Appendix F).

Within Treatment Group II, a total of 199 participants were included in the analysis of the influence of voice attributes on all the dependent variables used in this study. More specifically, 43 participants were assigned in vignette 1 (female and slow), 51 participants were assigned in vignette 2 (female and fast), 51 were assigned in vignette 3 (male and slow), 54 participants were assigned in vignette 4 (male and fast). This translates to the following participation allocations: as for the vignette of female/male voice, it comes down to 94 participants that were exposed to a female voice and 105 participants that were exposed to a male voice. As for the treatment of slow/fast voice, it comes down to 94 participants that were exposed to a slow voice and 105 participants that were exposed to a fast voice.

5.2 Participant characteristics

This section provides a comprehensive description of the participants' characteristics. Table 9 presents an overview of the socio-demographic attributes of the 601 participants, including gender, age, and education. The distribution shows that 47.59% of the participants identified as male, 47.75% as female, and the remaining respondents self-identified as other. A majority of the participants fell within the age group of 25-30, accounting for 57.57% of the total. Moreover, the majority of participants reported having completed high school or possessing a Bachelor's degree. Table 10 outlines the countries of residence for the participants, indicating that the largest representation came from the United Kingdom (32.11%), followed by South Africa (15.31%) and Spain (7.49%) in the second and third positions, respectively. Then, in Figure 18, 19 and 20 an overview is given of the average levels of all the measured dependent variables for the control group, Treatment Group I and Treatment Group II respectively. As illustrated in all three figures, the average levels for all the dependent variables tended to cluster around the mid-range. For detailed values of each dependent variable per research group, see Appendix G.

Table 9: Socio-demographics characteristics of all the participants.

Variables	Total (<i>N</i> = 601)	Affective Trust	Cognitive Trust	Intention to Use
	<i>N</i> (%)	Mean ± <i>SD</i>	Mean ± <i>SD</i>	Mean ± <i>SD</i>
<i>Gender</i>				
Male	286 (47.59)	172.10 ± 26.29	64.06 ± 9.33	61.99 ± 11.75
Female	287 (47.75)	175.11 ± 25.13	65.33 ± 8.55	62.93 ± 11.63
Other	28 (4.66)	164.14 ± 25.95	61.71 ± 8.89	58.36 ± 11.76
<i>Age</i>				
18-24	255 (42.43)	174.77 ± 25.45	65.31 ± 8.89	63.05 ± 11.56
25-30	346 (57.57)	171.93 ± 26.02	64.00 ± 9.01	61.69 ± 11.81
<i>Education</i>				
Secondary School	27 (4.49)	180.56 ± 25.34	67.59 ± 8.26	64.48 ± 11.91
High School	226 (37.60)	175.69 ± 25.90	65.11 ± 8.68	62.94 ± 11.63
Bachelor's Degree	251 (41.76)	172.53 ± 26.29	64.58 ± 9.32	62.54 ± 11.86
Master's Degree	90 (14.98)	167.40 ± 22.93	62.47 ± 8.39	59.73 ± 11.08
PhD or higher	7 (1.17)	157.57 ± 27.01	60.86 ± 11.70	54.86 ± 12.12

Table 10: Frequency Table Countries of Residence of all the participants.

Country	Country classification	Number of participants <i>N (%)</i>
United Kingdom	Developed	193 (32.11)
South Africa	Developing/Emerging	92 (15.31)
Spain	Developed	45 (7.49)
Poland	Developing/Emerging	35 (5.82)
Portugal	Developed	25 (4.16)
Germany	Developed	24 (3.99)
Italy	Developed	23 (3.83)
Netherlands	Developed	22 (3.66)
Greece	Developed	12 (2.00)
Australia	Developed	11 (1.83)
Belgium	Developed	10 (1.66)
Hungary	Developing/Emerging	10 (1.66)
Canada	Developed	8 (1.33)
France	Developed	8 (1.33)
Ireland	Developed	7 (1.17)
United States	Developed	6 (1.00)
Türkiye	Developing/Emerging	5 (0.83)
Latvia	Developed	4 (0.67)
Mexico	Developing/Emerging	4 (0.67)
Sweden	Developed	4 (0.67)
Indonesia	Developing/Emerging	3 (0.50)
New Zealand	Developed	3 (0.50)
China	Developing/Emerging	2 (0.33)
Finland	Developed	2 (0.33)
India	Developing/Emerging	2 (0.33)
Israel	Developed	2 (0.33)
Korea	Developed	2 (0.33)
Mongolia	Developing/Emerging	2 (0.33)
Nigeria	Developing/Emerging	2 (0.33)
Norway	Developed	2 (0.33)
Pakistan	Developing/Emerging	2 (0.33)
Russian Federation	Developing/Emerging	2 (0.33)
Slovenia	Developed	2 (0.33)
Sri Lanka	Developing/Emerging	2 (0.33)
Suriname	Developing/Emerging	2 (0.33)
Togo	Developing/Emerging	2 (0.33)
Ukraine	Developing/Emerging	2 (0.33)
Argentina	Developing/Emerging	1 (0.17)
Bangladesh	Developing/Emerging	1 (0.17)
Colombia	Developing/Emerging	1 (0.17)
Cyprus	Developed	1 (0.17)
Czech Republic	Developed	1 (0.17)
Estonia	Developed	1 (0.17)
Guyana	Developing/Emerging	1 (0.17)
Lithuania	Developed	1 (0.17)
Morocco	Developing/Emerging	1 (0.17)
Samoa	Developing/Emerging	1 (0.17)
Slovakia	Developed	1 (0.17)
Switzerland	Developed	1 (0.17)
Venezuela	Developing/Emerging	1 (0.17)

Note: Country classification data was sourced from the [International Monetary Fund \(IMF\)](#) (2022).

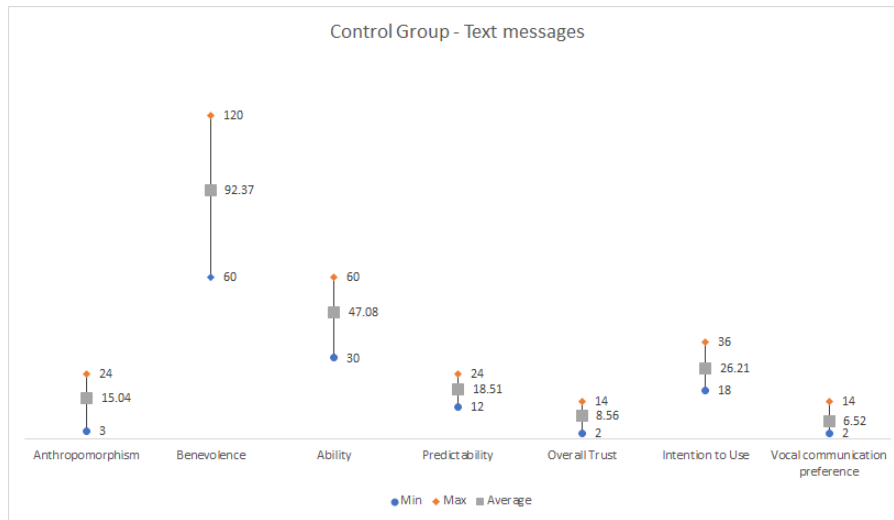


Figure 18: Overview of absolute participant measurements for the control group.

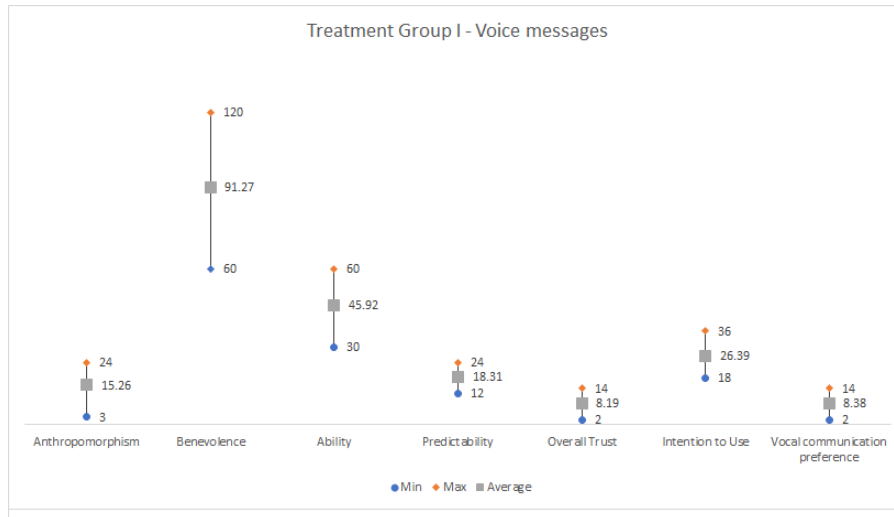


Figure 19: Overview of absolute participant measurements for Treatment Group I.

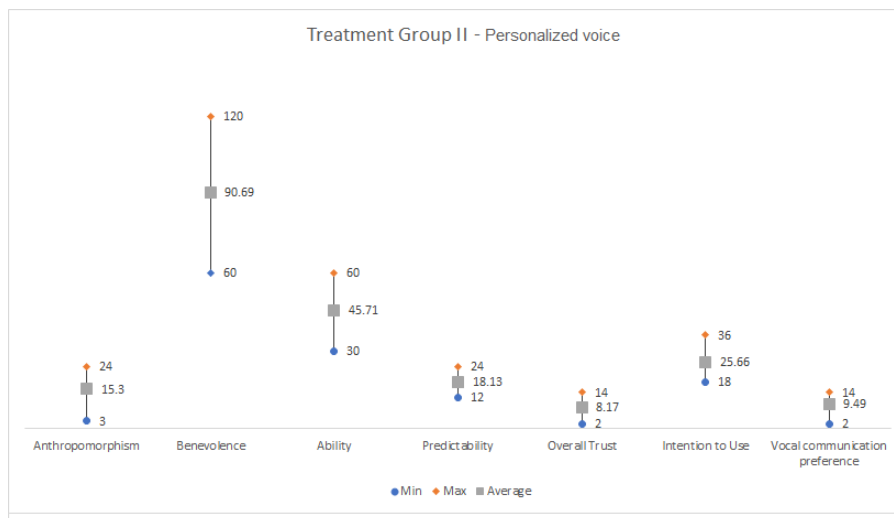


Figure 20: Overview of absolute participant measurements for Treatment Group II.

5.3 Findings

Chapter 3.2.5 provided a comprehensive overview of the data analysis strategy employed in this study. In this section, we discuss the results obtained from implementing this strategy. To facilitate clarity and relevance in addressing the sub-questions, the results are presented per hypothesis.

Hypothesis 1: Users with depression who interact with a chatbot that has vocal communication capabilities are more likely to have a higher perceived cognitive trust.

Hypothesis 1 pertains to the relationship between vocal communication (i.e., the use of voice messages) and cognitive trust. To examine this relationship, we assessed the direct impact of vocal communication on cognitive trust. The results of this analysis are presented in Table 11, which displays the relationships between vocal communication and the sub-components of cognitive trust. It is evident from the table that no significant relationship was observed ($p > 0.05$) between vocal communication and both ability and predictability. Moreover, when combining the results of ability and predictability to evaluate cognitive trust, no significant effect of vocal communication on cognitive trust was found ($p > 0.05$). These findings are summarized in Table 12.

Table 11: The influence of vocal communication on ability and predictability. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test.

Situation	Ability			Predictability		
	Significance (p)	W-value	Effect Size	Significance (p)	W-value	Effect Size
Voice messages	0.06475	17866	0.185	0.74390	19625	0.032

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 12: The influence of vocal communication on cognitive trust. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test.

Situation	Cognitive Trust (Ability + Predictability)		
	Significance (p)	W-value	Effect Size
Voice messages	0.12370	18221	0.154

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Further insights into the relationship between vocal communication and cognitive trust could be gained by exploring additional contributing factors. In Table 13, the influence of voice attributes, such as gender of voice and speech rate, on cognitive trust is examined using the Mann-Whitney U Test as the analytical tool. Notably, the female voice exhibited a significant impact on participants' cognitive trust ($W = 8015.5$, $p = 0.036$), with a small effect size indicated by Cohen's D of 0.151. Subsequently, ordinal logistic regression revealed a significant negative association between the female voice and cognitive trust ($\beta = -0.43$, $OR = 0.65$, $SE = 0.21$, $p = 0.0365$, $95\%CI = -0.84$ to -0.03). Conversely, the male voice did not exert a significant influence on cognitive trust ($p > 0.05$). Similarly, speech rate did not demonstrate a significant effect on cognitive trust ($p > 0.05$). The interaction effect of gender of voice and speech rate on cognitive trust was also explored. As such, the result of the interaction effect is presented in Table 14. As can be observed, no significant interaction was found ($\beta = -0.641$, $p > 0.05$). In Figure 21 the interaction effect of gender of voice and speech rate on cognitive trust is visually presented. Here, it can be seen that the red and blue lines do not clearly cross each other, which implies that there is indeed no interaction effect.

Table 13: The relationship between the voice attributes and cognitive trust. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test.

Variable	Cognitive Trust		
	Significance (<i>p</i>)	W-value	Effect Size
Gender of voice			
Female voice	0.0360*	8015.5	0.151
Male voice	0.6365	10205	0.054
Speech rate			
Slow voice	0.1958	8564	0.151
Fast voice	0.2226	9656.5	0.140

.*p*<0.10, **p*<.05, ***p*<.01, ****p*<.001

Table 14: Interaction effect of gender of voice and speech rate on cognitive trust. The results were derived by using data from Treatment Group I using ordinal logistic regression.

Dependent variable:	
Cognitive Trust	
Gender:Speed (interaction effect)	−0.641 (0.497)
Observations	199
Log Likelihood	−691.132

Note: **p*<0.1; ***p*<0.05; ****p*<0.01

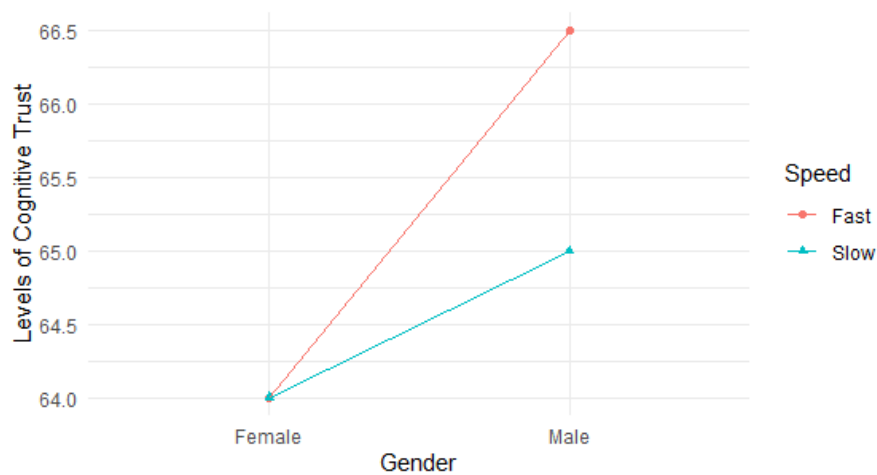


Figure 21: The interaction plot between gender of voice and speech rate on cognitive trust.

The insights on the relationship between vocal communication and cognitive trust could also be expanded by investigating the role of socio-demographic characteristics being part of the research context. Table 15 presents the effect of vocal communication on cognitive trust based on different socio-demographic characteristics. The results obtained using the Mann-Whitney U Test indicate that for participants residing in developed countries, vocal communication exhibited a significant effect on cognitive trust ($W = 11860$, $p = 0.0392$), with a small effect size reflected by Cohen's D of 0.226. Subsequently, ordinal logistic regression analysis revealed a significant negative association between participants living in developed countries and cognitive trust ($\beta = -0.15$, $OR = 0.86$, $SE = 0.21$, $p = 0.046$, $95\%CI = -0.56$ to -0.25). However, no significant influence of vocal communication on cognitive trust was observed for participants residing in developing countries ($p > 0.05$). Furthermore, the gender

of participants did not play a significant role in the relationship between vocal communication and cognitive trust ($p > 0.05$).

Table 15: The relationship between vocal communication and cognitive trust, in which participants are stratified into country of residence or gender. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test.

Variable	Cognitive Trust		
	Significance (p)	W-value	Effect Size
Treatment Group I			
Developed country	0.0392*	11860	0.226
Developing country	0.8149	6108	0.029
Female participant	0.3237	9156.5	0.114
Male participant	0.1453	8172.5	0.171

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Hypothesis 2: Users with depression who interact with a chatbot that has vocal communication capabilities are more likely to have a higher perceived affective trust.

Hypothesis 2 is about the relationship between vocal communication and affective trust. As such, in order to find relationships that could address hypothesis 2, the direct relationship was measured of vocal communication on affective trust. In Table 16, the findings of the relationship between vocal communication and affective trust are displayed. As can be seen, for vocal communication no significant relationship was found with affective trust ($p > 0.05$).

Table 16: The influence of vocal communication on affective trust. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test.

Situation	Affective Trust (Benevolence)		
	Significance (p)	W-value	Effect Size
Voice messages	0.46130	19148	0.074

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

The insights on the relationship between vocal communication and affective trust could be expanded by investigating other factors that could contribute to this relationship. In Table 17 it can be seen how the voice attributes gender of voice and speech rate could influence affective trust, using Mann-Whitney U Test as the analysis tool. As such, it could be observed that gender of voice and speech rate had no significant effect on affective trust ($p > 0.05$).

Table 17: The relationship between the voice attributes and affective trust. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test.

Variable	Affective Trust		
	Significance (p)	W-value	Effect Size
Gender of voice			
Female voice	0.2477	8657.5	0.135
Male voice	0.9333	10490	0.010
Speech rate			
Slow voice	0.2892	8723	0.124
Fast voice	0.9967	10549	0.001

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

The interaction effect of gender of voice and speech rate on affective trust was also explored. As such, the result of the interaction effect is presented in Table 18. As can be observed, no significant interaction was found ($\beta = -0.587$, $p > 0.05$). In Figure 22 the interaction effect of gender of voice and speech rate on affective trust is visually presented. Here, it can be seen that the red and blue lines do not clearly cross each other, which implies that there is indeed no interaction effect.

Table 18: Interaction effect of gender of voice and speech rate on affective trust. The results were derived by using data from Treatment Group I using ordinal logistic regression.

<i>Dependent variable:</i>	
Affective Trust	
Gender:Speed (interaction effect)	-0.587 (0.496)
Observations	199
Log Likelihood	-861.106

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

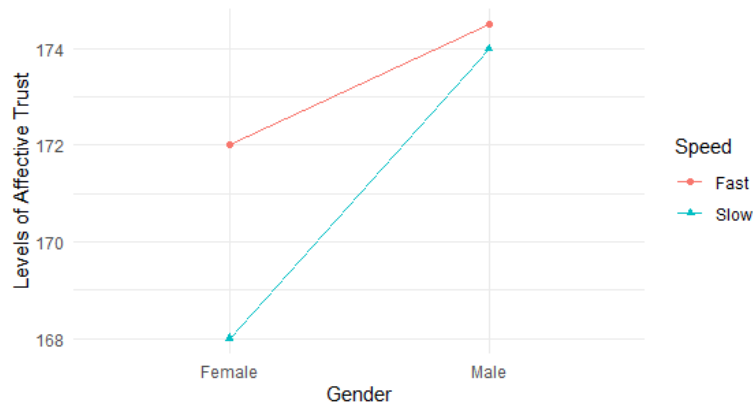


Figure 22: The interaction plot between gender of voice and speech rate on affective trust.

The insights on the relationship between vocal communication and affective trust could also be expanded by investigating the role of socio-demographic characteristics being part of the research context. Table 19 presents the effect of vocal communication on affective trust based on different socio-demographic characteristic. Using Mann-Whitney U Test as the analysis tool, it could be observed that participants living in a developed or developing country had no significant effect on affective trust ($p > 0.05$). Likewise, participants being male or female had no significant effect on affective trust ($p > 0.05$).

Table 19: The relationship between vocal communication and affective trust, in which participants are stratified into country of residence or gender. The results were derived by comparing data from the control group and Treatment Group I using Mann-Whitney U Test.

Variable	Affective Trust		
	Significance (p)	W-value	Effect Size
Treatment Group I			
Developed country	0.1210	12307	0.170
Developing country	0.1661	5505.5	0.172
Female participant	0.9353	9791.5	0.009
Male participant	0.4212	8607.5	0.094

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Hypothesis 3: Users with depression who interact with a chatbot that has personalized vocal communication capabilities are more likely to have a higher perceived cognitive trust.

Hypothesis 3 aims to investigate the connection between personalized vocal communication and cognitive trust. To explore this relationship, a direct analysis of the impact of personalized vocal communication on cognitive trust was conducted. The outcomes of this examination, specifically focusing on the sub-components of cognitive trust (ability and predictability), are presented in Table 20. The results demonstrate that personalized vocal communication significantly influenced the levels of ability ($W = 17675$, $p = 0.02997$) with a small effect size (Cohen's $D = 0.218$). Ordinal logistic regression confirmed a significant negative association between personalized vocal communication and ability ($\beta = -0.38$, $OR = 0.69$, $SE = 0.17$, $p = 0.0296$, $95\%CI = -0.72$ to -0.04). However, no significant relationship was observed for predictability ($p > 0.05$).

To provide a comprehensive understanding of the role of cognitive trust in this context, the combined impact of ability and predictability, referred to as cognitive trust, was further analyzed and presented in Table 21. The results indicated a significant effect of personalized vocal communication on cognitive trust ($W = 17718$, $p = 0.03295$), with a small effect size (Cohen's $D = 0.214$). Ordinal logistic regression supported the significant negative association between personalized vocal communication and cognitive trust ($\beta = -0.372$, $OR = 0.69$, $SE = 0.17$, $p = 0.0325$, $95\%CI = -0.71$ to -0.03). This association is visually depicted in Figure 23.

Table 20: The influence of personalized vocal communication on ability and predictability. The results were derived by comparing data from the control group and Treatment Group II using Mann-Whitney U Test.

Situation	Ability			Predictability		
	Significance (p)	W-value	Effect Size	Significance (p)	W-value	Effect Size
Personalized voice	0.02997*	17675	0.218	0.15230	18546	0.142

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 21: The influence of personalized vocal communication on cognitive trust. The results were derived by comparing data from the control group and Treatment Group II using Mann-Whitney U Test.

Situation	Cognitive Trust (Ability + Predictability)		
	Significance (p)	W-value	Effect Size
Personalized voice	0.03295*	17718	0.214

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

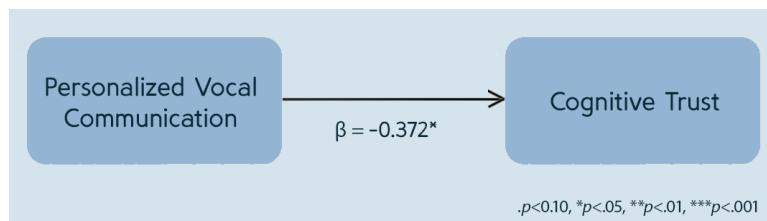


Figure 23: The association of personalized vocal communication on cognitive trust.

The investigation into the relationship between personalized vocal communication and cognitive trust can be further enriched by investigating the role of socio-demographic characteristics being part of the research context.

Table 22 presents the effect of personalized vocal communication on cognitive trust across various socio-demographic characteristics. The results obtained through the Mann-Whitney U Test revealed that when participants resided in developed countries, personalized vocal communication had a significant effect on cognitive trust ($W = 11275$, $p = 0.0018$), with a medium effect size (Cohen's $D = 0.343$). Subsequently, employing ordinal logistic regression, it was found that participants residing in developed countries were significantly negatively associated with

cognitive trust ($\beta = -0.35$, $OR = 0.71$, $SE = 0.21$, $p = 0.044$, $95\%CI = -0.75$ to -0.06). However, personalized vocal communication did not exert a significant influence on cognitive trust when participants resided in developing countries ($p > 0.05$). Likewise, participants being female had no significant role in the relationship between vocal communication and cognitive trust ($p > 0.05$). However, being male did, as it had a significant effect on cognitive trust ($W = 8106$, $p = 0.013$). A Cohen's D of 0.29 indicates a small effect size. When looking at the association relationship, it was found that there was a significant negative relationship present with cognitive trust for being male ($\beta = -0.35$, $OR = 0.71$, $SE = 0.21$, $p = 0.043$, $95\%CI = -0.75$ to -0.06).

Table 22: The relationship between personalized vocal communication and cognitive trust, in which participants are stratified into country of residence or gender. The results were derived by comparing data from the control group and Treatment Group II using Mann-Whitney U Test.

Variable	Cognitive Trust		
	Significance (p)	W-value	Effect Size
Treatment Group II			
Developed country	0.0018**	11275	0.343
Developing country	0.5479	5624	0.075
Female participant	0.6253	9211.5	0.057
Male participant	0.0130*	8106	0.290

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Hypothesis 4: Users with depression who interact with a chatbot that has personalized vocal communication capabilities are more likely to have a higher perceived affective trust.

Hypothesis 4 is about the relationship between personalized vocal communication and affective trust. As such, in order to find relationships that could address hypothesis 4, the direct relationship was measured of personalized vocal communication on affective trust. In Table 23, the findings of the relationship between personalized vocal communication and affective trust are displayed. As can be seen, for personalized vocal communication no significant relationship was found with affective trust ($p > 0.05$).

Table 23: The influence of personalized vocal communication on affective trust. The results were derived by comparing data from the control group and Treatment Group II using Mann-Whitney U Test.

Situation	Affective Trust (Benevolence)		
	Significance (p)	W-value	Effect Size
Personalized voice	0.19060	18676	0.131

Table 24: The relationship between personalized vocal communication and affective trust, in which participants are stratified into country of residence or gender. The results were derived by comparing data from the control group and Treatment Group II using Mann-Whitney U Test.

Variable	Affective Trust		
	Significance (p)	W-value	Effect Size
Treatment Group II			
Developed country	0.0258*	12074	0.243
Developing country	0.2567	5353	0.141
Female participant	0.8774	9441	0.018
Male participant	0.1344	8798	0.174

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

The insights on the relationship between personalized vocal communication and affective trust could also be expanded by investigating the role of socio-demographic characteristics being part of the research context. In Table 24 the effect of personalized vocal communication on affective trust is displayed per socio-demographic characteristic. Using Mann-Whitney U Test as the analysis tool, it could be observed that participants living in a developed country had significant effect on affective trust ($W = 12074$, $p = 0.0258$). The Cohen's D of 0.243 indicates a small effect size. When using ordinal logistic regression, results showed that living in a developing country was significantly negatively associated with affective trust ($\beta = -0.15$, $OR = 0.86$, $SE = 0.21$, $p = 0.0353$, $95\%CI = -0.55$ to -0.25). However, living in a developing country had no significant influence on affective trust ($p > 0.05$). Likewise, the gender of participants had no significant role in the relationship between personalized vocal communication and affective trust ($p > 0.05$).

Hypothesis 5: Depressed users with high perceived cognitive trust in the chatbot are more likely to have intention to use the chatbot.

Hypothesis 5 is about the relationship between cognitive trust and intention to use. In this study, the direct effect of cognitive trust on intention to use was not measured. Instead, mediation analysis is required to, not only investigate the relationship between cognitive trust and intention to use, but also study cognitive trust as the mediator between (personalized) vocal communication and intention to use. In Table 25, the findings of conducting mediation analysis is shown. It can be seen that cognitive trust was significantly positively associated with intention to use ($\beta = 0.31$, $OR = 1.36$, $SE = 0.02$, $p = <2.00E-16$, $95\%CI = 0.27$ to 0.34) for participants who used a chatbot with vocal communication. However, since vocal communication was not significantly associated with cognitive trust, it was identified that cognitive trust was not a mediator between vocal communication and intention to use.

When looking at participants who interacted with a chatbot that has personalized vocal communication capabilities, it can be seen that cognitive trust was significantly positively associated with intention to use ($\beta = 0.31$, $OR = 1.36$, $SE = 0.02$, $p = <2.00E-16$, $95\%CI = 0.28$ to 0.34). In addition, it could also be observed that personalized vocal communication was negatively associated with cognitive trust ($\beta = -0.37$, $OR = 0.69$, $SE = 0.17$, $p = 3.25E-02$, $95\%CI = -0.71$ to -0.03). The presence of these two significant relationships was enough to identify cognitive trust as a mediator, in which cognitive trust mediates indirectly the effect of personalized vocal communication on intention to use. In Figure 24 this relationship could be seen.

Table 25: Mediation Analysis: Overview of the individual associations of voice or personalized voice (IV), cognitive trust (M) and intention to use (DV).

Statistic	β	OR	SE	p	95%CI
Voice					
IV \rightarrow DV	0.08	1.08	0.17	6.67E-01	-0.27 to 0.42
Cognitive Trust					
IV \rightarrow M	-0.27	0.77	0.17	0.123	-0.61 to -0.07
M \rightarrow DV	0.31	1.36	0.02	<2.00E-16***	0.27 to 0.34
Personalized Voice					
IV \rightarrow DV	-0.19	0.83	0.17	2.69E-01	-0.53 to 0.15
Cognitive Trust					
IV \rightarrow M	-0.37	0.69	0.17	3.25E-02**	-0.71 to -0.03
M \rightarrow DV	0.31	1.36	0.02	<2.00E-16***	0.28 to 0.34

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

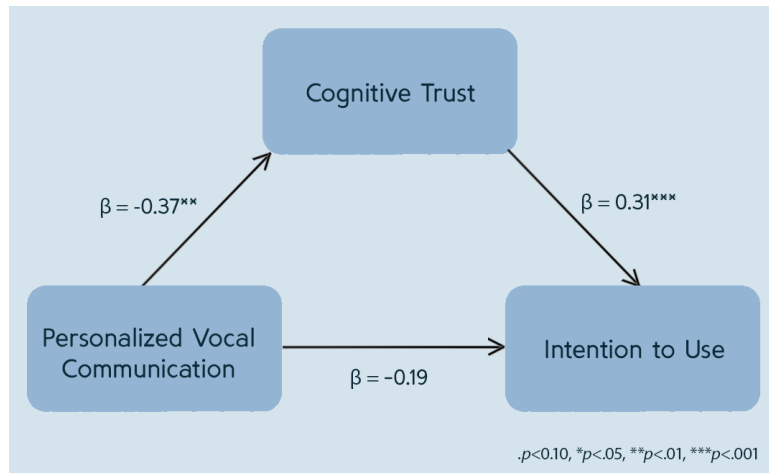


Figure 24: The relationship between personalized vocal communication, cognitive trust and intention to use, in which cognitive trust mediates indirectly the effect of personalized vocal communication on intention to use.

Hypothesis 6: Depressed users with high perceived affective trust in the chatbot are more likely to have intention to use the chatbot.

Hypothesis 6 is about the relationship between affective trust and intention to use. In this study, the direct effect of affective trust on intention to use was not measured. Instead, mediation analysis is required to, not only investigate the relationship between affective trust and intention to use, but also study affective trust as the mediator between (personalized) vocal communication and intention to use. In Table 26, the findings of conducting mediation analysis is shown. It can be seen that affective trust was significantly positively associated with intention to use ($\beta = 0.10$, $OR = 1.10$, $SE = 0.01$, $p = <2.00E-16$, $95\%CI = 0.09$ to 0.11) for participants who used a chatbot with vocal communication. However, since vocal communication was not significantly associated with affective trust, it was identified that affective trust was not a mediator between vocal communication and intention to use.

When looking at participants who interacted with a chatbot that has personalized vocal communication capabilities, it can be seen that affective trust was significantly positively associated with intention to use ($\beta = 0.09$, $OR = 1.10$, $SE = 0.01$, $p = <2.00E-16$, $95\%CI = 0.08$ to 0.11). However, since personalized vocal communication was not significantly associated with affective trust, it was identified that affective trust was not a mediator between personalized vocal communication and intention to use.

Table 26: Mediation Analysis: Overview of the individual associations of voice or personalized voice (IV), affective trust (M) and intention to use (DV).

Statistic	β	OR	SE	p	95%CI
Voice					
IV \rightarrow DV	0.08	1.08	0.17	6.67E-01	-0.27 to 0.42
Affective Trust					
IV \rightarrow M	-0.11	0.90	0.17	5.29E-01	-0.45 to 0.23
M \rightarrow DV	0.10	1.10	0.01	<2.00E-16***	0.09 to 0.11
Personalized Voice					
IV \rightarrow DV	-0.19	0.83	0.17	2.69E-01	-0.53 to 0.15
Affective Trust					
IV \rightarrow M	-0.21	0.81	0.17	2.23E-01	-0.55 to -0.13
M \rightarrow DV	0.09	1.10	0.01	<2.00E-16***	0.08 to 0.11

.p<0.10, *p<.05, **p<.01, ***p<.001

Additional findings to support answering the sub-questions

Thus far in this chapter, we have presented the findings that directly address all the hypotheses formulated in this study. These findings play a significant role in addressing the sub-questions and ultimately contribute to answering the main research question. In this section, additional insights are presented that could support answering the sub-questions (and thus address the main research question.)

Table 27 presents the individual effects of vocal communication on various outcomes, including overall trust, intention to use, anthropomorphism, and vocal communication preference. Notably, only vocal communication exhibited a significant influence on vocal communication preference ($W = 3093.5$, $p = 1.271\text{E-}04$), with a medium effect size indicated by a Cohen's D of 0.575. Employing ordinal logistic regression, the analysis revealed a significant positive association between vocal communication and vocal communication preference ($\beta = 1.00$, $OR = 2.72$, $SE = 0.26$, $p = 0.000123$, $95\%CI = 0.49$ to 1.52 , $95\%CI - OR = 1.64$ to 4.55).

The individual effects of personalized vocal communication on overall trust, intention to use, anthropomorphism, and vocal communication preference were further presented in Table 27. From all these dependent variables, only personalized vocal communication demonstrated a significant effect on vocal communication preference ($W = 2054.5$, $p = 2.297\text{E-}09$), indicating a large effect size as evidenced by a Cohen's D of 0.979. Employing ordinal logistic regression, the analysis revealed a significant positive association between personalized vocal communication and vocal communication preference ($\beta = 1.70$, $OR = 5.45$, $SE = 0.28$, $p = 2.11\text{E-}09$, $95\%CI = 1.15$ to 2.26 , $95\%CI - OR = 3.15$ to 9.58).

Table 27: The influence of (personalized) vocal communication on overall trust, intention to use, anthropomorphism and vocal communication preference. The results were derived by comparing data from the control group and Treatment Group I, and derived by comparing data from the control group and Treatment Group II. Here, the Mann-Whitney U Test was used.

Situation	Significance (<i>p</i>)	W-value	Effect Size
Voice messages			
Overall Trust	0.31740	18849	0.100
Intention to Use	0.66760	19504	0.043
Anthropomorphism	0.59420	19385	0.053
Vocal communication preference	1.271E-04***	3093.5	0.575
Personalized voice			
Overall Trust	0.21550	18765	0.123
Intention to Use	0.27030	18920	0.110
Anthropomorphism	0.62230	19628	0.049
Vocal communication preference	2.297E-09***	2054.5	0.979

. $p < 0.10$, * $p < .05$, ** $p < .01$, *** $p < .001$

The previous sections have primarily discussed the comparisons between the treatment groups and the control group. However, it is also insightful to compare the results between the two treatment groups, namely vocal communication and personalized vocal communication. Table 28 presents the differences between these two groups across various dimensions, including trust, anthropomorphism, overall trust, intention to use, and vocal communication preference. The analysis revealed a significant positive association in vocal communication preference for personalized vocal communication. When using ordinal logistic regression, a significant positive association was found ($\beta = 0.63$, $OR = 1.89$, $SE = 0.26$, $p = 0.013$, $95\%CI = 0.14$ to 1.14 , $95\%CI - OR = 1.15$ to 3.12). This indicates that participants showed a preference for personalized vocal communication over non-personalized vocal communication.

Table 28: The difference between vocal communication and personalized vocal communication in chatbots, in terms of dimensions of trust, anthropomorphism, overall trust, intention to use and vocal communication preference.

	Significance (<i>p</i>)	W-value	Effect Size
Affective Trust	0.5911	19378	0.054
Cognitive Trust	0.5763	19354	0.056
Benevolence	0.5911	19378	0.054
Ability	0.7477	19628	0.032
Predictability	0.2957	18800	0.104
Anthropomorphism	0.8736	19816	0.016
Overall Trust	0.8432	19772	0.020
Intention to Use	0.1284	18247	0.152
Vocal Communication Preference	0.0133*	3652.5	0.361

.*p*<0.10, **p*<.05, ***p*<.01, ****p*<.001

The examination of socio-demographic characteristics being part of the research context can be further expanded by exploring their relationship with vocal communication and intention to use. Table 29 presents the findings regarding this relationship, with a specific focus on country of residence and gender. It was observed that participants residing in developing countries had a significant influence on intention to use ($W = 4614$, $p = 0.002$), indicating a medium effect size (Cohen's $D = 0.338$). Ordinal logistic regression revealed a positive association between living in a developing country and intention to use ($\beta = 0.07$, $OR = 1.07$, $SE = 0.21$, $p = 0.0356$, $95\%CI = -0.33$ to -0.09). However, no significant influence on intention to use was found for participants living in developed countries ($p > 0.05$). Similarly, the gender of participants did not play a significant role in the relationship between vocal communication and intention to use ($p > 0.05$).

The examination of country of residence and gender as part of the research context also includes an exploration of their relationship with personalized vocal communication and intention to use, as detailed in Table 29. Notably, participants residing in developed countries were found to significantly influence intention to use ($W = 12053$, $p = 0.024$), indicating a small effect size (Cohen's $D = 0.246$). Through ordinal logistic regression, living in a developed country demonstrated a negative association with intention to use ($\beta = -0.13$, $OR = 0.88$, $SE = 0.21$, $p = 0.041$, $95\%CI = -0.53$ to -0.29). Conversely, living in a developing country had no significant influence on intention to use ($p > 0.05$). Furthermore, the gender of participants did not play a significant role in the relationship between personalized vocal communication and intention to use ($p > 0.05$).

Table 29: The relationship between (personalized) vocal communication and intention to use, in which participants are stratified into country of residence or gender. The results were derived by comparing data from the control group and Treatment Group I, and derived by comparing data from the control group and Treatment Group II. The Mann-Whitney U Test was used.

Variable	Intention to Use		
	Significance (<i>p</i>)	W-value	Effect Size
Treatment Group I			
Developed country	0.2377	12634	0.129
Developing country	0.0020**	4614	0.338
Female participant	0.4456	9314.5	0.088
Male participant	0.8872	9050.5	0.017
Treatment Group II			
Developed country	0.0240*	12053	0.246
Developing country	0.1188	5139	0.194
Female participant	0.9245	9482	0.011
Male participant	0.2117	8974.5	0.145

.*p*<0.10, **p*<.05, ***p*<.01, ****p*<.001

The preceding sections have explored the role of socio-demographic characteristics being part of the research context, and its relationship with (personalized) vocal communication, various dimensions of trust and intention to use. However, it is insightful to further investigate the impact of socio-demographic characteristics on these dimensions of trust and intention to use within each research group. Table 30 presents the findings of these relationships specifically related to affective trust and cognitive trust.

Notably, the country classification exhibited a significant influence on affective trust and cognitive trust for the control group. Through ordinal logistic regression, it was revealed that participants from developing countries exhibited notably higher levels of affective trust and cognitive trust. Similarly, a parallel trend was observed for participants assigned to the chatbot with personalized vocal communication (Treatment Group II). However, among participants assigned to the chatbot with vocal communication (Treatment Group I), only affective trust was significantly influenced by differences in country classification. The ordinal logistic regression analysis indicated that residing in a developing country significantly increased levels of affective trust. In contrast, gender differences did not yield a significant impact on affective trust and cognitive trust in either of the research groups.

Table 30: The influence of the difference in country classification and gender on affective trust and cognitive trust within a research group.

Variable	Affective Trust			Cognitive Trust		
	Significance (<i>p</i>)	W-value	Effect Size	Significance (<i>p</i>)	W-value	Effect Size
Control Group						
Country classification	0.00162**	2835	0.457	0.002906**	2899.5	0.431
Gender	0.7834	4453.5	0.04	0.8854	4503.5	0.021
Treatment Group I						
Country classification	0.01505*	3306.5	0.351	0.09649	3594.5	0.238
Gender	0.536	4226	0.09	0.6392	4282.5	0.068
Treatment Group II						
Country classification	0.01077*	3183.5	0.368	0.006388**	3118.5	0.394
Gender	0.2419	4200.5	0.169	0.07742	3970	0.256

.*p*<0.10, **p*<.05, ***p*<.01, ****p*<.001

The impact of the difference in country classification and gender on intention to use was also subjected to examination. The findings are presented in Table 31. It is evident that the variance in country classification wielded a significant influence on intention to use across all research groups. Utilizing ordinal logistic regression, it was demonstrated that participants residing in developing countries exhibited significantly higher levels of intention to use. Conversely, gender differences did not yield a significant impact on intention to use within any of the research groups.

Table 31: The influence of the difference in country classification and gender on intention to use within a research group.

Variable	Intention to Use		
	Significance (<i>p</i>)	W-value	Effect Size
Control Group			
Country classification	0.000111***	2578.5	0.567
Gender	0.9372	4528.5	0.012
Treatment Group I			
Country classification	0.000123***	2783.5	0.566
Gender	0.4344	4165.5	0.114
Treatment Group II			
Country classification	0.004565***	3080	0.410
Gender	0.3481	4291.5	0.135

.*p*<0.10, **p*<.05, ***p*<.01, ****p*<.001

A total of 215 participants had given open feedback. From the open feedback, several themes were identified in regards to all the chatbots used in this study. The first theme was chatbot competence. From all the participants who had given open feedback, more than half of them pointed out that the chatbot was not cognitive competent. In other words, participants felt not heard or they felt that the chatbot was not replying on a personalized matter:

"It felt like it was not actually reading what i wrote, rather it felt like a lecture. useful advice but it did not feel personal enough to satisfy me" [PT14]¹

When looking it at on the level of the research groups, it was found that this was equally true for the chatbots used in the control group and all the treatment groups. Furthermore, several participants compared the chatbots used in this study with ChatGPT, in which they pointed out that they expected that the chatbots in this study should be at least as cognitive competent as ChatGPT:

"After engaging in numerous conversations with ChatGPT, I found this chatbot to be lifeless and somewhat unhelpful. It didn't truly 'listen' to my concerns, problems, and needs." [PT8]

The second identified theme was technical performance. Here, several participants pointed out that the messages sent were too long:

"It could've given a bit less voice messages, some times it was too many." [PT419]

Another technical problem that participants pointed out was that they would prefer if the voice messages would play automatically, as currently participants were required to manually press on each voice message to play the audio:

"Voice messages could play automatically in order for a better experience." [PT397]

The third identified theme was voice characteristics. Several participants have pointed out that the voices could feel unnatural. Things like voices are too robotic or voices seem artificial were common trends:

"The voice seems very "robotic", it would be better to have a more sophisticated text-to-speech bot" [PT78]

¹PT('number') represents the unique ID of the participant of this study.

6 Discussion

In this chapter, the key findings from this study are reflected upon. Additionally, the theoretical, managerial, and ethical implications of the study's findings are explained, shedding light on their broader significance and potential impact. Moreover, the limitations of the research are addressed, highlighting any constraints or shortcomings that might have influenced the study's outcomes. As a result, recommendations for future research directions are also provided to address the gaps and extend the knowledge in this domain.

6.1 Reflections on the Research Questions

Following the design science approach for knowledge questions of [Wieringa \(2014\)](#), the sub-questions of this study are answered and discussed. This enables to address the main research question.

SQ1: What does the use of voice and its attributes do to the different trust dimensions on chatbots?

Many studies have shown how different features or aspects of conversational agents could change the level of anthropomorphism and how this could influence the perceived trust of users ([Følstad et al., 2018](#), [Hsiao and Chen, 2021](#)), which in turn, plays an important role in adoption ([Melián-González et al. \(2021\)](#), [Ng et al. \(2020\)](#)). For example, it was shown that anthropomorphism could increase trust in smart healthcare services like health chatbots ([Liu and Tao, 2022](#)). In this study, the influence of chatbots that were able to speak to its users (vocal communication) was tested on trustworthiness. Although other studies have shown how the voice used in voice assistants could influence perception of trust ([Elkins and Derrick, 2013](#), [Tolmeijer et al., 2021](#)), this has never been studied with chatbots, especially not with chatbots used to support people who experience depression.

According to [Laurel \(1997\)](#), anthropomorphism in conversational agents could direct the user's attention to the competency of the chatbot, which relates to the user's confidence in the chatbot's ability to perform its tasks due to human-like qualities. Another study has shown that anthropomorphism of intelligent personal assistants could induce cognitive trust ([Chen and Park, 2021](#)). In this study, giving the chatbot a voice was hypothesized to increase anthropomorphism, and thus cognitive trust could be affected by the voice. However, hypothesis 1 (*Users with depression who interact with a chatbot that has vocal communication capabilities are more likely to have a higher perceived cognitive trust.*) was rejected since no evidence was found for this relationship. In other words, cognitive trust does not significantly change by vocal communication. This could be explained by the fact that no significant difference in levels of anthropomorphism was measured to begin with, which means that the voices used in this study were not able to make a difference in the perception of human-likeness of its chatbot compared to chatbots that only communicate through text messages.

In this study, it was shown that chatbots that used voice messages as a mode of communication were not able to cause any significant changes in affective trust, which rejects hypothesis 2 (*Users with depression who interact with a chatbot that has vocal communication capabilities are more likely to have a higher perceived affective trust.*). In other words, the confidence of the user based on feelings was not different between this type of chatbot and the chatbot that uses text messages to communicate (the control group). This contradicted a previous study that showed, in the online setting, communication through voice increased trust more than through text messages ([Williams et al., 2007](#)). However, in that particular study, data of participants were collected three times over a period of a month, which indicates that time played a crucial role in building trust. According to [Strauss \(1978\)](#) trust is generated over time, which was also shown for building trust with one's health provider ([Khullar, 2019](#)) and building trust with voice assistants ([Elkins and Derrick, 2013](#)). Since, in this study, participants spend an average of 4 minutes with the chatbot, it is possible that time was too limited to build trust. Therefore, lack of bonding time could have resulted in no increased levels of anthropomorphism for chatbots with vocal communication compared to the control group. If more time was given, it could have also increased cognitive trust due to increased levels of anthropomorphism, although this is something that requires further studies. In addition, since the participants experienced depression, it could be that in general, it is hard for them to connect on an emotional level ([Achterbergh et al., 2020](#)), whether that is with a regular chatbot or a chatbot with a voice. This could also explain why no difference in affective trust could be observed.

From exploratory analysis, it was found that living in a developing country was able to increase affective and cognitive trust when participants were interacting with the chatbot with vocal communication capabilities. This could be explained by the idea that people from developing countries might be more open to technology adoption; at least this was the case in different technological contexts ([Takahashi et al., 2020](#)). However, no evidence was found that

living in a developing country could make a difference for chatbot adoption between text-based and voice-based communication.

SQ2: What does the personalization of the voice do to the different trust dimensions on chatbots?

Personalization of chatbots within the context of users experiencing depression could be regarded in two ways: personalization of appearance/presentation (Blom and Monk, 2003) and personalization in terms of machine intelligence (Lipschitz et al., 2019). In this study, voice is regarded as a form of presentation of the chatbot. However, since no research was done before on the effects of chatbots with vocal communication, the influence of personalization of the voice given to chatbots was tested on trustworthiness.

Since chatbots with a voice, randomly allocated to participants, did not change levels of affective trust, it was no surprise to find no significant changes in affective trust by introducing the option to personalize the voice, as vocal communication already showed to have no effect on affective trust to begin with. This resulted in the rejection of hypothesis 4 (*Users with depression who interact with a chatbot that has personalized vocal communication capabilities are more likely to have a higher perceived affective trust.*). However, with no changes in affective trust, it was also unclear if participants (who experienced depression) could have found it uncomfortable to make a decision about which voice they wanted to have for the chatbot. People with depression usually have a harder time making decisions in general (Bishop and Gagne, 2018). Furthermore, previous studies have shown that personalization induces behavior to select aspects that are familiar, increasing the emotional state of the user (Song and Shin, 2022). Perhaps the personalization options in this study were too limited to find familiarity within the voices.

This study has shown that cognitive trust significantly decreases when given the option to personalize the voice of the chatbot, resulting in the rejection of hypothesis 3 (*Users with depression who interact with a chatbot that has personalized vocal communication capabilities are more likely to have a higher perceived cognitive trust.*). More specifically, it was the ability part of cognitive trust that had the biggest contribution to this relationship. There were several reasons that could explain this result. One of which is the fact that the chatbot was not able to provide true personalized answers but rather was only able to provide pre-recorded answers following a rigid script. Interestingly, this is in line with the literature, which suggested that when chatbots receive negative reviews from users, the complaints include topics about the lack of creative responses (Ahmed et al., 2022). According to Ma and Liu (2020), users of voice assistants might have higher expectations of its capabilities than with text-based search. However, if these expectations were not met, users were less forgiving and lose trust in its ability. As such, when users were able to choose a voice in this study for the chatbot, and thus build up the expectation of advanced interactions to come, it could be a huge loss of confidence when the chatbot does not deliver by responding in an advanced manner. It might also be possible that with the introduction of ChatGPT on the market, expectations in general were higher for chatbots. In fact, participants in this study have consistently given feedback that they compared the chatbot from this study with the performance of ChatGPT, and expected the chatbot in this study to at least perform as cognitively. This was in line with literature, which suggested personalization in terms of machine intelligence was crucial for adoption of chatbots for mental health (Liu and Tao, 2022). Another reason for lower levels of cognitive trust is that participants who had chosen a desired voice might not have experienced the chosen voice as expected when actually interacting with the chatbot, which refers back to the expectation problem and why cognitive trust could decrease. In fact, in this study, it was observed that the female voice was significantly negatively associated with cognitive trust. Although the effect size was small (<0.2), it might have contributed to the expectation problem. Interestingly, it were specifically male participants who experienced significantly lower cognitive trust with chatbots providing personalized vocal communication. Since certain negative stereotypes exist towards gender roles of women that could surface by hearing a female voice (West et al., 2019), the use of a female voice could have contributed to lack of confidence, although more research is required. However, if this is true, then this could open discussion about the ethical implications of using gender-specific voice-based chatbots since reinforcement of negative stereotyping could occur (West et al., 2019).

From exploratory analysis, it was found that lower levels of affective trust and cognitive trust were significantly associated with participants who were living in developed countries. These relationships could be explained by the idea that participants from developed countries might have higher expectations in general for technology due to their economic advances. This also translated into the observation that participants from developed countries have significantly lower intention to use chatbots with personalized vocal communication than participants from developing countries. Moreover, it was also observed that living in a developing country was able to increase affective and cognitive trust when participants were interacting with the chatbot with personalized vocal communication capabilities, compared to participants from developed countries. As mentioned before, this could be explained by

the idea that people from developing countries might be more open to technology adoption; at least this was the case in different technological contexts (Takahashi et al., 2020). However, no evidence was found that living in a developing country could make a difference for chatbot adoption between text-based and personalized voice-based communication.

SQ3: What is the influence of the different dimensions of trust on intention to use of chatbots?

Previous studies have shown how the adoption of technology depends on the perception of trust (Bahmanziari et al., 2003), which translates into intention to use (De Cicco et al., 2022). This has also been shown to be the case in the mental health context (Kasilingam, 2020, Sitthipon et al., 2022, Zhu et al., 2022). Since trust could be divided into affective and cognitive trust, studies were done to show that the two dimensions of trust (affective and cognitive) are required to increase intention to use healthcare-related digital applications (Meng et al., 2022). However, this relationship has never been tested before with chatbots that use (personalized) speech to communicate with its user. As such, this was tested in this study.

Thus far it was found that there were no different levels of affective trust measured for both treatment groups. Similarly, no difference was observed in cognitive trust with vocal communication, although cognitive trust was negatively affected by personalized vocal communication. Overall trust was also directly measured, which resulted in no difference either for both treatment groups. Therefore, it could be suggested that the chatbots from the treatment groups were not able to affect trust levels, or the chatbots were impacting it negatively. If the findings from literature holds, which states that increased levels of trust are required for adoption of chatbots (De Cicco et al., 2022), then it was no surprise that in this study, no direct relationship was found between (personalized) vocal communication and intention to use since no increased levels of trust were measured. With no increase levels of trust, participants (who were all of young age) could also have found it more difficult to increase intent to use the chatbot since young people require feeling emotionally comfortable enough to seek mental help to express their feelings with a potential help provider (Rickwood et al., 2007). However, in this study, it was found that for both treatment groups, affective and cognitive trust were positively associated with intention to use, suggesting that levels of intention to use would increase with higher levels of trust. This was in line with findings from the literature (Meng et al., 2022). In other words, hypothesis 5 (*Depressed users with high perceived cognitive trust in the chatbot are more likely to have intention to use the chatbot.*) and hypothesis 6 (*Depressed users with high perceived affective trust in the chatbot are more likely to have intention to use the chatbot.*) were accepted.

Although it was found that affective trust had no mediating role, cognitive trust might have. It was shown that cognitive trust did not mediate in the total effect between personalized vocal communication of chatbots and its intention to use. However, it was found that cognitive trust significantly contributed to the indirect effect of these two variables. Since no significant direct effect was also found, it was possible that other unknown variables could have mediated the relationship positively between personalized vocal communication and intention to use. This could be explained using mediation theory, which states according to Hayes (2009), "two or more indirect effects with opposite signs can cancel each other out, producing a total effect and perhaps even a total indirect effect that is not detectably different from zero, in spite of the existence of specific indirect effects that are not zero". In other words, future research should consider testing other variables to close this gap.

Through exploratory analysis, it was found that there were no differences in affective trust, cognitive trust, and intention to use between chatbots with vocal communication and chatbots with personalized vocal communication. However, it was found that participants who interacted with the chatbot where the voice could be chosen significantly preferred this type over chatbots in which no personalization of the voice could occur. In fact, for both treatment groups, it was suggested that participants significantly preferred a chatbot that could speak over regular text-based chatbots. This suggests that participants like the idea of chatbots with vocal communication, but it requires building trust to actually use it. Since there is demand for chatbots with personalized vocal communication, there is sufficient ground for further research. Furthermore, when looking at the role of socio-demographic characteristics, it was shown that participants who were living in developing countries were significantly associated with higher level of intention to use than participants from developed countries. However, it was not shown whether country classification could make a difference in intention of use between (personalized) vocal communication and text-based messages. Since these findings were exploratory in nature, more research is required to understand the impact of country classification on chatbot adoption and why this relationship exists. Finally, it seems that speech rate of the voice does not matter for intention to use.

How does vocal communication and its personalization in chatbots affect the intention to use of young adults with depression?

By answering all three sub-questions, it is possible to provide an answer to the main research question. This is achieved by following the problem decomposition outlined in Chapter 1.2.

To begin with, this study revealed that chatbots that were able to communicate through speech were not able to change affective and cognitive trust significantly. This implies that using chatbots with vocal communication does not necessarily change levels of trust in comparison with text-based chatbots. However, since levels of anthropomorphism were not different from the control group, it was suggested, based on literature, that participants in this study might not have enough time to build trust with the chatbot with vocal communication. Additionally, it was proposed that due to the nature of being depressed, affective trust could be harder to increase. These reasons could also explain why affective trust did not change in this study when giving the option to personalize the voice of a chatbot. As such, this study revealed that personalizing the voice does not help with increasing trust based on emotions with the chatbot. However, cognitive trust was negatively affected by personalization. With help from literature, it was suggested that this might be attributed to the fact that the chatbot was not able to reply in a personalized manner, while providing the option to choose a voice gave the impression of advanced interactions. Other factors such as prior experience with ChatGPT and gender bias might have also altered expectations with the chatbot in this study. Specifically the latter could be of ethical concern.

Findings from this study also revealed that cognitive and affective trust were positively associated with intention to use. In other words, this study revealed that a high intention to use chatbots with (personalized) vocal communication required increased levels of trust. Additionally, cognitive trust was able to indirectly mediate the relationship between personalized vocal communication and intention to use. Since no direct mediating effect was found, it was suggested from literature that there might be other variables that could have been mediating this relationship as well.

Finally, it was found from exploratory analysis that users from developed countries were associated with lower cognitive and affective trust than users from developing countries when it comes to personalization of voice. Also, users from developing countries were observed to be associated with higher intention to use chatbots that communicate through speech, although more research is needed to understand why this is the case.

6.2 Reflections of Implications

The findings from this study were able to contribute to existing literature in many areas of research, as well as open ethical discussions. In addition, the findings from this study were valuable for corporations and society.

6.2.1 Theoretical Implications

To begin with, this study was able to expand on the knowledge about the adoption of conversational agents that could provide mental health support. From the literature, it was already known that people who experience depression were open to using these types of chatbots (Rathnayaka et al., 2022). However, personalization in terms of machine intelligence was needed to increase adoption of chatbots for mental health services (Liu and Tao, 2022). What was not clear is how to increase adoption by young adults, since the adoption rate for young adults could be troublesome given that they were more reluctant to seek help for mental health (Bae Brandtzaeg et al., 2021). Therefore, there was a demand to find ways to increase adoption for young adults who experience depression. In this study, the use of speech as a mode of communication of chatbots was addressed, and it was hypothesized that it could make it more emotionally and rationally comfortable for young adults to use. Moreover, previous research usually focused on studying the adoption in developed countries. In fact, not much research has been done about technology acceptance in developing countries (Averweg, 2005), while developing countries were more likely to have less access to healthcare (Kamal et al., 2020). This study was able to expand this area of knowledge by giving more insights about the adoption of chatbots for mental health services within developing countries.

Furthermore, this study was able to expand on the knowledge about trust and its impact on adoption of technology. Previous studies have shown that the right levels of anthropomorphism could increase trust towards chatbots (Følstad et al., 2018, Hsiao and Chen, 2021), which has also been studied in the context of health chatbots (Liu and Tao, 2022). Additionally, the impact of voice assistants on trust has also been studied (Elkins and Derrick, 2013, Leitman et al., 2010, Seymour and Van Kleek, 2021, Tolmeijer et al., 2021). However, what had never been studied before was using voice to increase anthropomorphism levels in order to increase trust in chatbots, and this study provided more insights into this relationship. Likewise, providing personalization options for the voice

and its relationship with trust and chatbot adoption had also never been tested within the context of supporting mental health chatbots. From the literature, it was known that the aspect of personalization could contribute to the adoption of chatbots, which could be about personalization in terms of machine intelligence (Lipschitz et al., 2019) and personalization in terms of appearance/presentation (Blom and Monk, 2003). Although this study was able to gain more insights about personalization in terms of appearance/presentation, results from this study were also able to gain more insights about personalization in terms of machine intelligence. This latter could be observed with the decreased cognitive trust found with personalized vocal communication and how this potentially could be attributed to high expectations of chatbot performance due to presentation of the chatbot.

Finally, this research goes beyond just examining the use of voice and its relationship with trust and intention to use. Previous literature has shown concerns about the bias that could arise with gender-specific voices (Chin et al., 2020, West et al., 2019). This study acknowledges the potential gender-based biases that could arise with selecting gender-specific voices for a chatbot, which means this study was able to contribute to the knowledge around ethics in chatbot technology.

6.2.2 Managerial & Societal Implications

The COVID-19 pandemic has shown that depression and anxiety became the most common mental disorder among young-aged workers (Campbell and Gavett, 2021, Hawes et al., 2022, Rathnayaka et al., 2022). After the COVID-19 pandemic, it was clear that depression cases remained high (Wang et al., 2022). This is problematic for young adults, as they tend to be more reluctant to seek help for mental health issues (Bae Brandtzæg et al., 2021). As a result, the findings from this study could be of help.

Firstly, many companies are developing chatbots for the healthcare sector, not only for mental health applications but also for other purposes like disease prevention and diagnosis, all of which face adoption challenges (Laumer et al., 2019). However, it is important to note that this study only considers users who experience depression, so the findings may not necessary apply to other patient groups. Although patients, in general, could have mental health problems as well due to being a patient (Louvardi et al., 2020). Thus, with the findings of this study, chatbot builders could use the insights of this study to potentially get closer to the ideal design to improve adoption of chatbots for mental healthcare purposes, and potentially also for the broader healthcare sector.

Furthermore, given the shortage of mental health professionals (Darzi and Evans, 2016), more people with mental health problems (like depression) could be supported when more chatbots are available in the future, that only are able to support their problems but also in which the users feel comfortable enough to continue using the chatbot. With its success, this could also relieve work pressure on mental health professionals, as there is a shortage of mental health professionals.

Moreover, once these chatbots are introduced to the market, users from developing countries could also benefit from these chatbots since they are online accessible and developing countries usually have a lack of access to healthcare (Kamal et al., 2020).

In addition, once these mental health chatbots are on the market, this could also contribute to the long-term well-being of society. People with depression usually have a lower quality of life, which could result in economic damages (Cutler and Summers, 2020).

Finally, the findings from this study could also guide the government and health authorities to create policies around the use of chatbots ethically and safely, especially if the chatbot is integrated with artificial intelligence and uses gender-specific voices.

6.2.3 Ethical Implications

Although using chatbots to support people who experience depression could offer many benefits, such as potentially increasing access to mental health professionals, lowering costs, reducing error rates, and improving outcomes for patients, there are ethical concerns that need to be taken into account. With the current state-of-the-art AI-driven chatbot technology, mistakes can still occur, such as being insulting and hurtful to users' emotions, misinterpreting their inquiries, and failing to provide the advantages and services they were promised. When people with depression use such chatbot, this could potentially be more harmful to the user due to the sensitivity of the people the chatbots are intended to help. From a virtuous point of view, these chatbots should not even be used until they are fully capable of interacting in a manner that is ethically right. Similarly, from a Kantian perspective, these chatbots

should be thoroughly tested before they come to the market, because if the chatbot does not speak the truth or misleads, this interferes with one's autonomy and goes against the proposition of the universal law.

Another concern for health chatbots is the issue of privacy. The conversations with users with depression could contain highly personal and sensitive information that needs to be processed by the chatbot. For the chatbot to be able to serve for the long-term, it should be able to store this data as well in order to respond back intelligently and personalized. If this data is leaked or stolen, this raises ethical issue that could harm the user.

A third potential problem to consider with mental health chatbots is when there is unexpected disclosure of crime. Usually, when a patient admits to an act of crime during a therapy session, the human mental health professional is obligated by law to report suspected abuse (Coghlan et al., 2023). However, in a situation where the chatbot is the medical caregiver, this jurisdiction does not hold since the owner of the chatbot is usually a company. Principles of Kantianism suggest that at least an ethical duty of this kind is needed where the revelation is credible.

A fourth potential problem to consider is whether it is ethical to let a patient develop a relationship with a chatbot. However, considering the lack of mental health professionals and the ability to scale the chatbot (assuming that the chatbot functions at least as well as a human), from a utilitarian point of view, the chatbot could bring more benefits to the most number of people.

Insights from this study has also brought to light potential ethical concerns when giving chatbots a voice. From the literature, it was already known that artificial intelligence could bring racial, religious and gender bias due to the limited number of available data sets to train the chatbot (Kundi et al., 2023). However, this problem is likely to also occur the other way around when giving the chatbot a voice, as observed from the findings of this study. Namely, in this study, it was found that male participants had lower cognitive trust after interacting with the chatbot that is able to communicate through speech. Since it was also observed that cognitive trust was lower when the voice of the chatbot is female, this could imply that male participants could have a negative gender-based bias when it comes to hearing a female voice. As a result, reinforcement of negative stereotyping could occur (West et al., 2019). For example, in the long-term, male participants could actually think that human woman are less competent due to their bad experience with the chatbot using a female voice. This mirroring phenomenon was already shown with voice assistants; given that users of voice assistants interact with a commanding tone, this behavior could be mirrored in conversations with women (Chin et al., 2020). However, this ethical concern could be solved by using gender-ambiguous voices (Tolmeijer et al., 2021).

6.3 Limitations & Future Research

As mentioned in Chapter 1.1, it has been shown that the use of AI to detect depression in the user's voice is becoming a reality (Rejaibi et al., 2022), which makes it imaginable that chatbots could use this technique to assess the progress of depression treatment by listening to the voice of users. To make the conversational interaction more user-friendly, it would make sense for chatbots in this context to gain speech-to-text and text-to-speech capabilities. The collected data in this study is in line with this idea since this study has suggested that participants would prefer a chatbot that is able to communicate using a (personalized) voice over text-based communication. This means that there is ground to continue studying chatbots with at least vocal communication capabilities. For example, trustworthiness and intention to use could be tested by using chatbots that can not only communicate through speech but are also able to listen to the user's voice.

Findings and limitations of this study has also inspired further research ideas. To begin with, the external validity was limited since it was a lab experiment. In other words, applying the findings from this study to other contexts and the real-life setting should be done cautiously. To increase external validity, future study could be conducted as field experiments. In a field experiment, a real mental health professional could recommend using the chatbot to its patients who are actually in need of help at the time of the study. Instead of a cross-sectional study, it would become a longitudinal study, allowing for measurements of affective trust, cognitive trust, and intention to use over a longer period of time. Moreover, this would also allow testing if levels of anthropomorphism increase, which could result in higher trust for both chatbots with vocal communication and chatbots with personalized vocal communication.

Furthermore, external validity was also questionable due to the technology used in the chatbot. Since the responses of the chatbot were limited to a pre-defined script, the chatbot does not truly represent how it should interact in a real-life setting (which is more personalized). Future research should consider implementing existing

well-trained AI-driven chatbot technology like OpenAI's ChatGPT or Google's Bard to personalize the responses of the chatbot. If the technology is not available, another method is to capture a video of someone interacting with the chatbot, in which it seems that the chatbot is interacting on a highly personalized matter. In reality, someone else is interacting on behalf of the chatbot, or video post-production could make it seem that the chatbot is interacting naturally and intelligent. Then participants of the study would be able to watch the video to get an idea of the experience, after which the levels of trust and intention to use could be measured. It is expected that, unlike the results from this study, cognitive trust could increase by using chatbots with personalized vocal communication.

In addition, findings from this study suggested that personalized vocal communication had no effect on affective trust. It could be interesting to explore why this could be the case. This study hypothesizes that it could be due to the nature of people with depression, making it harder for them to make decisions in general (Bishop and Gagne, 2018). Therefore, in further research, this aspect needs to be taken into account in the study design. Furthermore, more options to personalize the voice could be given to the participants in future research, since this should increase affective trust due to increased chances of finding familiarity in a voice (Song and Shin, 2022).

Moreover, it was found that cognitive trust was negatively affected by personalized vocal communication. This study suspects that this is due to a letdown of expectations of high performance (Ma and Liu, 2020). To actually understand the reason behind this relationship, future study could investigate whether prior experience with advanced chatbot technologies like ChatGPT could have been the reason. Although, as mentioned before, future research should implement chatbots that could reply in a personalized manner to begin with, it would still be recommended to rule out the influence of prior experience with advanced chatbot technologies like ChatGPT. Another factor to explore is whether there is an increased level of performance expectations based on the personalized voice chosen and whether this could influence cognitive trust. Future research could also explore the role of gender-specific voices on performance expectations of the chatbot, which also relates to the ethical aspects of using chatbots, given gender-specific voices could reinforce negative stereotypes (West et al., 2019).

Since this study found that there could be other mediating factors between personalized vocal communication and intention to use, future research could explore this further.

Also, future research could focus on the influence of demographic features of participants on the two dimensions of trust and intention to use of chatbots that communicates through speech. For example, since this study observed that participants from developing countries were more willing to use chatbots with personalized vocal communication, it could be interesting to find out why. In addition, since this study focused on young adults, in future research, older people could also be tested to see if vocal communication could be of value for them as well. Finally, future study could also compare the effects of trust and intention to use between participants who experience depression and those who are patients in general, but not necessarily depressed. This could improve design guidelines for building chatbots in general for the healthcare sector.

7 Conclusion

Although many young adults experience depression, it remains a challenge to encourage young adults to seek help for mental health. Digitalizing psychotherapies by means of mental health chatbot services could be a potential solution. It was identified that voice could influence trust, which could potentially encourage adoption of chatbots to support people with depression. Therefore, the main research question of this study is formulated as:

"How does vocal communication and its personalization in chatbots affect the intention to use of young adults with depression?"

Since the main research question is a knowledge problem, it was decided to follow the design science methodology of [Wieringa \(2014\)](#), to conduct this study. In this chapter all the key takeaways and recommendations are provided.

7.1 Key Takeaways

This study explored the influence of chatbots with vocal communication capabilities on perceived trust and intention to use among individuals experiencing depression. Despite previous research highlighting the impact of anthropomorphism on trust in conversational agents, the study did not find significant changes in affective and cognitive trust with the introduction of vocal communication in the chatbots. This could have been due to lack of time to build trust with the chatbot, which is something that requires more research.

Personalization of the chatbot's voice was also investigated, but no significant effect on affective trust was observed, possibly due to the limited range of voice options provided that in this case could decrease chances to find familiarity. As such, this requires further research. However, participants who experienced depression displayed a decrease in cognitive trust when given the option to personalize the voice, possibly due to unmet expectations and the inability of the chatbot to provide truly personalized responses. As such, more research is needed to confirm these explanations.

The study confirmed that cognitive and affective trust were positively associated with intention to use, aligning with previous research on the importance of trust for technology adoption. However, no direct relationship between vocal communication and intention to use was found, and mediation analysis suggested the presence of unknown variables impacting this relationship. To fill this knowledge gap, more research is required.

Exploratory analysis revealed a preference for chatbots with (personalized) vocal communication capabilities over text-based communication, indicating a potential demand for such features. Participants from developing countries were associated with higher intention to use of chatbots with vocal communication, while those from developed countries displayed an association of lower cognitive trust. For personalized vocal communication, participants from developed countries were associated with lower cognitive, affective trust and intention to use. Since these findings were exploratory, additional research could provide more insights about the underlying mechanism.

Overall, the findings highlight the complexities of trust perception and intention to use in the context of chatbots for mental health support. Further research is necessary to understand the impact of voice personalization, user's country classification, and other variables on chatbot adoption and to explore ways to enhance trust and utilization of vocal communication in mental health chatbots.

7.2 Recommendations

Companies building chatbots

The findings from this study can be valuable for guiding the development and adoption of chatbots for mental healthcare purposes, particularly for young adults experiencing depression. To maximize the benefits of chatbots in addressing mental health challenges like depression, the following recommendations are suggested:

- **Ideal Design for Adoption:** Chatbot builders in the healthcare sector should leverage the insights from this study to design chatbots that are user-friendly and effective for supporting mental health. Although speech as the mean of communication for chatbots were not able to increase trustworthiness, and thus intention to use, there is still opportunity to further investigate the use of speech by studying the long-term effects on intention to use. In the end, participants of this study had pointed out to have a preference for chatbots that are able to speak than to only send text messages. This also provides more ground to study the effects on

trustworthiness and intention to use with chatbots that are not only able to communicate through speech but also able to listen to the users' voice.

- **Global Accessibility:** The online accessibility of chatbots makes them a potentially valuable resource for people in developing countries where access to healthcare is limited. By introducing these chatbots to the market, not only could individuals in underserved regions can benefit from mental health support, but chatbot builders could commercialize on this market.
- **Ethical Development and Testing:** Chatbot technology should undergo rigorous testing and development before being deployed for users, adhering to virtuous principles and Kantian ethics. Developers must ensure that chatbots are fully capable of interacting ethically and truthfully.

Policy makers

From the findings of this study, it is clear that ethical issues could arise with using chatbots providing health services. Although the current state-of-the-art knowledge is still premature, it is a matter of time until chatbots are ready for the market. Policy makers like the government and health authorities should therefore design, or at least reflect on, policies that address the ethical implications of using chatbots to support people with depression.

While using chatbots to support people with depression can offer many benefits, such as increased access to mental health professionals, cost reduction, lower error rates, and improved patient outcomes, it is essential to address the ethical concerns associated with their use. To ensure responsible implementation and maximize positive impact, the following recommendations are suggested:

- **Privacy Protection:** To maintain user trust and confidentiality, chatbot conversations containing highly personal and sensitive information should be securely processed and stored. Developers must implement robust data protection measures to prevent any data leaks or breaches.
- **Handling Unexpected Disclosures:** To address situations involving unexpected disclosure of crimes during chatbot interactions, developers should consider implementing ethical guidelines. While the chatbot may not have the same legal obligations as human professionals, it should have an ethical duty to handle credible revelations responsibly.
- **Balancing Patient Relationships:** The ethical implications of patients developing relationships with chatbots should be carefully considered. In situations where mental health professionals are scarce, and chatbots can effectively provide support, developers should assess the potential benefits for the majority while ensuring that users' autonomy and well-being are safeguarded.
- **Addressing Bias in Chatbot Voices:** To mitigate biases in chatbot voices, particularly those related to gender, developers should use gender-ambiguous voices when possible. By avoiding gender-specific voices, the risk of reinforcing negative stereotypes can be minimized.

References

- Achterbergh, L., Pitman, A., Birken, M., Pearce, E., Sno, H. and Johnson, S. (2020), 'The experience of loneliness among young people with depression: a qualitative meta-synthesis of the literature', *BMC psychiatry* **20**(1), 1–23.
- Adamopoulou, E. and Moussiades, L. (2020), An overview of chatbot technology, in 'Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II 16', Springer, pp. 373–383.
- Adolphs, R. (2002), 'Trust in the brain', *Nature neuroscience* **5**(3), 192–193.
- Ahmed, A., Aziz, S., Khalifa, M., Shah, U., Hassan, A., Abd-Alrazaq, A. and Househ, M. (2022), 'Thematic analysis on user reviews for depression and anxiety chatbot apps: Machine learning approach', *JMIR Formative Research* **6**(3), e27654.
- Al Shamsi, J. H., Al-Emran, M. and Shaalan, K. (2022), 'Understanding key drivers affecting students' use of artificial intelligence-based voice assistants', *Education and Information Technologies* **27**(6), 8071–8091.
- Alqahtani, F. and Orji, R. (2020), 'Insights from user reviews to improve mental health apps', *Health informatics journal* **26**(3), 2042–2066.
- Aoki, N. (2020), 'An experimental study of public trust in ai chatbots in the public sector', *Government Information Quarterly* **37**(4), 101490.
- Araujo, T. (2018), 'Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions', *Computers in Human Behavior* **85**, 183–189.
- Averweg, U. R. (2005), 'Applicability of the technology acceptance model in three developing countries: Saudi arabia, malaysia and south africa', *Alternation* **12**(1a), 331–343.
- Bae Brandtzæg, P. B., Skjuve, M., Kristoffer Dysthe, K. K. and Følstad, A. (2021), When the social becomes non-human: Young people's perception of social support in chatbots, in 'Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems', pp. 1–13.
- Bahmanziari, T., Pearson, J. M. and Crosby, L. (2003), 'Is trust important in technology adoption? a policy capturing approach', *Journal of Computer Information Systems* **43**(4), 46–54.
- BarNir, A., Watson, W. E. and Hutchins, H. M. (2011), 'Mediation and moderated mediation in the relationship among role models, self-efficacy, entrepreneurial career intention, and gender', *Journal of applied social psychology* **41**(2), 270–297.
- Bartneck, C., Kulić, D., Croft, E. and Zoghbi, S. (2009), 'Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots', *International journal of social robotics* **1**, 71–81.
- Beck, A. J., Manderscheid, R. W. and Buerhaus, P. (2018), 'The future of the behavioral health workforce: optimism and opportunity', *American journal of preventive medicine* **54**(6), S187–S189.
- Beck, A. T. (1979), *Cognitive therapy of depression*, Guilford press.
- Bishop, S. J. and Gagne, C. (2018), 'Anxiety, depression, and decision making: a computational perspective', *Annual review of neuroscience* **41**, 371–388.
- Blom, J. O. and Monk, A. F. (2003), 'Theory of personalization of appearance: Why users personalize their pcs and mobile phones', *Human-computer interaction* **18**(3), 193–228.
- Bor, W., Dean, A. J., Najman, J. and Hayatbakhsh, R. (2014), 'Are child and adolescent mental health problems increasing in the 21st century? a systematic review', *Australian & New Zealand journal of psychiatry* **48**(7), 606–616.
- Botsman, R. (2017), *Who can you trust?: how technology brought us together—and why it could drive us apart*, Penguin UK.

- Boyd, K., Potts, C., Bond, R., Mulvenna, M., Broderick, T., Burns, C., Bickerdike, A., Mctear, M., Kostenius, C., Vakaloudis, A. et al. (2022), Usability testing and trust analysis of a mental health and wellbeing chatbot, in 'Proceedings of the 33rd European Conference on Cognitive Ergonomics', pp. 1–8.
- Braun, V. and Clarke, V. (2012), *Thematic analysis.*, American Psychological Association.
- Butryn, T., Bryant, L., Marchionni, C., Sholevar, F. et al. (2017), 'The shortage of psychiatrists and other mental health providers: causes, current state, and potential solutions', *International Journal of Academic Medicine* **3**(1), 5.
- Caldarini, G., Jaf, S. and McGarry, K. (2022), 'A literature survey of recent advances in chatbots', *Information* **13**(1), 41.
- Calefato, F., Lanubile, F. and Novielli, N. (2015), 'The role of social media in affective trust building in customer–supplier relationships', *Electronic Commerce Research* **15**, 453–482.
- Cambria, E. and White, B. (2014), 'Jumping nlp curves: A review of natural language processing research', *IEEE Computational intelligence magazine* **9**(2), 48–57.
- Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., Armour, C. and McTear, M. (2019), Assessing the usability of a chatbot for mental health care, in 'Internet Science: INSCI 2018 International Workshops, St. Petersburg, Russia, October 24–26, 2018, Revised Selected Papers 5', Springer, pp. 121–132.
- Campbell, M. and Gavett, G. (2021), 'What covid-19 has done to our well-being, in 12 charts', *Harvard Business Review* **10**.
- Carlson, K. D. and Herdman, A. O. (2012), 'Understanding the impact of convergent validity on research results', *Organizational Research Methods* **15**(1), 17–32.
- Charness, G., Gneezy, U. and Kuhn, M. A. (2012), 'Experimental methods: Between-subject and within-subject design', *Journal of economic behavior & organization* **81**(1), 1–8.
- Chen, Q. Q. and Park, H. J. (2021), 'How anthropomorphism affects trust in intelligent personal assistants', *Industrial Management & Data Systems* .
- Chin, H., Molefi, L. W. and Yi, M. Y. (2020), Empathy is all you need: How a conversational agent should respond to verbal abuse, in 'Proceedings of the 2020 CHI conference on human factors in computing systems', pp. 1–13.
- Coghlan, S., Leins, K., Sheldrick, S., Cheong, M., Gooding, P. and D'Alfonso, S. (2023), 'To chat or bot to chat: Ethical issues with using chatbots in mental health', *DIGITAL HEALTH* **9**, 20552076231183542.
- Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D. and Bandeira, N. (2017), " what can i help you with?" infrequent users' experiences of intelligent personal assistants, in 'Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services', pp. 1–12.
- Cowell, A. J. and Stanney, K. M. (2005), 'Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility', *International journal of human-computer studies* **62**(2), 281–306.
- Cutler, D. M. and Summers, L. H. (2020), 'The covid-19 pandemic and the \$16 trillion virus', *Jama* **324**(15), 1495–1496.
- Darzi, A. and Evans, T. (2016), 'The global shortage of health workers-an opportunity to transform care'.
- Davis, N., Olsen, N., Perry, V. G., Stewart, M. M. and White, T. B. (2023), 'I'm only human? the role of racial stereotypes, humanness, and satisfaction in transactions with anthropomorphic sales bots', *Journal of the Association for Consumer Research* **8**(1), 000–000.
- De Cicco, R., Iacobucci, S., Aquino, A., Romana Alparone, F. and Palumbo, R. (2022), Understanding users' acceptance of chatbots: An extended tam approach, in 'Chatbot Research and Design: 5th International Workshop, CONVERSATIONS 2021, Virtual Event, November 23–24, 2021, Revised Selected Papers', Springer, pp. 3–22.
- Diederich, S., Brendel, A. B., Morana, S. and Kolbe, L. (2022), 'On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research', *Journal of the Association for Information Systems* **23**(1), 96–138.

- Dimidjian, S. and Davis, K. J. (2009), 'Newer variations of cognitive-behavioral therapy: behavioral activation and mindfulness-based cognitive therapy', *Current psychiatry reports* **11**(6), 453.
- Dosovitsky, G. and Bunge, E. (2022), 'Development of a chatbot for depression: adolescent perceptions and recommendations', *Child and Adolescent Mental Health*.
- Duffy, B. R. (2003), 'Anthropomorphism and the social robot', *Robotics and autonomous systems* **42**(3-4), 177–190.
- Duffy, B. R., Joue, G. and Bourke, J. (2002), Issues in assessing performance of social robots, in 'WSEAS International Conference on Robotics'.
- Elkins, A. C. and Derrick, D. C. (2013), 'The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents', *Group decision and negotiation* **22**(5), 897–913.
- Fitzner, K. (2007), 'Reliability and validity a quick review', *The Diabetes Educator* **33**(5), 775–780.
- Fitzpatrick, K. K., Darcy, A. and Vierhile, M. (2017), 'Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial', *JMIR mental health* **4**(2), e7785.
- Følstad, A., Nordheim, C. B. and Bjørkli, C. A. (2018), What makes users trust a chatbot for customer service? an exploratory interview study, in 'International conference on internet science', Springer, pp. 194–208.
- Goodman, K. L. and Mayhorn, C. B. (2023), 'It's not what you say but how you say it: Examining the influence of perceived voice assistant gender and pitch on trust and reliance', *Applied Ergonomics* **106**, 103864.
- Gotthardt, M., Striegl, J., Loitsch, C. and Weber, G. (2022), Voice assistant-based cbt for depression in students: effects of empathy-driven dialog management, in 'Computers Helping People with Special Needs: 18th International Conference, ICCHP-AAATE 2022, Lecco, Italy, July 11–15, 2022, Proceedings, Part I', Springer, pp. 451–461.
- Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H.-C. and Jeste, D. V. (2019), 'Artificial intelligence for mental health and mental illnesses: an overview', *Current psychiatry reports* **21**, 1–18.
- Gusenbauer, M. (2019), 'Google scholar to overshadow them all? comparing the sizes of 12 academic search engines and bibliographic databases', *Scientometrics* **118**(1), 177–214.
- Gutfleisch, T., Samuel, R. and Sacchi, S. (2021), 'The application of factorial surveys to study recruiters' hiring intentions: comparing designs based on hypothetical and real vacancies', *Quality & Quantity* **55**(3), 775–804.
- Hanson, V. L. (2010), 'Influencing technology adoption by older adults', *Interacting with Computers* **22**(6), 502–509.
- Harilal, N., Shah, R., Sharma, S. and Bhutani, V. (2020), Caro: an empathetic health conversational chatbot for people with major depression, in 'Proceedings of the 7th ACM IKDD CoDS and 25th COMAD', pp. 349–350.
- Hawes, M. T., Szenczy, A. K., Klein, D. N., Hajcak, G. and Nelson, B. D. (2022), 'Increases in depression and anxiety symptoms in adolescents and young adults during the covid-19 pandemic', *Psychological medicine* **52**(14), 3222–3230.
- Hayes, A. F. (2009), 'Beyond baron and kenny: Statistical mediation analysis in the new millennium', *Communication monographs* **76**(4), 408–420.
- Hengstler, M., Enkel, E. and Duelli, S. (2016), 'Applied artificial intelligence and trust—the case of autonomous vehicles and medical assistance devices', *Technological Forecasting and Social Change* **105**, 105–120.
- Hirsch, E. D., Kett, J. F. and Trefil, J. S. (2002), *The new dictionary of cultural literacy*, Houghton Mifflin Harcourt.
- Høiland, C. G., Følstad, A. and Karahasanovic, A. (2020), 'Hi, can i help? exploring how to design a mental health chatbot for youths', *Human Technology* **16**(2), 139.
- Hopko, D. R., Lejuez, C., Ruggiero, K. J. and Eifert, G. H. (2003), 'Contemporary behavioral activation treatments for depression: Procedures, principles, and progress', *Clinical psychology review* **23**(5), 699–717.

- Hossain, M. M., Tasnim, S., Sultana, A., Faizah, F., Mazumder, H., Zou, L., McKyer, E. L. J., Ahmed, H. U. and Ma, P. (2020), 'Epidemiology of mental health problems in covid-19: a review', *F1000Research* **9**.
- Hsiao, K.-L. and Chen, C.-C. (2021), 'What drives continuance intention to use a food-ordering chatbot? an examination of trust and satisfaction', *Library Hi Tech*.
- Huckvale, K., Nicholas, J., Torous, J. and Larsen, M. E. (2020), 'Smartphone apps for the treatment of mental health conditions: status and considerations', *Current opinion in psychology* **36**, 65–70.
- Inkster, B., Sarda, S., Subramanian, V. et al. (2018), 'An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study', *JMIR mHealth and uHealth* **6**(11), e12106.
- International Monetary Fund (IMF) (2022), 'World economic and financial surveys, world economic outlook database—weo groups and aggregates information'.
- Ivarsson, A., Andersen, M. B., Johnson, U. and Lindwall, M. (2013), 'To adjust or not adjust: Nonparametric effect sizes, confidence intervals, and real-world meaning', *Psychology of sport and exercise* **14**(1), 97–102.
- Jaccard, J. and Turrisi, R. (2003), *Interaction effects in multiple regression*, number 72, Sage.
- Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C. and Vukovic, J. (2010), 'A domain-specific opposite-sex bias in human preferences for manipulated voice pitch', *Animal Behaviour* **79**(1), 57–62.
- Kamal, S. A., Shafiq, M. and Kakria, P. (2020), 'Investigating acceptance of telemedicine services through an extended technology acceptance model (tam)', *Technology in Society* **60**, 101212.
- Kanter, J. W., Manos, R. C., Bowe, W. M., Baruch, D. E., Busch, A. M. and Rusch, L. C. (2010), 'What is behavioral activation?: A review of the empirical literature', *Clinical psychology review* **30**(6), 608–620.
- Kasilingam, D. L. (2020), 'Understanding the attitude and intention to use smartphone chatbots for shopping', *Technology in Society* **62**, 101280.
- Khullar, D. (2019), 'Building trust in health care—why, where, and how', *Jama* **322**(6), 507–509.
- Kleinbaum, D. G. and Klein, M. (2010), Ordinal logistic regression, in 'Logistic regression', Springer, pp. 463–488.
- Kleine, R., Galimov, A., Hanewinkel, R., Unger, J., Sussman, S. and Hansen, J. (2023), 'Impact of the covid-19 pandemic on young people with and without pre-existing mental health problems', *Scientific Reports* **13**(1), 6111.
- Klopfenstein, L. C., Delpriori, S., Malatini, S. and Bogliolo, A. (2017), The rise of bots: A survey of conversational interfaces, patterns, and paradigms, in 'Proceedings of the 2017 conference on designing interactive systems', pp. 555–565.
- Komiak, S. Y. and Benbasat, I. (2006), 'The effects of personalization and familiarity on trust and adoption of recommendation agents', *MIS quarterly* pp. 941–960.
- Kruskal, W. H. (1958), 'Ordinal measures of association', *Journal of the American Statistical Association* **53**(284), 814–861.
- Kundi, B., El Morr, C., Gorman, R. and Dua, E. (2023), 'Artificial intelligence and bias: A scoping review', *AI and Society* pp. 199–215.
- Lappeman, J., Marlie, S., Johnson, T. and Poggenpoel, S. (2022), 'Trust and digital privacy: willingness to disclose personal information to banking chatbot services', *Journal of Financial Services Marketing* pp. 1–21.
- Laumer, S., Maier, C. and Gubler, F. T. (2019), 'Chatbot acceptance in healthcare: Explaining user adoption of conversational agents for disease diagnosis'.
- Laurel, B. (1997), 'Interface agents: Metaphors with character', *Human Values and the design of Computer Technology* pp. 207–219.
- Lee, J. D. and See, K. A. (2004), 'Trust in automation: Designing for appropriate reliance', *Human factors* **46**(1), 50–80.
- Lee, O.-K. D., Ayyagari, R., Nasirian, F. and Ahmadian, M. (2021), 'Role of interaction quality and trust in use of ai-based voice-assistant systems', *Journal of Systems and Information Technology* **23**(2), 154–170.

- Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughhead, J., Valdez, J. N., Javitt, D. C., Turetsky, B. and Gur, R. (2010), '" it's not what you say, but how you say it": a reciprocal temporo-frontal network for affective prosody', *Frontiers in human neuroscience* p. 19.
- Lewinsohn, P. M. (1974), 'A behavioral approach to depression', *Essential papers on depression* pp. 150–172.
- Lewis, J. D. and Weigert, A. (1985), 'Trust as a social reality', *Social forces* **63**(4), 967–985.
- Lewis, R., Ferguson, C., Wilks, C., Jones, N. and Picard, R. W. (2022), Can a recommender system support treatment personalisation in digital mental health therapy? a quantitative feasibility assessment using data from a behavioural activation therapy app, in 'CHI Conference on Human Factors in Computing Systems Extended Abstracts', pp. 1–8.
- Liew, T. W., Tan, S.-M. and Ismail, H. (2017), 'Exploring the effects of a non-interactive talking avatar on social presence, credibility, trust, and patronage intention in an e-commerce website', *Human-centric Computing and Information Sciences* **7**, 1–21.
- Lipschitz, J., Miller, C. J., Hogan, T. P., Burdick, K. E., Lippin-Foster, R., Simon, S. R., Burgess, J. et al. (2019), 'Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement', *JMIR mental health* **6**(1), e11334.
- Liu, K. and Tao, D. (2022), 'The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services', *Computers in Human Behavior* **127**, 107026.
- Louvardi, M., Pelekasis, P., Chrousos, G. P. and Darviri, C. (2020), 'Mental health in chronic disease patients during the covid-19 quarantine in greece', *Palliative & Supportive Care* **18**(4), 394–399.
- Lucas, G. M., Gratch, J., King, A. and Morency, L.-P. (2014), 'It's only a computer: Virtual humans increase willingness to disclose', *Computers in Human Behavior* **37**, 94–100.
- Ma, X. and Liu, A. (2020), Challenges in supporting exploratory search through voice assistants, in 'Proceedings of the 2nd Conference on Conversational User Interfaces', pp. 1–3.
- Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S. and Söllner, M. (2019), 'Ai-based digital assistants: Opportunities, threats, and research perspectives', *Business & Information Systems Engineering* **61**, 535–544.
- Maedche, A., Morana, S., Schacht, S., Werth, D. and Krumeich, J. (2016), 'Advanced user assistance systems', *Business & Information Systems Engineering* **58**, 367–370.
- Martell, C. R., Addis, M. E. and Jacobson, N. S. (2001), *Depression in context: Strategies for guided action.*, WW Norton & Co.
- Mayer, R. C., Davis, J. H. and Schoorman, F. D. (1995), 'An integrative model of organizational trust', *Academy of management review* **20**(3), 709–734.
- McKnight, P. E. and Najab, J. (2010), 'Mann-whitney u test', *The Corsini encyclopedia of psychology* pp. 1–1.
- Melcher, J., Camacho, E., Lagan, S. and Torous, J. (2022), 'College student engagement with mental health apps: analysis of barriers to sustained use', *Journal of American College Health* **70**(6), 1819–1825.
- Melián-González, S., Gutiérrez-Taño, D. and Bulchand-Gidumal, J. (2021), 'Predicting the intentions to use chatbots for travel and tourism', *Current Issues in Tourism* **24**(2), 192–210.
- Meng, F., Guo, X., Peng, Z., Ye, Q. and Lai, K.-H. (2022), 'Trust and elderly users' continuance intention regarding mobile health services: the contingent role of health and technology anxieties', *Information Technology & People* **35**(1), 259–280.
- Mnasri, M. (2019), 'Recent advances in conversational nlp: Towards the standardization of chatbot building', *arXiv preprint arXiv:1903.09025* .
- Mori, M., MacDorman, K. F. and Kageki, N. (2012), 'The uncanny valley [from the field]', *IEEE Robotics & automation magazine* **19**(2), 98–100.
- Moriuchi, E. (2021), 'An empirical study on anthropomorphism and engagement with disembodied ais and consumers' re-use behavior', *Psychology & Marketing* **38**(1), 21–42.

- Mostafa, R. B. and Kasamani, T. (2021), 'Antecedents and consequences of chatbot initial trust', *European Journal of Marketing* .
- Müller, L., Mattke, J., Maier, C., Weitzel, T. and Graser, H. (2019), Chatbot acceptance: A latent profile analysis on individuals' trust in conversational agents, in 'Proceedings of the 2019 on Computers and People Research Conference', pp. 35–42.
- Murthi, B. and Sarkar, S. (2003), 'The role of the management sciences in research on personalization', *Management Science* **49**(10), 1344–1362.
- Nadarzynski, T., Miles, O., Cowie, A. and Ridge, D. (2019), 'Acceptability of artificial intelligence (ai)-led chatbot services in healthcare: A mixed-methods study', *Digital health* **5**, 2055207619871808.
- Nasirian, F., Ahmadian, M. and Lee, O.-K. D. (2017), 'Ai-based voice assistant systems: Evaluating from the interaction and trust perspectives'.
- Nass, C. and Moon, Y. (2000), 'Machines and mindlessness: Social responses to computers', *Journal of social issues* **56**(1), 81–103.
- Ng, M., Coopamootoo, K. P., Toreini, E., Aitken, M., Elliot, K. and van Moorsel, A. (2020), Simulating the effects of social presence on trust, privacy concerns & usage intentions in automated bots for finance, in '2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)', IEEE, pp. 190–199.
- Nienaber, A.-M. and Schewe, G. (2014), 'Enhancing trust or reducing perceived risk, what matters more when launching a new product?', *International Journal of Innovation Management* **18**(01), 1450005.
- Nowak, K. L. (2004), 'The influence of anthropomorphism and agency on social judgment in virtual environments', *Journal of Computer-Mediated Communication* **9**(2), JCMC925.
- O'Connor, J. J. and Barclay, P. (2017), 'The influence of voice pitch on perceptions of trustworthiness across social contexts', *Evolution and human behavior* **38**(4), 506–512.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E. et al. (2021), 'The prisma 2020 statement: an updated guideline for reporting systematic reviews', *International journal of surgery* **88**, 105906.
- Palan, S. and Schitter, C. (2018), 'Prolific. ac—a subject pool for online experiments', *Journal of Behavioral and Experimental Finance* **17**, 22–27.
- Parnell, S. I., Klein, S. H. and Gaiser, F. (2022), Do we know and do we care? algorithms and attitude towards conversational user interfaces: Comparing chatbots and voice assistants, in 'Proceedings of the 4th Conference on Conversational User Interfaces', pp. 1–6.
- Peer, E., Brandimarte, L., Samat, S. and Acquisti, A. (2017), 'Beyond the turk: Alternative platforms for crowd-sourcing behavioral research', *Journal of Experimental Social Psychology* **70**, 153–163.
- Perski, O., Blandford, A., West, R. and Michie, S. (2017), 'Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis', *Translational behavioral medicine* **7**(2), 254–267.
- Peters, D. H., Garg, A., Bloom, G., Walker, D. G., Brieger, W. R. and Hafizur Rahman, M. (2008), 'Poverty and access to health care in developing countries', *Annals of the new York Academy of Sciences* **1136**(1), 161–171.
- Pfeuffer, N., Benlian, A., Gimpel, H. and Hinz, O. (2019), 'Anthropomorphic information systems', *Business & Information Systems Engineering* **61**, 523–533.
- Poola, I. (2017), 'How artificial intelligence in impacting real life everyday', *International Journal for Advance Research and Development* **2**(10), 96–100.
- Pousttchi, K. and Wiedemann, D. G. (2007), 'What influences consumers' intention to use mobile payments', *LA Global Mobility Round table* pp. 1–16.
- Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N. and Taylor, S. H. (2017), " alexa is my new bff" social roles, user satisfaction, and personification of the amazon echo, in 'Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems', pp. 2853–2859.

- Qiu, L. and Benbasat, I. (2009), 'Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems', *Journal of management information systems* **25**(4), 145–182.
- Rathnayaka, P., Mills, N., Burnett, D., De Silva, D., Alahakoon, D. and Gray, R. (2022), 'A mental health chatbot with cognitive skills for personalised behavioural activation and remote health monitoring', *Sensors* **22**(10), 3653.
- Raunio, K. (2021), Chatbot anthropomorphism: Adoption and acceptance in customer service, Master's thesis, University of Twente.
- Redmiles, D. D. and Schumann, J. (2011), 'Supporting initial trust in distributed idea generation and evaluation'.
- Rejaibi, E., Komaty, A., Meriaudeau, F., Agrebi, S. and Othmani, A. (2022), 'Mfcc-based recurrent neural network for automatic clinical depression recognition and assessment from speech', *Biomedical Signal Processing and Control* **71**, 103107.
- Rempel, J. K., Holmes, J. G. and Zanna, M. P. (1985), 'Trust in close relationships.', *Journal of personality and social psychology* **49**(1), 95.
- Richards, D. A., Ekers, D., McMillan, D., Taylor, R. S., Byford, S., Warren, F. C., Barrett, B., Farrand, P. A., Gilbody, S., Kuyken, W. et al. (2016), 'Cost and outcome of behavioural activation versus cognitive behavioural therapy for depression (cobra): a randomised, controlled, non-inferiority trial', *The Lancet* **388**(10047), 871–880.
- Rickwood, D. J., Deane, F. P. and Wilson, C. J. (2007), 'When and how do young people seek professional help for mental health problems?', *Medical journal of Australia* **187**(S7), S35–S39.
- Rodríguez Cardona, D., Janssen, A., Guhr, N., Breitner, M. H. and Milde, J. (2021), A matter of trust? examination of chatbot usage in insurance business, in 'proceedings of the 54th Hawaii international conference on system sciences', p. 556.
- Rousseau, D. M., Sitkin, S. B., Burt, R. S. and Camerer, C. (1998), 'Not so different after all: A cross-discipline view of trust', *Academy of management review* **23**(3), 393–404.
- Roy, R. and Naidoo, V. (2021), 'Enhancing chatbot effectiveness: The role of anthropomorphic conversational styles and time orientation', *Journal of Business Research* **126**, 23–34.
- Royston, P. (1992), 'Approximating the shapiro-wilk w-test for non-normality', *Statistics and computing* **2**(3), 117–119.
- Rusman, E., Van Bruggen, J., Sloep, P. and Koper, R. (2010), 'Fostering trust in virtual project teams: Towards a design framework grounded in a trustworthiness antecedents (twan) schema', *International journal of human-computer studies* **68**(11), 834–850.
- Sah, Y. J. and Peng, W. (2015), 'Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website', *Computers in Human Behavior* **45**, 392–401.
- Santos, J. R. A. (1999), 'Cronbach's alpha: A tool for assessing the reliability of scales', *Journal of extension* **37**(2), 1–5.
- Schirmer, A., Feng, Y., Sen, A. and Penney, T. B. (2019), 'Angry, old, male—and trustworthy? how expressive and person voice characteristics shape listener trust', *PloS one* **14**(1), e0210555.
- Schoorman, F. D., Mayer, R. C. and Davis, J. H. (2007), 'An integrative model of organizational trust: Past, present, and future'.
- Schreuter, D., van der Putten, P. and Lamers, M. H. (2021), 'Trust me on this one: conforming to conversational assistants', *Minds and Machines* **31**, 535–562.
- Schroeder, J. and Epley, N. (2016), 'Mistaking minds and machines: How speech affects dehumanization and anthropomorphism.', *Journal of Experimental Psychology: General* **145**(11), 1427.
- Seymour, W. and Van Kleek, M. (2021), 'Exploring interactions between trust, anthropomorphism, and relationship development in voice assistants', *Proceedings of the ACM on Human-Computer Interaction* **5**(CSCW2), 1–16.
- Shamon, H. and Berning, C. (2019), 'Attention check items and instructions in online surveys with incentivized and non-incentivized samples: Boon or bane for data quality?', *Shamon, H., & Berning, CC (2020). Attention Check Items and Instructions in Online Surveys with Incentivized and Non-Incentivized Samples: Boon or Bane for Data Quality* pp. 55–77.

- Simpson, S. S. and Piquero, N. L. (2002), 'Low self-control, organizational theory, and corporate crime', *Law and Society review* pp. 509–548.
- Sitthipon, T., Siripipatthanakul, S., Phayaphrom, B., Siripipattanakul, S. and Limna, P. (2022), 'Determinants of customers' intention to use healthcare chatbots and apps in bangkok, thailand', *International Journal of Behavioral Analytics* **2**(2), 1–15.
- Skjuve, M. and Brandzaeg, P. B. (2019), Measuring user experience in chatbots: An approach to interpersonal communication competence, in 'Internet Science: INSCI 2018 International Workshops, St. Petersburg, Russia, October 24–26, 2018, Revised Selected Papers 5', Springer, pp. 113–120.
- Söllner, M., Hoffmann, A. and Leimeister, J. M. (2016), 'Why different trust relationships matter for information systems users', *European Journal of Information Systems* **25**(3), 274–287.
- Song, S. W. and Shin, M. (2022), 'Uncanny valley effects on chatbot trust, purchase intention, and adoption intention in the context of e-commerce: The moderating role of avatar familiarity', *International Journal of Human–Computer Interaction* pp. 1–16.
- South, A., Joharatnam-Hogan, N., Purvis, C., James, E. C., Diaz-Montana, C., Cragg, W. J., Tweed, C., Macnair, A., Sydes, M. R., Snowdon, C. et al. (2021), 'Testing approaches to sharing trial results with participants: The show respect cluster randomised, factorial, mixed methods trial', *PLoS medicine* **18**(10), e1003798.
- Steiner, P. M., Atzmüller, C. and Su, D. (2016), 'Designing valid and reliable vignette experiments for survey research: A case study on the fair gender income gap', *Journal of Methods and Measurement in the Social Sciences* **7**(2), 52–94.
- Stieglitz, S., Brachten, F. and Kissmer, T. (2018), 'Defining bots in an enterprise context'.
- Storrie, K., Ahern, K. and Tuckett, A. (2010), 'A systematic review: students with mental health problems—a growing problem', *International journal of nursing practice* **16**(1), 1–6.
- Strauss, A. L. (1978), 'Negotiations: Varieties, contexts, processes, and social order', (*No Title*) .
- Striegl, J., Gotthardt, M., Loitsch, C. and Weber, G. (2022), Investigating the usability of voice assistant-based cbt for age-related depression, in 'Computers Helping People with Special Needs: 18th International Conference, ICCHP-AAATE 2022, Lecco, Italy, July 11–15, 2022, Proceedings, Part I', Springer, pp. 432–441.
- Takahashi, K., Muraoka, R. and Otsuka, K. (2020), 'Technology adoption, impact, and extension in developing countries' agriculture: A review of the recent literature', *Agricultural Economics* **51**(1), 31–45.
- Terblanche, N. and Kidd, M. (2022), 'Adoption factors and moderating effects of age and gender that influence the intention to use a non-directive reflective coaching chatbot', *SAGE Open* **12**(2), 21582440221096136.
- The Lancet Global Health (2020), 'Mental health matters', *The Lancet. Global Health* **8**(11), e1352.
- Toader, D.-C., Boca, G., Toader, R., Măcelaru, M., Toader, C., Ighian, D. and Rădulescu, A. T. (2019), 'The effect of social presence and chatbot errors on trust', *Sustainability* **12**(1), 256.
- Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J. S., Leimeister, J. M. M. and Bernstein, A. (2021), Female by default?—exploring the effect of voice assistant gender and pitch on trait and trust attribution, in 'Extended abstracts of the 2021 CHI conference on human factors in computing systems', pp. 1–7.
- Troshani, I., Rao Hill, S., Sherman, C. and Arthur, D. (2021), 'Do we trust in ai? role of anthropomorphism and intelligence', *Journal of Computer Information Systems* **61**(5), 481–491.
- Twenge, J. M., Joiner, T. E., Rogers, M. L. and Martin, G. N. (2018), 'Increases in depressive symptoms, suicide-related outcomes, and suicide rates among us adolescents after 2010 and links to increased new media screen time', *Clinical Psychological Science* **6**(1), 3–17.
- Ursachi, G., Horodnic, I. A. and Zait, A. (2015), 'How reliable are measurement scales? external factors with indirect influence on reliability estimators', *Procedia Economics and Finance* **20**, 679–686.
- van Hooijdonk, C. and Liebrecht, C. (2018), '"wat vervelend dat de fiets niet is opgeruimd! heb je een zaaknummer voor mij?^ ek" conversational human voice in webcare van nederlandse gemeenten'.

- Vesanen, J. (2007), 'What is personalization? a conceptual framework', *European Journal of Marketing* **41**(5/6), 409–418.
- Vesanen, J. and Raulas, M. (2006), 'Building bridges for personalization: a process model for marketing', *Journal of Interactive marketing* **20**(1), 5–20.
- Wald, R., Heijlselaar, E. and Bosse, T. (2021), Make your own: The potential of chatbot customization for the development of user trust, in 'Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization', pp. 382–387.
- Wallander, L. (2009), '25 years of factorial surveys in sociology: A review', *Social science research* **38**(3), 505–520.
- Wang, S., Chen, L., Ran, H., Che, Y., Fang, D., Sun, H., Peng, J., Liang, X. and Xiao, Y. (2022), 'Depression and anxiety among children and adolescents pre and post covid-19: A comparative meta-analysis', *Frontiers in psychiatry* p. 1752.
- Watson, D. (1992), 'Correcting for acquiescent response bias in the absence of a balanced scale: An application to class consciousness', *Sociological Methods & Research* **21**(1), 52–88.
- West, M., Kraut, R. and Ei Chew, H. (2019), 'I'd blush if i could: closing gender divides in digital skills through education'.
- WHO (2022), 'Mental disorders'. (accessed: 20.01.2023).
URL: <https://www.who.int/news-room/fact-sheets/detail/mental-disorders>
- Wienrich, C., Ebner, F. and Carolus, A. (2022), Giving alexa a face-implementing a new research prototype and examining the influences of different human-like visualizations on the perception of voice assistants, in 'Human-Computer Interaction. User Experience and Behavior: Thematic Area, HCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part III', Springer, pp. 605–625.
- Wieringa, R. J. (2014), *Design science methodology for information systems and software engineering*, Springer.
- Williams, D., Caplan, S. and Xiong, L. (2007), 'Can you hear me now? the impact of voice in an online gaming community', *Human communication research* **33**(4), 427–449.
- Xu, K. (2020), 'Language, modality, and mobile media use experiences: Social responses to smartphone cues in a task-oriented context', *Telematics and Informatics* **48**, 101344.
- Zhu, Y., Wang, R. and Pu, C. (2022), "'i am chatbot, your virtual mental health adviser.'" what drives citizens' satisfaction and continuance intention toward mental health chatbots during the covid-19 pandemic? an empirical study in china', *Digital Health* **8**, 20552076221090031.

A The script of all the chatbots

[BA-welcome] Hi, I am your personal therapist. What is your name?

⇒ Input text box.

[BA-AM-1a] Nice to meet you!

[BA-AM-1b] I was told that you haven't been feeling well lately. But don't worry, I am here to help. Actually, I have designed a plan for us to go through together today, to make you feel better about yourself.

[BA-AM-1c] It's called behavioral activation, you might have heard about it. But, just in case, I am going to take you through the plan together to get the most out of it! How does that sound?

[BA-AM-1d] So... , my first question for you is, could you think of activities you are doing in your life? Think of what you are doing within a normal week. It could be anything. For example, going to work, playing sports, playing with your kids, reading the newspaper, etcetera.

[BA-AM-1e] For this session, please mention just 1 activity. You can type your answer in this chat.

⇒ Input text box.

[BA-AM-2a] Ah, thank you for sharing. Now, please also assign the mood that you have when you do the mentioned activity. For example, when you are doing grocery shopping, it could make you feel happy. Or, when you are working, it could make you feel sad. Again, you can type your answer in this chat.

⇒ Input text box.

[BA-AM-2b] Thank you for sharing all of this with me. Now, allow me to explain why I asked you these questions and how this could bring you a step closer to feeling more positive in life.

[BA-AM-2c] So... , when we feel sad or depressed, we tend to feel less motivated to do activities and we prefer just doing nothing. And by doing nothing, this by itself could also make you feel even worse about yourself. And this cycle could continue forever.

[BA-AM-2d] In order to break the cycle, we could minimize the activities that make you feel sad and increase the activities that make you feel happy. As a result, you could feel a bit happier in life. And this positive reinforcement could make you to be more open for more activities to do. Eventually, you will enter this positive reinforcement cycle that could improve your general mood. And this also helps you to cope with your negativity.

[BA-AM-2e] So, it's important to be aware of what you are doing throughout the day, and how this affects your mood. Once we find out which activities make you feel happy, you should try to do these more often. Unfortunately, there is no easy fix for depression, but we can take it step by step.

[BA-AM-2f] Does this concept make sense to you?

⇒ Option: Yes (-continue-) or No ([BA-AM-2g] "It's okay. I think you'll be able to understand the concept after our session. So for now, let's move on.")

[BA-AM-3a] Now, let's try to find the ideal activities for you to do, using your values and life goals.

[BA-AM-3b] While some behaviors, like exercise and meditation, can be used right away to improve your mood, many of the behaviors that are likely to help us are those that align with the things that are important to us.

[BA-AM-3c] For example, the value statement "I value being a good parent" can offer activation tasks like spending more time with your kids or fixing your kid's damaged toy.

[BA-AM-3d] So, try to think of some values and goals that you have. For this session, please just share the 1 thing

that you value the most. And again, you may type your answer in this chat.

⇒ Input text box.

[BA-AS-1] What an amazing life goal! Thanks for sharing. So, my homework for you the coming weeks is to try to do more activities that contribute to the life goal that you just mentioned to me. After doing the activity, please keep track of how you felt.

[BA-end] In our next session, you could tell me all about it and how it went. For now, have a nice day, and let's talk again soon!

B Questionnaires

Anthropomorphism

Q1. I find the chatbot acting natural.

Q2. I find the chatbot humanlike.

Q3. I find the chatbot lifelike.

Affective trust

Benevolence

B1. This chatbot puts my interests first.

B2. This chatbot is genuinely interested in my wellbeing.

B3. If problems arise, one can expect to be treated fairly by this chatbot.

B4. If I got into difficult times in my life or I required help, this chatbot would do its best to help me.

B5. I feel like it could be hard for me to connect with this chatbot.

B6. I would describe the chatbot as a "warm" communication partner.

B7. The chatbot made me feel like it cared about me.

B8. This chatbot makes an effort to understand what I have to say.

B9. The chatbot is friendly and approachable.

B10. This chatbot keeps me updated about what is in its plans.

B11. The chatbot made me feel close to the chatbot.

Cognitive trust

Ability

A1. This chatbot is very competent.

A2. This chatbot could be able to fully satisfy my needs.

A3. One can expect good advice from this chatbot.

A4. In its job this chatbot seems to work efficiently.

A5. I have confidence in the skills of this chatbot.

A6. This chatbot does not provides its services with skills.

A7. This chatbot does things competently/in a capable manner.

Predictability

P1. This chatbot's methods of operation are unclear.

P2. I would rely on advice from this chatbot.

P3. This chatbot does things that promises to do for me.

Overall Trust

Q1. Overall, right now I trust the chatbot.

Q2. Overall, I could build trust with the chatbot in the long-term.

Intention to Use

Q1. I intend to use or to continue using this chatbot in the future.

Q2. When required, I will use this chatbot. Q3. I intend to use this chatbot in the future. Q4. I think that more and more people will use this chatbot.

Vocal Communication Preference

Q1. I would prefer this chatbot to talk to me using voice messages over text messages.

Q2. I foresee benefits if this chatbot communicates using voice messages.

C Survey Design



Welcome to the research study!

You are being invited to participate in a research study titled "Intention to use of chatbots for people experiencing depression". This study is being done by Kevin Doan from the TU Delft.

The purpose of this research study is to understand the preference for chatbot design better of people experiencing depression, in which the data collected from this online survey design will help to gain better understanding. This study takes you approximately 8 minutes to complete. Conducting this research study is for my master's thesis, which is a requirement for my graduation. We will be asking you to interact with a chatbot, after which your opinion will be asked.

As with any online activity the risk of a breach is always possible. To the best of our ability your answers in this study will remain confidential. We will minimize any risks by irreversibly anonymizing collected data. Furthermore, data will be stored in line with the General Data Protection Regulation (GDPR). In addition, age, gender, and occupation will be asked in this study, which is personal data. As such, data from this study will be only used for this study and stored securely offline, and will not be shared with any other party.

Your participation in this study is entirely voluntary and you can withdraw at any time during the survey. You are free to omit any questions. After finishing the survey, it is not possible to remove the data given collected data is irreversibly anonymized.

For further questions, please contact k.q.doan@student.tudelft.nl

Please indicate if you agree to the terms of consent:

I consent, begin the study

I do not consent, I do not wish to participate

Figure 25: The start of study, in which informed consent is asked to the participant. If not consent is given, the study will end immediately for the participant.



What is your Prolific ID? *Please note that this response should auto-fill with the correct ID*



Figure 26: Since Prolific was used to find participants, their Prolific ID was asked in order to compensate the participant.



Do you experience depression?

No

Yes

Do you have any hearing loss or hearing difficulties?

No

Yes

Are you 30 or younger?

Yes

No



Figure 27: To align with Prolific's pre-screening requirements, these are also asked in this section to make sure only eligible participant can enter the study. If not eligible, the study will end immediately for the participant.

Imagine that you are currently having a depressed mood and struggling with your mental health. This makes you feel sad and lose interest in doing your daily activities. However, you have been informed that a new chatbot has been introduced on the market that is able to help you feel more positive in the long run. Normally, it requires you to join multiple sessions with the chatbot in order to notice improvements with your mental health. In other words, using this chatbot long-term could encourage you to get your motivation back for doing activities.

You have been invited to try out this chatbot today to see if you feel comfortable using the chatbot. It will be an introduction session to get informed on how the chatbot can help you. To benefit the most from the chatbot, it is required to be in a silent area/environment to interact with the chatbot.

Are you in a silent area right now?

No, end study

Yes, continue with the survey

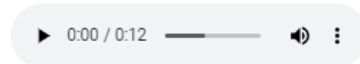


Figure 28: The study starts by introducing the participant with an fictitious situation. In addition, participant was asked to go to a silent area. If the participant was not able to be in a silent area, the study will end immediately for the participant.

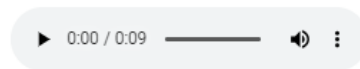
Before you can interact with the chatbot, you are allowed to personalize the chatbot to your taste. In this situation, the chatbot is able to speak to you using a voice. You may now choose one of the following voice options:

Female voice

Slow



Fast

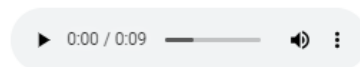


Male voice

Slow



Fast



What type of voice would you like to have for your chatbot?

Female - Slow
Female - Fast
Male - Slow
Male - Fast



Figure 29: This is the screen for Treatment Group II, in which the participant is able to listen and choose the preferred voice for the chatbot to have. The Control Group and Treatment Group I would not see this screen.

The study will now proceed in a chatbot environment

You can now access the chatbot interaction in a new window. Be aware that the chatbot will interact with you by sending voice messages, which is why it is important for you to be in a silent area before starting the chatbot. When the chatbot asks you a question, you can reply by typing in the chat box.

Please note:

1. It is important to **keep the current window open** while participating in the chatbot interaction.
2. You will have to **return to this window after completing the chatbot** interaction to proceed with the study.

To start the interaction, please click on the link below:

[START chatbot interaction](#)

Did you complete the chatbot interaction **as instructed above**?

No, end study

Yes, continue with the survey



Figure 30: This is the screen in which all participants will be directed to the Botstar chatbot. Depending on the allocated treatment/control group (and if applicable, depending on the chosen voice), the participant is will be directed to the chatbot accordingly.

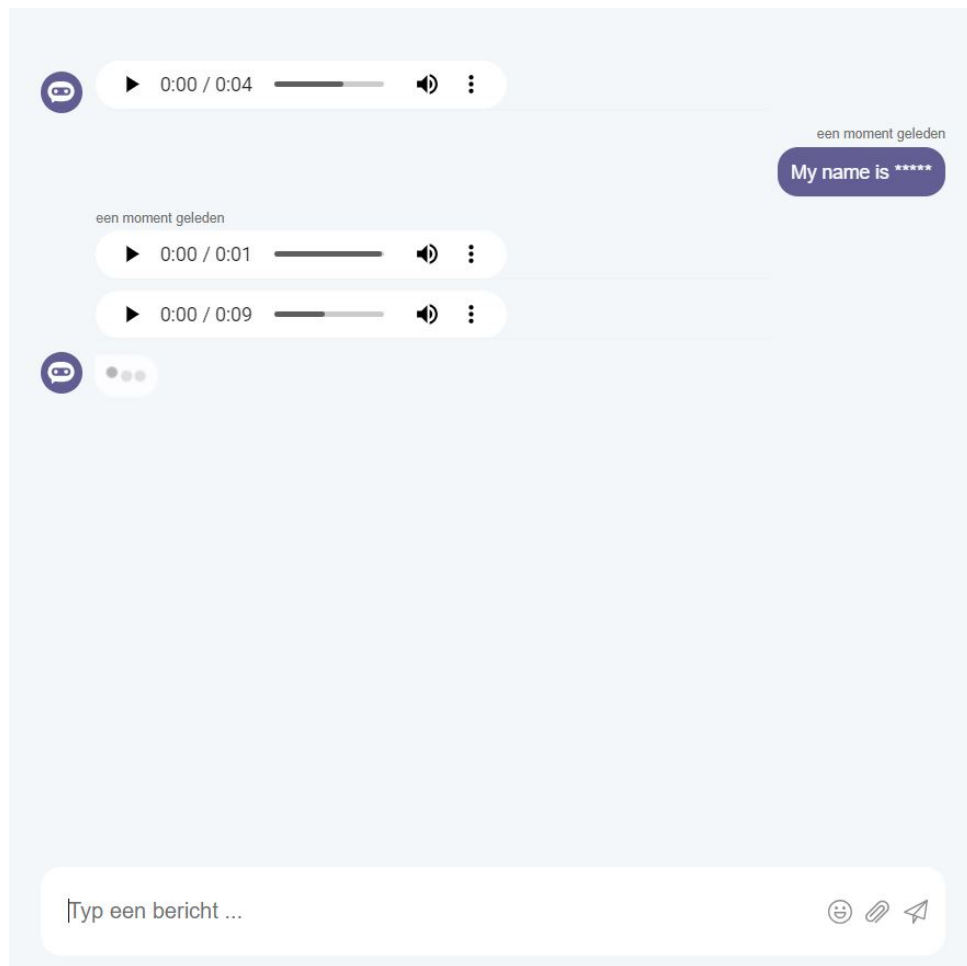


Figure 31: This is how the chabot would look like for Treatment Group I and II, in which voice messages are used to communicate with the participant. The control group will see text messages instead of voice messages.



Now we would like to ask you some questions about your experience with the chatbot.



Figure 32: After interacting with the chatbot, the participant is made aware of the questionnaires that would come next.

Please rate the chatbot you just interacted with on the following items:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I find the chatbot acting natural.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I find the chatbot humanlike.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I find the chatbot lifelike.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Figure 33: Here participants are asked to answer the questions of the anthropomorphism scale.

Please rate the chatbot you just interacted with on the following items:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
This chatbot puts my interests first.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This chatbot is genuinely interested in my wellbeing.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If problems arise, one can expect to be treated fairly by this chatbot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If I got into difficult times in my life or I required help, this chatbot would do its best to help me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I feel like it could be hard for me to connect with this chatbot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would describe the chatbot as a "warm" communication partner.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The chatbot made me feel like it cared about me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This chatbot makes an effort to understand what I have to say.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
The chatbot is friendly and approachable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This chatbot keeps me updated about what is in its plans.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The chatbot made me feel close to the chatbot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Figure 34: Here participants are asked to answer the questions of the benevolence scale.

Please rate the chatbot you just interacted with on the following items:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
This chatbot is very competent.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This chatbot could be able to fully satisfy my needs.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
One can expect good advice from this chatbot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In its job this chatbot seems to work efficiently.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have confidence in the skills of this chatbot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This chatbot does not provides its services with skills.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This chatbot does things competently/in a capable manner.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Figure 35: Here participants are asked to answer the questions of the ability scale.

You just interacted with a chatbot. What was the given advice for you?

To meditate for half an hour a day.

Think of activities to do that align with my goals and values.

Increase mental health by taking more rest.

Please rate the chatbot you just interacted with on the following items:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
This chatbot's methods of operation are unclear.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would rely on advice from this chatbot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This chatbot does things that promises to do for me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Figure 36: Here participants are asked to answer the questions of the predictability scale. In addition, an attention check is applied to check whether the participant had paid attention to the messages of the chatbot.

Please rate the chatbot you just interacted with on the following items:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Overall, right now I trust the chatbot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall, I could build trust with the chatbot in the long-term.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 37: Here participants are asked to answer the questions of the overall trust scale.

Please rate the chatbot you just interacted with on the following items:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I intend to use or to continue using this chatbot in the future.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When required, I will use this chatbot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I intend to use this chatbot in the future.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think that more and more people will use this chatbot.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 38: Here participants are asked to answer the questions of the intention to use scale.

Please rate the chatbot you just interacted with on the following items:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I would prefer this chatbot to talk to me using voice messages over text messages.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I foresee benefits if this chatbot communicates using voice messages.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Figure 39: Here participants are asked to answer the questions of the voice communication preference scale.

What is your age?

17 or younger

18-24

25-30

31 or older

What is your gender?

Male

Female

Other

What is the highest degree or level of education you have completed?

Secondary School

High School

Bachelor's Degree

Master's Degree

PhD or higher



Figure 40: Here participants are asked about their demographic information.



Do you have any feedback or thoughts on this survey?

Do you have any feedback or thoughts on the chatbot you interacted with?



Figure 41: Here participants are free to provide feedback/comment about their experience with the survey and chatbot.



Thank you for taking part in this study. Please click the button below to be redirected back to Prolific and register your submission.



Figure 42: The end of the study.

D Shapiro-Wilk test results

Table 32: Shapiro-Wilk test on the Anthropomorphism scale.

Construct item	Significance (p)	Calc W
Q1	3.15E-12	0.95709
Q2	2.20E-16	0.92446
Q3	4.37E-16	0.93044

Table 33: Shapiro-Wilk test on the Benevolence scale.

Construct item	Significance (p)	Calc W
Q1	5.12E-05	0.98749
Q2	3.44E-16	0.92959
Q3	2.20E-16	0.91209
Q4	2.20E-16	0.91350
Q5	2.20E-16	0.87892
Q6	2.20E-16	0.92269
Q7	1.03E-15	0.93342
Q8	2.20E-16	0.92618
Q9	2.20E-16	0.88492
Q10	2.20E-16	0.87286
Q11	6.48E-15	0.93948

Table 34: Shapiro-Wilk test on the Ability scale.

Construct item	Significance (p)	Calc W
Q1	1.78E-06	0.98296
Q2	2.20E-16	0.91630
Q3	2.20E-16	0.91183
Q4	2.20E-16	0.89917
Q5	4.77E-15	0.93849
Q6	1.07E-13	0.94797
Q7	2.20E-16	0.91605

Table 35: Shapiro-Wilk test on the Predictability scale.

Construct item	Significance (p)	Calc W
Q1	2.82E-09	0.97212
Q2	4.25E-16	0.93034
Q3	2.20E-16	0.91807

Table 36: Shapiro-Wilk test on the Overall Trust scale.

Construct item	Significance (p)	Calc W
Q1	5.46E-13	0.95251
Q2	2.20E-16	0.92335

Table 37: Shapiro-Wilk test on the Intention to Use scale.

Construct item	Significance (p)	Calc W
Q1	1.37E-11	0.96069
Q2	2.20E-16	0.90763
Q3	2.20E-16	0.92176
Q4	2.20E-16	0.90662

Table 38: Shapiro-Wilk test on the Voice Preference scale.

Construct item	Significance (p)	Calc W
Q1	9.94E-08	0.95441
Q2	9.52E-12	0.91301

E Spearman's rank correlation coefficient results

Table 40: Spearman's rank correlation coefficients of the Ability scale.

Relationship	Coefficient (r)	Significance (p)
Q1Q2	0.8589427	2.20E-16
Q1Q3	0.8211076	2.20E-16
Q1Q4	0.8534740	2.20E-16
Q1Q5	0.9297600	2.20E-16
Q1Q6	0.4208093	2.20E-16
Q1Q7	-0.8159133	2.20E-16
Q2Q3	0.5941217	2.20E-16
Q2Q4	0.6103795	2.20E-16
Q2Q5	0.7688836	2.20E-16
Q2Q6	0.3469101	2.20E-16
Q2Q7	-0.6791306	2.20E-16
Q3Q4	0.7140450	2.20E-16
Q3Q5	0.7275590	2.20E-16
Q3Q6	0.3658440	2.20E-16
Q3Q7	-0.6534733	2.20E-16
Q4Q5	0.7721698	2.20E-16
Q4Q6	0.3852630	2.20E-16
Q4Q7	-0.7309399	2.20E-16
Q5Q6	0.4098038	2.20E-16
Q5Q7	-0.7688138	2.20E-16
Q6Q7	-0.3803352	2.20E-16

Table 41: Spearman's rank correlation coefficients of the Predictability scale.

Relationship	Coefficient (r)	Significance (p)
Q1Q2	0.9084031	2.20E-16
Q1Q3	0.8218462	2.20E-16
Q2Q3	0.5353007	2.20E-16

Table 42: Spearman's rank correlation coefficients of the Overall Trust scale.

Relationship	Coefficient (r)	Significance (p)
Q1Q2	0.9539096	2.20E-16

Table 39: Spearman's rank correlation coefficients of the Benevolence scale.

Relationship	Coefficient (<i>r</i>)	Significance (<i>p</i>)
Q1Q2	0.8449081	2.20E-16
Q1Q3	0.7382517	2.20E-16
Q1Q4	0.7748715	2.20E-16
Q1Q5	0.7518510	2.20E-16
Q1Q6	0.7861899	2.20E-16
Q1Q7	0.8997270	2.20E-16
Q1Q8	0.8483525	2.20E-16
Q1Q9	0.7508062	2.20E-16
Q1Q10	0.4955351	2.20E-16
Q1Q11	0.8634064	2.20E-16
Q2Q3	0.5824058	2.20E-16
Q2Q4	0.5889917	2.20E-16
Q2Q5	0.5854024	2.20E-16
Q2Q6	0.6514617	2.20E-16
Q2Q7	0.7479659	2.20E-16
Q2Q8	0.6858160	2.20E-16
Q2Q9	0.5843355	2.20E-16
Q2Q10	0.3834188	2.20E-16
Q2Q11	0.6931417	2.20E-16
Q3Q4	0.6493782	2.20E-16
Q3Q5	0.4655181	2.20E-16
Q3Q6	0.5099279	2.20E-16
Q3Q7	0.6264850	2.20E-16
Q3Q8	0.5612684	2.20E-16
Q3Q9	0.5540214	2.20E-16
Q3Q10	0.3871435	2.20E-16
Q3Q11	0.5653134	2.20E-16
Q4Q5	0.5457350	2.20E-16
Q4Q6	0.5086394	2.20E-16
Q4Q7	0.6407351	2.20E-16
Q4Q8	0.6211366	2.20E-16
Q4Q9	0.5815461	2.20E-16
Q4Q10	0.5090091	2.20E-16
Q4Q11	0.6108371	2.20E-16
Q5Q6	0.5520970	2.20E-16
Q5Q7	0.6495100	2.20E-16
Q5Q8	0.6074029	2.20E-16
Q5Q9	0.4609700	2.20E-16
Q5Q10	0.2778987	4.04E-12
Q5Q11	0.6623488	2.20E-16
Q6Q7	0.7487468	2.20E-16
Q6Q8	0.6094797	2.20E-16
Q6Q9	0.6087822	2.20E-16
Q6Q10	0.3491039	2.20E-16
Q6Q11	0.6612216	2.20E-16
Q7Q8	0.7424067	2.20E-16
Q7Q9	0.6819293	2.20E-16
Q7Q10	0.4329850	2.20E-16
Q7Q11	0.7756998	2.20E-16
Q8Q9	0.5875346	2.20E-16
Q8Q10	0.4164301	2.20E-16
Q8Q11	0.7525347	2.20E-16
Q9Q10	0.5434638	2.20E-16
Q9Q11	0.6264360	2.20E-16
Q10Q11	0.4250609	2.20E-16

Table 43: Spearman's rank correlation coefficients of the Intention to Use scale.

Relationship	Coefficient (<i>r</i>)	Significance (<i>p</i>)
Q1Q2	0.9064506	2.20E-16
Q1Q3	0.9614756	2.20E-16
Q1Q4	0.8678428	2.20E-16
Q2Q3	0.8379929	2.20E-16
Q2Q4	0.7045930	2.20E-16
Q3Q4	0.7753650	2.20E-16

Table 44: Spearman's rank correlation coefficients of the Anthropomorphism scale.

Relationship	Coefficient (<i>r</i>)	Significance (<i>p</i>)
Q1Q2	0.9355990	2.20E-16
Q1Q3	0.9399552	2.20E-16
Q2Q3	0.8343632	2.20E-16

Table 45: Spearman's rank correlation coefficients of the speech delivery Preference scale.

Relationship	Coefficient (<i>r</i>)	Significance (<i>p</i>)
Q1Q2	0.9075022	2.20E-16

F Demographics Characteristics per treatment/control group

Table 46: Demographics characteristics - Chatbot communicates using text messages.

Variables	Total (<i>N</i> = 201)	Affective Trust	Cognitive Trust	Intention to Use
	<i>N</i> (%)	<i>Mean</i> ± <i>SD</i>	<i>Mean</i> ± <i>SD</i>	<i>Mean</i> ± <i>SD</i>
<i>Gender</i>				
Male	97 (48.25)	174.70 ± 26.81	65.53 ± 9.50	63.23 ± 11.44
Female	94 (46.77)	176.15 ± 25.34	66.03 ± 8.48	62.62 ± 12.03
Other	10 (4.98)	162.7 ± 23.74	62.00 ± 8.67	56.90 ± 12.83
<i>Age</i>				
18-24	82 (40.80)	177.54 ± 25.78	66.88 ± 9.35	63.12 ± 12.13
25-30	119 (59.20)	172.88 ± 26.12	64.70 ± 8.66	62.29 ± 11.61
<i>Education</i>				
Secondary School	8 (3.99)	193.25 ± 23.63	71.00 ± 7.71	68.50 ± 11.15
High School	83 (41.29)	178.90 ± 25.51	66.54 ± 8.83	63.92 ± 11.53
Bachelor's Degree	83 (41.29)	171.28 ± 27.20	64.81 ± 9.44	60.90 ± 12.34
Master's Degree	26 (12.94)	167.54 ± 20.83	63.19 ± 7.78	62.04 ± 10.79
PhD or higher	1 (0.49)	164 ± 0.00	70.00 ± 0.00	67.00 ± 0.00

Table 47: Demographics characteristics - Chatbot communicates using voice messages.

Variables	Total (N = 199)	Affective Trust	Cognitive Trust	Intention to Use
	<i>N (%)</i>	<i>Mean ± SD</i>	<i>Mean ± SD</i>	<i>Mean ± SD</i>
<i>Gender</i>				
Male	91 (45.73)	171.54 ± 26.37	63.69 ± 9.52	61.75 ± 12.61
Female	98 (49.25)	174.75 ± 24.58	64.80 ± 8.50	63.76 ± 10.67
Other	10 (5.02)	167.20 ± 35.30	63.60 ± 11.56	63.00 ± 12.66
<i>Age</i>				
18-24	75 (37.69)	175.97 ± 23.87	65.43 ± 8.40	65.01 ± 10.60
25-30	124 (62.31)	171.04 ± 27.04	63.51 ± 9.47	61.46 ± 12.12
<i>Education</i>				
Secondary School	11 (5.52)	172.46 ± 21.07	64.73 ± 6.05	61.46 ± 11.47
High School	74 (37.19)	175.38 ± 24.22	64.87 ± 8.07	63.65 ± 11.09
Bachelor's Degree	78 (39.20)	173.49 ± 28.44	64.81 ± 10.09	64.18 ± 12.22
Master's Degree	33 (16.58)	167.30 ± 23.99	61.79 ± 9.21	58.61 ± 11.09
PhD or higher	3 (1.51)	159.67 ± 40.25	58.67 ± 14.47	57.00 ± 13.00

Table 48: Demographics characteristics - Chatbot communicates using voice messages, in which voice attributes could be personalized.

Variables	Total (N = 201)	Affective Trust	Cognitive Trust	Intention to Use
	<i>N (%)</i>	<i>Mean ± SD</i>	<i>Mean ± SD</i>	<i>Mean ± SD</i>
<i>Gender</i>				
Male	98 (48.76)	170.04 ± 25.74	62.94 ± 8.86	60.99 ± 11.23
Female	95 (47.26)	174.28 ± 25.70	65.18 ± 8.72	62.38 ± 12.24
Other	8 (3.98)	162.13 ± 15.80	59.00 ± 4.78	54.38 ± 8.00
<i>Age</i>				
18-24	98 (48.76)	171.54 ± 26.25	63.90 ± 8.71	61.49 ± 11.66
25-30	103 (51.24)	171.91 ± 24.83	63.79 ± 8.87	61.28 ± 11.75
<i>Education</i>				
Secondary School	8 (3.98)	179.00 ± 30.04	68.13 ± 10.72	64.63 ± 13.48
High School	69 (34.33)	172.17 ± 27.93	63.65 ± 8.96	61.00 ± 12.23
Bachelor's Degree	90 (44.78)	172.86 ± 23.62	64.17 ± 8.57	62.63 ± 11.00
Master's Degree	31 (15.42)	167.39 ± 24.18	62.58 ± 8.19	59.00 ± 11.38
PhD or higher	3 (1.49)	153.33 ± 22.68	60.00 ± 12.29	48.67 ± 11.55

G Descriptive statistics of measurement scales

Table 49: Descriptive statistics of measurement scales.

Statistic	Mean	SD	Min	Median	Max
Text messages					
Anthropomorphism	15.04	5.59	3	14	24
Benevolence	92.37	13.65	60	91	120
Ability	47.08	6.86	30	46	60
Predictability	18.51	2.55	12	18	24
Overall Trust	8.56	3.20	2	9	14
Intention to Use	26.21	5.26	18	25	36
Vocal communication preference	6.52	3.16	2	6	14
Voice messages					
Anthropomorphism	15.26	5.75	3	14	24
Benevolence	91.27	13.61	60	89.5	120
Ability	45.92	6.90	30	45	60
Predictability	18.31	2.64	12	18	24
Overall Trust	8.19	3.28	2	8	14
Intention to Use	26.39	5.26	18	25	36
Vocal communication preference	8.38	3.34	2	8.5	14
Personalized voice					
Anthropomorphism	15.30	5.28	3	15.5	24
Benevolence	90.69	13.34	60	91.5	120
Ability	45.71	6.86	30	47	60
Predictability	18.13	2.51	12	18	24
Overall Trust	8.17	3.17	2	9	14
Intention to Use	25.66	5.20	18	25	36
Vocal communication preference	9.49	3.04	2	10	14

H R-Code

```

1 # rm(list = ls())
2 # CTRL + L
3
4 # Set working directory
5 setwd("C:/Users/Kevin/Desktop/Thesis TUD/6. Statistics")
6
7
8
9 # Install/load packages
10 #install.packages("stargazer", dependencies = TRUE)
11 library("stargazer")
12 #install.packages("reshape", dependencies = TRUE)
13 library("reshape")
14
15 library(ggplot2)
16 # install.packages("tidyverse")
17 library(tidyverse)
18 library(dplyr)
19 #install.packages("psych")
20 library(psych)
21 #install.packages("stats")
22 library(stats)
23 #install.packages("nFactors")
24 require(nFactors)
25
26
27 ##### Read your data #####
28

```

```

29 #constructs <- read.csv("official_dataset_onlyyoungs_extra - Master.csv", header = TRUE, sep
    =",")
30 #constructs <- read.csv("official_dataset_onlyyoungs_extra - Master - V1V2.csv", header = TRUE,
    sep =",")
31 #constructs <- read.csv("official_dataset_onlyyoungs_extra - Master - V1V3.csv", header = TRUE,
    sep =",")
32 #constructs <- read.csv("official_dataset_onlyyoungs_extra - Master - V2V3.csv", header = TRUE,
    sep =",")
33 constructs <- read.csv("gender_experiment.csv", header = TRUE, sep =",")
34
35 # Reverse the scales
36 myscale <- 6:12
37
38 # The tidyverse way
39 constructs <- constructs %>%
40   mutate(Q8_5 = min(myscale) - Q8_5 + max(myscale))
41
42
43
44 # The tidyverse way
45 constructs <- constructs %>%
46   mutate(Q9_6 = min(myscale) - Q9_6 + max(myscale))
47
48
49 # The tidyverse way
50 constructs <- constructs %>%
51   mutate(Q9_7 = min(myscale) - Q9_7 + max(myscale))
52
53
54 # The tidyverse way
55 constructs <- constructs %>%
56   mutate(Q11_1 = min(myscale) - Q11_1 + max(myscale))
57
58
59
60
61
62
63
64
65
66 ### Validity Analysis
67
68
69
70 ## Randomization Test
71
72 # Age
73
74 randomtestAge <- data.frame(constructs$Q13, constructs$AllConcat)
75
76 randomtestAge<-randomtestAge[randomtestAge$constructs.AllConcat=="1",]
77
78
79 agemean <- mean(as.numeric(randomtestAge$constructs.Q13), na.rm=TRUE)
80 print(agemean)
81 agesd <- sd(as.numeric(randomtestAge$constructs.Q13), na.rm = TRUE)
82 print(agesd)
83
84
85
86 plot(constructs.Q13 ~ constructs.AllConcat, data = randomtestAge)
87 randomtestAge.aov <-aov(constructs.Q13 ~ constructs.AllConcat, data = randomtestAge)
88 summary(randomtestAge.aov)
89
90
91 # Gender
92
93
94
95 randomtestGender <- data.frame(constructs$Q14, constructs$AllConcat)
96
97 randomtestGender<-randomtestGender[randomtestGender$constructs.AllConcat=="3",]

```

```

98
99
100
101 gendermean <- mean(as.numeric(randomtestGender$constructs.Q14),na.rm=TRUE)
102 print(gendermean)
103 gendersd <- sd(as.numeric(randomtestGender$constructs.Q14), na.rm = TRUE)
104 print(gendersd)
105
106
107 plot(constructs.Q14 ~ constructs.AllConcat, data = randomtestGender)
108 randomtestGender.aov <-aov(constructs.Q14 ~ constructs.AllConcat, data = randomtestGender)
109 summary(randomtestGender.aov)
110
111
112
113 # Education
114
115
116 randomtestEdu <- data.frame(constructs$Q15,constructs$AllConcat)
117
118
119 randomtestEdu<-randomtestEdu[randomtestEdu$constructs.AllConcat=="3",]
120
121
122 edumean <- mean(as.numeric(randomtestEdu$constructs.Q15),na.rm=TRUE)
123 print(edumean)
124 edusd <- sd(as.numeric(randomtestEdu$constructs.Q15), na.rm = TRUE)
125 print(edusd)
126
127
128 plot(constructs.Q15 ~ constructs.AllConcat, data = randomtestEdu)
129 randomtestEdu.aov <-aov(constructs.Q15 ~ constructs.AllConcat, data = randomtestEdu)
130 summary(randomtestEdu.aov)
131
132
133
134 # Per condition
135
136 # Age
137 randomtestAge <- data.frame(constructs$Q13,constructs$Gender,constructs$Speed, constructs$
  AllConcat, constructs$V2Merge)
138
139
140 randomtestAge<-randomtestAge[randomtestAge$constructs.AllConcat=="V2" & randomtestAge$
  constructs.Gender=="Female" & randomtestAge$constructs.Speed=="Slow",]
141 randomtestAge<-randomtestAge[randomtestAge$constructs.AllConcat=="V2" & randomtestAge$constructs
  .Gender=="Female" & randomtestAge$constructs.Speed=="Fast",]
142 randomtestAge<-randomtestAge[randomtestAge$constructs.AllConcat=="V2" & randomtestAge$constructs
  .Gender=="Male" & randomtestAge$constructs.Speed=="Slow",]
143 randomtestAge<-randomtestAge[randomtestAge$constructs.AllConcat=="V2" & randomtestAge$constructs
  .Gender=="Male" & randomtestAge$constructs.Speed=="Fast",]
144
145
146 agemean <- mean(as.numeric(randomtestAge$constructs.Q13),na.rm=TRUE)
147 print(agemean)
148 agesd <- sd(as.numeric(randomtestAge$constructs.Q13), na.rm = TRUE)
149 print(agesd)
150
151
152 plot(constructs.Q13 ~ constructs.Gender + constructs.Speed, data = randomtestAge)
153 randomtestAge.aov <-aov(constructs.Q13 ~ constructs.V2Merge, data = randomtestAge)
154 summary(randomtestAge.aov)
155
156
157
158 # Gender
159
160 randomtestGender <- data.frame(constructs$Q14,constructs$Gender,constructs$Speed, constructs$
  AllConcat, constructs$V2Merge)
161 randomtestGender<-randomtestGender[randomtestAge$constructs.AllConcat=="V2" & randomtestGender$
  constructs.Gender=="Female" & randomtestGender$constructs.Speed=="Slow",]
162 randomtestGender<-randomtestGender[randomtestAge$constructs.AllConcat=="V2" & randomtestGender$
  constructs.Gender=="Female" & randomtestGender$constructs.Speed=="Fast",]

```

```

163 randomtestGender<-randomtestGender[randomtestAge$constructs.AllConcat=="V2" & randomtestGender$
    constructs.Gender=="Male" & randomtestGender$constructs.Speed=="Slow",]
164 randomtestGender<-randomtestGender[randomtestAge$constructs.AllConcat=="V2" & randomtestGender$
    constructs.Gender=="Male" & randomtestGender$constructs.Speed=="Fast",]
165
166
167 gendermean <- mean(as.numeric(randomtestGender$constructs.Q14),na.rm=TRUE)
168 print(gendermean)
169 gendersd <- sd(as.numeric(randomtestGender$constructs.Q14), na.rm = TRUE)
170 print(gendersd)
171
172 plot(constructs.Q13 ~ constructs.Gender + constructs.Speed, data = randomtestGender)
173 randomtestGender.aov <-aov(constructs.Q14 ~ constructs.V2Merge, data = randomtestGender)
174 summary(randomtestGender.aov)
175
176
177 # Education
178 randomtestEdu <- data.frame(constructs$Q15,constructs$Gender,constructs$Speed, constructs$
    AllConcat, constructs$V2Merge)
179 randomtestEdu<-randomtestEdu[randomtestAge$constructs.AllConcat=="V2" & randomtestEdu$constructs
    .Gender=="Female" & randomtestEdu$constructs.Speed=="Slow",]
180 randomtestEdu<-randomtestEdu[randomtestAge$constructs.AllConcat=="V2" & randomtestEdu$constructs
    .Gender=="Female" & randomtestEdu$constructs.Speed=="Fast",]
181 randomtestEdu<-randomtestEdu[randomtestAge$constructs.AllConcat=="V2" & randomtestEdu$constructs
    .Gender=="Male" & randomtestEdu$constructs.Speed=="Slow",]
182 randomtestEdu<-randomtestEdu[randomtestAge$constructs.AllConcat=="V2" & randomtestEdu$constructs
    .Gender=="Male" & randomtestEdu$constructs.Speed=="Fast",]
183
184
185
186 edumean <- mean(as.numeric(randomtestEdu$constructs.Q15),na.rm=TRUE)
187 print(edumean)
188 edusd <- sd(as.numeric(randomtestEdu$constructs.Q15), na.rm = TRUE)
189 print(edusd)
190
191 plot(constructs.Q13 ~ constructs.Gender + constructs.Speed, data = randomtestEdu)
192 randomtestEdu.aov <-aov(constructs.Q15 ~ constructs.V2Merge, data = randomtestEdu)
193 summary(randomtestEdu.aov)
194
195
196
197
198
199
200
201 ## Creating relevant dataframe
202 dfQ7 <- data.frame(constructs$Q7_1,constructs$Q7_2,constructs$Q7_3)
203 dfQ8 <- data.frame(constructs$Q8_1,constructs$Q8_2,constructs$Q8_3,constructs$Q8_4,constructs$
    Q8_5,constructs$Q8_6,constructs$Q8_7,constructs$Q8_8,constructs$Q8_9,constructs$Q8_10,
    constructs$Q8_11)
204 dfQ9 <- data.frame(constructs$Q9_1,constructs$Q9_2,constructs$Q9_3,constructs$Q9_4,constructs$
    Q9_5,constructs$Q9_6,constructs$Q9_7)
205 dfQ11 <- data.frame(constructs$Q11_1,constructs$Q11_2,constructs$Q11_3)
206 dfQ11b <- data.frame(constructs$Q11b_1,constructs$Q11b_2)
207 dfQ12 <- data.frame(constructs$Q12_1,constructs$Q12_2,constructs$Q12_3,constructs$Q12_4)
208 dfQ12b <- data.frame(constructs$Q12b_1,constructs$Q12b_2)
209
210
211
212 ## Reliability Analysis
213
214
215 # Nonparametric data/normality test
216 shapiro.test(constructs$Q7_1)
217 shapiro.test(constructs$Q7_2)
218 shapiro.test(constructs$Q7_3)
219
220
221 shapiro.test(constructs$Q8_1)
222 shapiro.test(constructs$Q8_2)
223 shapiro.test(constructs$Q8_3)
224 shapiro.test(constructs$Q8_4)
225 shapiro.test(constructs$Q8_5)

```

```

226 shapiro.test(constructs$Q8_6)
227 shapiro.test(constructs$Q8_7)
228 shapiro.test(constructs$Q8_8)
229 shapiro.test(constructs$Q8_9)
230 shapiro.test(constructs$Q8_10)
231 shapiro.test(constructs$Q8_11)
232
233 shapiro.test(constructs$Q9_1)
234 shapiro.test(constructs$Q9_2)
235 shapiro.test(constructs$Q9_3)
236 shapiro.test(constructs$Q9_4)
237 shapiro.test(constructs$Q9_5)
238 shapiro.test(constructs$Q9_6)
239 shapiro.test(constructs$Q9_7)
240
241
242 shapiro.test(constructs$Q11_1)
243 shapiro.test(constructs$Q11_2)
244 shapiro.test(constructs$Q11_3)
245
246
247
248 shapiro.test(constructs$Q11b_1)
249 shapiro.test(constructs$Q11b_2)
250
251
252 shapiro.test(constructs$Q12_1)
253 shapiro.test(constructs$Q12_2)
254 shapiro.test(constructs$Q12_3)
255 shapiro.test(constructs$Q12_4)
256
257 shapiro.test(constructs$Q12b_1)
258 shapiro.test(constructs$Q12b_2)
259
260
261 # Convergent validity
262
263 # Anthropomorphism
264 corr_Q7_1Q7_2 <- cor.test(x=constructs$Q7_1, y=constructs$Q7_2, method = 'spearman', exact =
FALSE)
265 print(corr_Q7_1Q7_2)
266
267 corr_Q7_1Q7_3 <- cor.test(x=constructs$Q7_1, y=constructs$Q7_3, method = 'spearman', exact =
FALSE)
268 print(corr_Q7_1Q7_3)
269
270 corr_Q7_2Q7_3 <- cor.test(x=constructs$Q7_2, y=constructs$Q7_3, method = 'spearman', exact =
FALSE)
271 print(corr_Q7_2Q7_3)
272
273
274
275 # Benevolence
276 corr_Q8_1Q8_2 <- cor.test(x=constructs$Q8_1, y=constructs$Q8_2, method = 'spearman', exact =
FALSE)
277 print(corr_Q8_1Q8_2)
278
279 corr_Q8_1Q8_3 <- cor.test(x=constructs$Q8_1, y=constructs$Q8_3, method = 'spearman', exact =
FALSE)
280 print(corr_Q8_1Q8_3)
281
282 corr_Q8_1Q8_4 <- cor.test(x=constructs$Q8_1, y=constructs$Q8_4, method = 'spearman', exact =
FALSE)
283 print(corr_Q8_1Q8_4)
284
285 corr_Q8_1Q8_5 <- cor.test(x=constructs$Q8_1, y=constructs$Q8_5, method = 'spearman', exact =
FALSE)
286 print(corr_Q8_1Q8_5)
287
288 corr_Q8_1Q8_6 <- cor.test(x=constructs$Q8_1, y=constructs$Q8_6, method = 'spearman', exact =
FALSE)
289 print(corr_Q8_1Q8_6)
290

```

```

291 corr_Q8_1Q8_7 <- cor.test(x=constructs$Q8_1, y=constructs$Q8_7, method = 'spearman', exact =
    FALSE)
292 print(corr_Q8_1Q8_7)
293
294 corr_Q8_1Q8_8 <- cor.test(x=constructs$Q8_1, y=constructs$Q8_8, method = 'spearman', exact =
    FALSE)
295 print(corr_Q8_1Q8_8)
296
297 corr_Q8_1Q8_9 <- cor.test(x=constructs$Q8_1, y=constructs$Q8_9, method = 'spearman', exact =
    FALSE)
298 print(corr_Q8_1Q8_9)
299
300 corr_Q8_1Q8_10 <- cor.test(x=constructs$Q8_1, y=constructs$Q8_10, method = 'spearman', exact =
    FALSE)
301 print(corr_Q8_1Q8_10)
302
303 corr_Q8_1Q8_11 <- cor.test(x=constructs$Q8_1, y=constructs$Q8_11, method = 'spearman', exact =
    FALSE)
304 print(corr_Q8_1Q8_11)
305
306
307
308 corr_Q8_2Q8_3 <- cor.test(x=constructs$Q8_2, y=constructs$Q8_3, method = 'spearman', exact =
    FALSE)
309 print(corr_Q8_2Q8_3)
310
311 corr_Q8_2Q8_4 <- cor.test(x=constructs$Q8_2, y=constructs$Q8_4, method = 'spearman', exact =
    FALSE)
312 print(corr_Q8_2Q8_4)
313
314 corr_Q8_2Q8_5 <- cor.test(x=constructs$Q8_2, y=constructs$Q8_5, method = 'spearman', exact =
    FALSE)
315 print(corr_Q8_2Q8_5)
316
317 corr_Q8_2Q8_6 <- cor.test(x=constructs$Q8_2, y=constructs$Q8_6, method = 'spearman', exact =
    FALSE)
318 print(corr_Q8_2Q8_6)
319
320 corr_Q8_2Q8_7 <- cor.test(x=constructs$Q8_2, y=constructs$Q8_7, method = 'spearman', exact =
    FALSE)
321 print(corr_Q8_2Q8_7)
322
323 corr_Q8_2Q8_8 <- cor.test(x=constructs$Q8_2, y=constructs$Q8_8, method = 'spearman', exact =
    FALSE)
324 print(corr_Q8_2Q8_8)
325
326 corr_Q8_2Q8_9 <- cor.test(x=constructs$Q8_2, y=constructs$Q8_9, method = 'spearman', exact =
    FALSE)
327 print(corr_Q8_2Q8_9)
328
329 corr_Q8_2Q8_10 <- cor.test(x=constructs$Q8_2, y=constructs$Q8_10, method = 'spearman', exact =
    FALSE) # rho = 0.4297884
330 print(corr_Q8_2Q8_10)
331
332 corr_Q8_2Q8_11 <- cor.test(x=constructs$Q8_2, y=constructs$Q8_11, method = 'spearman', exact =
    FALSE)
333 print(corr_Q8_2Q8_11)
334
335
336
337
338 corr_Q8_3Q8_4 <- cor.test(x=constructs$Q8_3, y=constructs$Q8_4, method = 'spearman', exact =
    FALSE)
339 print(corr_Q8_3Q8_4)
340
341 corr_Q8_3Q8_5 <- cor.test(x=constructs$Q8_3, y=constructs$Q8_5, method = 'spearman', exact =
    FALSE)
342 print(corr_Q8_3Q8_5)
343
344 corr_Q8_3Q8_6 <- cor.test(x=constructs$Q8_3, y=constructs$Q8_6, method = 'spearman', exact =
    FALSE)
345 print(corr_Q8_3Q8_6)
346

```

```

347 corr_Q8_3Q8_7 <- cor.test(x=constructs$Q8_3, y=constructs$Q8_7, method = 'spearman', exact =
      FALSE)
348 print(corr_Q8_3Q8_7)
349
350 corr_Q8_3Q8_8 <- cor.test(x=constructs$Q8_3, y=constructs$Q8_8, method = 'spearman', exact =
      FALSE)
351 print(corr_Q8_3Q8_8)
352
353 corr_Q8_3Q8_9 <- cor.test(x=constructs$Q8_3, y=constructs$Q8_9, method = 'spearman', exact =
      FALSE)
354 print(corr_Q8_3Q8_9)
355
356 corr_Q8_3Q8_10 <- cor.test(x=constructs$Q8_3, y=constructs$Q8_10, method = 'spearman', exact =
      FALSE) # 0.3872125
357 print(corr_Q8_3Q8_10)
358
359 corr_Q8_3Q8_11 <- cor.test(x=constructs$Q8_3, y=constructs$Q8_11, method = 'spearman', exact =
      FALSE)
360 print(corr_Q8_3Q8_11)
361
362
363
364
365
366
367 corr_Q8_4Q8_5 <- cor.test(x=constructs$Q8_4, y=constructs$Q8_5, method = 'spearman', exact =
      FALSE)
368 print(corr_Q8_4Q8_5)
369
370 corr_Q8_4Q8_6 <- cor.test(x=constructs$Q8_4, y=constructs$Q8_6, method = 'spearman', exact =
      FALSE) # 0.499814
371 print(corr_Q8_4Q8_6)
372
373 corr_Q8_4Q8_7 <- cor.test(x=constructs$Q8_4, y=constructs$Q8_7, method = 'spearman', exact =
      FALSE)
374 print(corr_Q8_4Q8_7)
375
376 corr_Q8_4Q8_8 <- cor.test(x=constructs$Q8_4, y=constructs$Q8_8, method = 'spearman', exact =
      FALSE)
377 print(corr_Q8_4Q8_8)
378
379 corr_Q8_4Q8_9 <- cor.test(x=constructs$Q8_4, y=constructs$Q8_9, method = 'spearman', exact =
      FALSE)
380 print(corr_Q8_4Q8_9)
381
382 corr_Q8_4Q8_10 <- cor.test(x=constructs$Q8_4, y=constructs$Q8_10, method = 'spearman', exact =
      FALSE)
383 print(corr_Q8_4Q8_10)
384
385 corr_Q8_4Q8_11 <- cor.test(x=constructs$Q8_4, y=constructs$Q8_11, method = 'spearman', exact =
      FALSE)
386 print(corr_Q8_4Q8_11)
387
388
389
390
391
392 corr_Q8_5Q8_6 <- cor.test(x=constructs$Q8_5, y=constructs$Q8_6, method = 'spearman', exact =
      FALSE)
393 print(corr_Q8_5Q8_6)
394
395 corr_Q8_5Q8_7 <- cor.test(x=constructs$Q8_5, y=constructs$Q8_7, method = 'spearman', exact =
      FALSE)
396 print(corr_Q8_5Q8_7)
397
398 corr_Q8_5Q8_8 <- cor.test(x=constructs$Q8_5, y=constructs$Q8_8, method = 'spearman', exact =
      FALSE)
399 print(corr_Q8_5Q8_8)
400
401 corr_Q8_5Q8_9 <- cor.test(x=constructs$Q8_5, y=constructs$Q8_9, method = 'spearman', exact =
      FALSE)
402 print(corr_Q8_5Q8_9)
403

```



```

404 corr_Q8_5Q8_10 <- cor.test(x=constructs$Q8_5, y=constructs$Q8_10, method = 'spearman', exact =
      FALSE) # -0.3666307
405 print(corr_Q8_5Q8_10)
406
407 corr_Q8_5Q8_11 <- cor.test(x=constructs$Q8_5, y=constructs$Q8_11, method = 'spearman', exact =
      FALSE)
408 print(corr_Q8_5Q8_11)
409
410
411
412
413
414 corr_Q8_6Q8_7 <- cor.test(x=constructs$Q8_6, y=constructs$Q8_7, method = 'spearman', exact =
      FALSE)
415 print(corr_Q8_6Q8_7)
416
417 corr_Q8_6Q8_8 <- cor.test(x=constructs$Q8_6, y=constructs$Q8_8, method = 'spearman', exact =
      FALSE)
418 print(corr_Q8_6Q8_8)
419
420 corr_Q8_6Q8_9 <- cor.test(x=constructs$Q8_6, y=constructs$Q8_9, method = 'spearman', exact =
      FALSE)
421 print(corr_Q8_6Q8_9)
422
423 corr_Q8_6Q8_10 <- cor.test(x=constructs$Q8_6, y=constructs$Q8_10, method = 'spearman', exact =
      FALSE) # 0.369301
424 print(corr_Q8_6Q8_10)
425
426 corr_Q8_6Q8_11 <- cor.test(x=constructs$Q8_6, y=constructs$Q8_11, method = 'spearman', exact =
      FALSE)
427 print(corr_Q8_6Q8_11)
428
429
430
431
432
433 corr_Q8_7Q8_8 <- cor.test(x=constructs$Q8_7, y=constructs$Q8_8, method = 'spearman', exact =
      FALSE)
434 print(corr_Q8_7Q8_8)
435
436 corr_Q8_7Q8_9 <- cor.test(x=constructs$Q8_7, y=constructs$Q8_9, method = 'spearman', exact =
      FALSE)
437 print(corr_Q8_7Q8_9)
438
439 corr_Q8_7Q8_10 <- cor.test(x=constructs$Q8_7, y=constructs$Q8_10, method = 'spearman', exact =
      FALSE) # 0.4539196
440 print(corr_Q8_7Q8_10)
441
442 corr_Q8_7Q8_11 <- cor.test(x=constructs$Q8_7, y=constructs$Q8_11, method = 'spearman', exact =
      FALSE)
443 print(corr_Q8_7Q8_11)
444
445
446
447
448 corr_Q8_8Q8_9 <- cor.test(x=constructs$Q8_8, y=constructs$Q8_9, method = 'spearman', exact =
      FALSE)
449 print(corr_Q8_8Q8_9)
450
451 corr_Q8_8Q8_10 <- cor.test(x=constructs$Q8_8, y=constructs$Q8_10, method = 'spearman', exact =
      FALSE) # 0.4446777
452 print(corr_Q8_8Q8_10)
453
454 corr_Q8_8Q8_11 <- cor.test(x=constructs$Q8_8, y=constructs$Q8_11, method = 'spearman', exact =
      FALSE)
455 print(corr_Q8_8Q8_11)
456
457
458
459
460 corr_Q8_9Q8_10 <- cor.test(x=constructs$Q8_9, y=constructs$Q8_10, method = 'spearman', exact =
      FALSE)
461 print(corr_Q8_9Q8_10)

```

```

462
463 corr_Q8_9Q8_11 <- cor.test(x=constructs$Q8_9, y=constructs$Q8_11, method = 'spearman', exact =
      FALSE)
464 print(corr_Q8_9Q8_11)
465
466
467
468 corr_Q8_10Q8_11 <- cor.test(x=constructs$Q8_10, y=constructs$Q8_11, method = 'spearman', exact =
      FALSE) # 0.441014
469 print(corr_Q8_10Q8_11)
470
471 ## conclusion: remove Q8_10
472
473
474
475 # Ability
476 corr_Q9_1Q9_2 <- cor.test(x=constructs$Q9_1, y=constructs$Q9_2, method = 'spearman', exact =
      FALSE)
477 print(corr_Q9_1Q9_2)
478
479 corr_Q9_1Q9_3 <- cor.test(x=constructs$Q9_1, y=constructs$Q9_3, method = 'spearman', exact =
      FALSE)
480 print(corr_Q9_1Q9_3)
481
482 corr_Q9_1Q9_4 <- cor.test(x=constructs$Q9_1, y=constructs$Q9_4, method = 'spearman', exact =
      FALSE)
483 print(corr_Q9_1Q9_4)
484
485 corr_Q9_1Q9_5 <- cor.test(x=constructs$Q9_1, y=constructs$Q9_5, method = 'spearman', exact =
      FALSE)
486 print(corr_Q9_1Q9_5)
487
488 corr_Q9_1Q9_6 <- cor.test(x=constructs$Q9_1, y=constructs$Q9_6, method = 'spearman', exact =
      FALSE) # 0.3681498
489 print(corr_Q9_1Q9_6)
490
491 corr_Q9_1Q9_7 <- cor.test(x=constructs$Q9_1, y=constructs$Q9_7, method = 'spearman', exact =
      FALSE) # -0.7195305
492 print(corr_Q9_1Q9_7)
493
494
495
496 corr_Q9_2Q9_3 <- cor.test(x=constructs$Q9_2, y=constructs$Q9_3, method = 'spearman', exact =
      FALSE)
497 print(corr_Q9_2Q9_3)
498
499 corr_Q9_2Q9_4 <- cor.test(x=constructs$Q9_2, y=constructs$Q9_4, method = 'spearman', exact =
      FALSE)
500 print(corr_Q9_2Q9_4)
501
502 corr_Q9_2Q9_5 <- cor.test(x=constructs$Q9_2, y=constructs$Q9_5, method = 'spearman', exact =
      FALSE)
503 print(corr_Q9_2Q9_5)
504
505 corr_Q9_2Q9_6 <- cor.test(x=constructs$Q9_2, y=constructs$Q9_6, method = 'spearman', exact =
      FALSE) # 0.3089537
506 print(corr_Q9_2Q9_6)
507
508 corr_Q9_2Q9_7 <- cor.test(x=constructs$Q9_2, y=constructs$Q9_7, method = 'spearman', exact =
      FALSE) # -0.6609055
509 print(corr_Q9_2Q9_7)
510
511
512
513 corr_Q9_3Q9_4 <- cor.test(x=constructs$Q9_3, y=constructs$Q9_4, method = 'spearman', exact =
      FALSE)
514 print(corr_Q9_3Q9_4)
515
516 corr_Q9_3Q9_5 <- cor.test(x=constructs$Q9_3, y=constructs$Q9_5, method = 'spearman', exact =
      FALSE)
517 print(corr_Q9_3Q9_5)
518
519 corr_Q9_3Q9_6 <- cor.test(x=constructs$Q9_3, y=constructs$Q9_6, method = 'spearman', exact =

```

```

FALSE) # 0.3750324
520 print(corr_Q9_3Q9_6)
521
522 corr_Q9_3Q9_7 <- cor.test(x=constructs$Q9_3, y=constructs$Q9_7, method = 'spearman', exact =
FALSE) # -0.6512416
523 print(corr_Q9_3Q9_7)
524
525
526
527 corr_Q9_4Q9_5 <- cor.test(x=constructs$Q9_4, y=constructs$Q9_5, method = 'spearman', exact =
FALSE)
528 print(corr_Q9_4Q9_5)
529
530 corr_Q9_4Q9_6 <- cor.test(x=constructs$Q9_4, y=constructs$Q9_6, method = 'spearman', exact =
FALSE) # 0.3827814
531 print(corr_Q9_4Q9_6)
532
533 corr_Q9_4Q9_7 <- cor.test(x=constructs$Q9_4, y=constructs$Q9_7, method = 'spearman', exact =
FALSE) # -0.7352883
534 print(corr_Q9_4Q9_7)
535
536
537
538
539
540 corr_Q9_5Q9_6 <- cor.test(x=constructs$Q9_5, y=constructs$Q9_6, method = 'spearman', exact =
FALSE) # 0.387117
541 print(corr_Q9_5Q9_6)
542
543 corr_Q9_5Q9_7 <- cor.test(x=constructs$Q9_5, y=constructs$Q9_7, method = 'spearman', exact =
FALSE) # -0.7808872
544 print(corr_Q9_5Q9_7)
545
546
547
548
549 corr_Q9_6Q9_7 <- cor.test(x=constructs$Q9_6, y=constructs$Q9_7, method = 'spearman', exact =
FALSE) # -0.4013122
550 print(corr_Q9_6Q9_7)
551
552
553
554 ## Conclusion: remove Q9_6, Q9_7
555
556
557 # Predictability
558 corr_Q11_1Q11_2 <- cor.test(x=constructs$Q11_1, y=constructs$Q11_2, method = 'spearman', exact =
FALSE) # 0.3894545
559 print(corr_Q11_1Q11_2)
560
561 corr_Q11_1Q11_3 <- cor.test(x=constructs$Q11_1, y=constructs$Q11_3, method = 'spearman', exact =
FALSE) # 0.248144
562 print(corr_Q11_1Q11_3)
563
564 corr_Q11_2Q11_3 <- cor.test(x=constructs$Q11_2, y=constructs$Q11_3, method = 'spearman', exact =
FALSE)
565 print(corr_Q11_2Q11_3)
566
567 ## Conclusion: remove Q11_1
568
569
570 # Overall Trust
571 corr_Q11b_1Q11b_2 <- cor.test(x=constructs$Q11b_1, y=constructs$Q11b_2, method = 'spearman',
exact = FALSE)
572 print(corr_Q11b_1Q11b_2)
573
574
575 # Intention to use
576 corr_Q12_1Q12_2 <- cor.test(x=constructs$Q12_1, y=constructs$Q12_2, method = 'spearman', exact =
FALSE)
577 print(corr_Q12_1Q12_2)
578
579 corr_Q12_1Q12_3 <- cor.test(x=constructs$Q12_1, y=constructs$Q12_3, method = 'spearman', exact =

```

```

FALSE)
580 print(corr_Q12_1Q12_3)
581
582 corr_Q12_1Q12_4 <- cor.test(x=constructs$Q12_1, y=constructs$Q12_4, method = 'spearman',exact =
FALSE)
583 print(corr_Q12_1Q12_4)
584
585
586
587 corr_Q12_2Q12_3 <- cor.test(x=constructs$Q12_2, y=constructs$Q12_3, method = 'spearman',exact =
FALSE)
588 print(corr_Q12_2Q12_3)
589
590 corr_Q12_2Q12_4 <- cor.test(x=constructs$Q12_2, y=constructs$Q12_4, method = 'spearman',exact =
FALSE)
591 print(corr_Q12_2Q12_4)
592
593
594 corr_Q12_3Q12_4 <- cor.test(x=constructs$Q12_3, y=constructs$Q12_4, method = 'spearman',exact =
FALSE)
595 print(corr_Q12_3Q12_4)
596
597
598 # Voice preference
599 corr_Q12b_1Q12b_2 <- cor.test(x=constructs$Q12b_1, y=constructs$Q12b_2, method = 'spearman',
exact = FALSE)
600 print(corr_Q12b_1Q12b_2)
601
602
603
604 # Measure internal consistency of all the scales using Cronbach's alpha
605 psych::alpha(dfQ7, check.keys=TRUE)
606 psych::alpha(dfQ8, check.keys=TRUE) # drop 8_10
607 psych::alpha(dfQ9, check.keys=TRUE) # drop 9_6, 9_7
608 psych::alpha(dfQ11, check.keys=TRUE) # drop 11_1
609 psych::alpha(dfQ11b, check.keys=TRUE)
610 psych::alpha(dfQ12, check.keys=TRUE) # drop 12_4
611 psych::alpha(dfQ12b, check.keys=TRUE)
612
613
614
615
616
617
618 ## New relevant dataframe
619 dfQ7 <- data.frame(constructs$Q7_1,constructs$Q7_2,constructs$Q7_3)
620 dfQ8 <- data.frame(constructs$Q8_1,constructs$Q8_2,constructs$Q8_3,constructs$Q8_4,constructs$
Q8_5,constructs$Q8_6,constructs$Q8_7,constructs$Q8_8,constructs$Q8_9,constructs$Q8_11)
621 dfQ9 <- data.frame(constructs$Q9_1,constructs$Q9_2,constructs$Q9_3,constructs$Q9_4,constructs$
Q9_5)
622 dfQ11 <- data.frame(constructs$Q11_2,constructs$Q11_3)
623 dfQ11b <- data.frame(constructs$Q11b_1,constructs$Q11b_2)
624 dfQ12 <- data.frame(constructs$Q12_1,constructs$Q12_2,constructs$Q12_3)
625 dfQ12b <- data.frame(constructs$Q12b_1,constructs$Q12b_2)
626
627
628
629
630 #
631 # mydata <- data.frame(dfQ9)
632 #
633 # KMO(mydata)
634 # bartlett.test(mydata)
635 #
636 #
637 # ev <-eigen(cor(mydata))
638 # nS<-nScree(x=ev$values)
639 # plotnScree(nS,legend = F)
640 #
641 # print(ev$values)
642 #
643 # fit<-factanal(mydata,2,rotation = "varimax")
644 # print(fit,digits = 2, cutoff = 0.3, sort = T)

```

```

645 #
646 #
647 #
648 #
649 #
650 #
651 #
652 #
653 # cor(mydata) # the average correlation needs to be higher than 0.3, or lower than -0.3. Are
        the variables highly correlated among one another, and hence eligible for PCA.
654 # mean(cor(mydata))
655 #
656 # PCA = princomp(mydata)
657 #
658 # PCA$loadings
659 #
660 # PC = PCA$scores
661 # cor(PC)
662
663
664
665
666
667
668 ### Data preparation
669
670 # Process Benevolence scores
671 colnms1=c("Q7_1", "Q7_2", "Q7_3")
672 constructs$Q7_1<-rowSums(constructs[,colnms1])
673
674
675
676 colnms1=c("Q9_1", "Q9_2", "Q9_3", "Q9_4", "Q9_5") # removed Q9_6, Q9_7
677 constructs$Q9_1<-rowSums(constructs[,colnms1])
678
679
680 colnms1=c("Q11_2", "Q11_3") # removed Q11_1
681 constructs$Q11_1<-rowSums(constructs[,colnms1])
682
683
684
685 colnms1=c("Q11b_1", "Q11b_2")
686 constructs$Q11b_1<-rowSums(constructs[,colnms1])
687
688
689 colnms1=c("Q12_1", "Q12_2", "Q12_3") # removed Q12_4
690 constructs$Q12_1<-rowSums(constructs[,colnms1])
691
692 # Intention
693 constructs$intention<-rowSums(constructs[,colnms1])
694
695
696 colnms1=c("Q12b_1", "Q12b_2")
697 constructs$Q12b_1<-rowSums(constructs[,colnms1])
698
699
700
701 colnms1=c("Q8_1", "Q8_2", "Q8_3", "Q8_4", "Q8_5", "Q8_6", "Q8_7", "Q8_8", "Q8_9", "Q8_11") #
        removed Q8_10
702 constructs$Q8_1<-rowSums(constructs[,colnms1])
703
704
705
706 ## Create 1 score for affective and cognitive
707
708 # Affective
709 constructs$affective<-rowSums(constructs[,colnms1])
710
711 # Cognitive
712 colnms1=c("Q9_1", "Q11_1")
713 constructs$cognitive<-rowSums(constructs[,colnms1])
714
715

```

```

716 # Affective + Cognitive
717 colnms1=c("affective", "cognitive")
718 constructs$combined<-rowSums(constructs[,colnms1])
719
720
721
722
723
724
725 # Filter out the rows that we do not need
726 V1<-constructs[constructs$V1Q1=="1",]
727 V1<-constructs[!is.na(constructs$V1Q1), ]
728
729
730 V2<-constructs[constructs$V2Concat=="2",]
731 V2<-constructs[!is.na(constructs$V2Concat), ]
732
733
734 V3<-constructs[constructs$V3Concat=="3",]
735 V3<-constructs[!is.na(constructs$V3Concat), ]
736
737
738
739 ### Demographics
740
741 ## Entire Dataset
742
743 Male<-constructs[constructs$Q14=="1",]
744
745 malemean <- mean(as.numeric(Male$affective),na.rm=TRUE)
746 print(malemean)
747 malesd <- sd(as.numeric(Male$affective), na.rm = TRUE)
748 print(malesd)
749
750
751
752 malemean <- mean(as.numeric(Male$cognitive),na.rm=TRUE)
753 print(malemean)
754 malesd <- sd(as.numeric(Male$cognitive), na.rm = TRUE)
755 print(malesd)
756
757
758 malemean <- mean(as.numeric(Male$intention),na.rm=TRUE)
759 print(malemean)
760 malesd <- sd(as.numeric(Male$intention), na.rm = TRUE)
761 print(malesd)
762
763
764
765 Female<-constructs[constructs$Q14=="2",]
766
767
768 femalemean <- mean(as.numeric(Female$affective),na.rm=TRUE)
769 print(femalemean)
770 femalesd <- sd(as.numeric(Female$affective), na.rm = TRUE)
771 print(femalesd)
772
773
774
775 femalemean <- mean(as.numeric(Female$cognitive),na.rm=TRUE)
776 print(femalemean)
777 femalesd <- sd(as.numeric(Female$cognitive), na.rm = TRUE)
778 print(femalesd)
779
780
781 femalemean <- mean(as.numeric(Female$intention),na.rm=TRUE)
782 print(femalemean)
783 femalesd <- sd(as.numeric(Female$intention), na.rm = TRUE)
784 print(femalesd)
785
786
787
788

```

```

789 Other<-constructs[constructs$Q14=="3",]
790
791
792 othermean <- mean(as.numeric(Other$affective),na.rm=TRUE)
793 print(othermean)
794 othersd <- sd(as.numeric(Other$affective), na.rm = TRUE)
795 print(othersd)
796
797
798
799 othermean <- mean(as.numeric(Other$cognitive),na.rm=TRUE)
800 print(othermean)
801 othersd <- sd(as.numeric(Other$cognitive), na.rm = TRUE)
802 print(othersd)
803
804
805 othermean <- mean(as.numeric(Other$intention),na.rm=TRUE)
806 print(othermean)
807 othersd <- sd(as.numeric(Other$intention), na.rm = TRUE)
808 print(othersd)
809
810
811
812
813
814 age1<-constructs[constructs$Q13=="2",]
815
816
817 age1mean <- mean(as.numeric(age1$affective),na.rm=TRUE)
818 print(age1mean)
819 age1sd <- sd(as.numeric(age1$affective), na.rm = TRUE)
820 print(age1sd)
821
822
823
824 age1mean <- mean(as.numeric(age1$cognitive),na.rm=TRUE)
825 print(age1mean)
826 age1sd <- sd(as.numeric(age1$cognitive), na.rm = TRUE)
827 print(age1sd)
828
829
830 age1mean <- mean(as.numeric(age1$intention),na.rm=TRUE)
831 print(age1mean)
832 age1sd <- sd(as.numeric(age1$intention), na.rm = TRUE)
833 print(age1sd)
834
835
836 age2<-constructs[constructs$Q13=="3",]
837
838
839 age2mean <- mean(as.numeric(age2$affective),na.rm=TRUE)
840 print(age2mean)
841 age2sd <- sd(as.numeric(age2$affective), na.rm = TRUE)
842 print(age2sd)
843
844
845
846 age2mean <- mean(as.numeric(age2$cognitive),na.rm=TRUE)
847 print(age2mean)
848 age2sd <- sd(as.numeric(age2$cognitive), na.rm = TRUE)
849 print(age2sd)
850
851
852 age2mean <- mean(as.numeric(age2$intention),na.rm=TRUE)
853 print(age2mean)
854 age2sd <- sd(as.numeric(age2$intention), na.rm = TRUE)
855 print(age2sd)
856
857
858
859
860 ed1<-constructs[constructs$Q15=="1",]
861

```

```

862
863 ed1mean <- mean(as.numeric(ed1$affective),na.rm=TRUE)
864 print(ed1mean)
865 ed1sd <- sd(as.numeric(ed1$affective), na.rm = TRUE)
866 print(ed1sd)
867
868
869
870 ed1mean <- mean(as.numeric(ed1$cognitive),na.rm=TRUE)
871 print(ed1mean)
872 ed1sd <- sd(as.numeric(ed1$cognitive), na.rm = TRUE)
873 print(ed1sd)
874
875
876 ed1mean <- mean(as.numeric(ed1$intention),na.rm=TRUE)
877 print(ed1mean)
878 ed1sd <- sd(as.numeric(ed1$intention), na.rm = TRUE)
879 print(ed1sd)
880
881
882
883
884 ed2<-constructs[constructs$Q15=="2",]
885
886
887 ed2mean <- mean(as.numeric(ed2$affective),na.rm=TRUE)
888 print(ed2mean)
889 ed2sd <- sd(as.numeric(ed2$affective), na.rm = TRUE)
890 print(ed2sd)
891
892
893
894 ed2mean <- mean(as.numeric(ed2$cognitive),na.rm=TRUE)
895 print(ed2mean)
896 ed2sd <- sd(as.numeric(ed2$cognitive), na.rm = TRUE)
897 print(ed2sd)
898
899
900 ed2mean <- mean(as.numeric(ed2$intention),na.rm=TRUE)
901 print(ed2mean)
902 ed2sd <- sd(as.numeric(ed2$intention), na.rm = TRUE)
903 print(ed2sd)
904
905
906
907 ed3<-constructs[constructs$Q15=="3",]
908
909
910 ed3mean <- mean(as.numeric(ed3$affective),na.rm=TRUE)
911 print(ed3mean)
912 ed3sd <- sd(as.numeric(ed3$affective), na.rm = TRUE)
913 print(ed3sd)
914
915
916
917 ed3mean <- mean(as.numeric(ed3$cognitive),na.rm=TRUE)
918 print(ed3mean)
919 ed3sd <- sd(as.numeric(ed3$cognitive), na.rm = TRUE)
920 print(ed3sd)
921
922
923 ed3mean <- mean(as.numeric(ed3$intention),na.rm=TRUE)
924 print(ed3mean)
925 ed3sd <- sd(as.numeric(ed3$intention), na.rm = TRUE)
926 print(ed3sd)
927
928
929
930 ed4<-constructs[constructs$Q15=="4",]
931
932
933 ed4mean <- mean(as.numeric(ed4$affective),na.rm=TRUE)
934 print(ed4mean)

```



```

935 ed4sd <- sd(as.numeric(ed4$affective), na.rm = TRUE)
936 print(ed4sd)
937
938
939
940 ed4mean <- mean(as.numeric(ed4$cognitive),na.rm=TRUE)
941 print(ed4mean)
942 ed4sd <- sd(as.numeric(ed4$cognitive), na.rm = TRUE)
943 print(ed4sd)
944
945
946 ed4mean <- mean(as.numeric(ed4$intention),na.rm=TRUE)
947 print(ed4mean)
948 ed4sd <- sd(as.numeric(ed4$intention), na.rm = TRUE)
949 print(ed4sd)
950
951
952
953 ed5<-constructs[constructs$Q15=="5",]
954
955
956 ed5mean <- mean(as.numeric(ed5$affective),na.rm=TRUE)
957 print(ed5mean)
958 ed5sd <- sd(as.numeric(ed5$affective), na.rm = TRUE)
959 print(ed5sd)
960
961
962
963 ed5mean <- mean(as.numeric(ed5$cognitive),na.rm=TRUE)
964 print(ed5mean)
965 ed5sd <- sd(as.numeric(ed5$cognitive), na.rm = TRUE)
966 print(ed5sd)
967
968
969 ed5mean <- mean(as.numeric(ed5$intention),na.rm=TRUE)
970 print(ed5mean)
971 ed5sd <- sd(as.numeric(ed5$intention), na.rm = TRUE)
972 print(ed5sd)
973
974
975
976
977
978 ## V1
979
980 Male<-V1[V1$Q14=="1",]
981
982 malemean <- mean(as.numeric(Male$affective),na.rm=TRUE)
983 print(malemean)
984 malesd <- sd(as.numeric(Male$affective), na.rm = TRUE)
985 print(malesd)
986
987
988
989 malemean <- mean(as.numeric(Male$cognitive),na.rm=TRUE)
990 print(malemean)
991 malesd <- sd(as.numeric(Male$cognitive), na.rm = TRUE)
992 print(malesd)
993
994
995 malemean <- mean(as.numeric(Male$intention),na.rm=TRUE)
996 print(malemean)
997 malesd <- sd(as.numeric(Male$intention), na.rm = TRUE)
998 print(malesd)
999
1000
1001
1002
1003
1004 Female<-V1[V1$Q14=="2",]
1005
1006
1007 femalemean <- mean(as.numeric(Female$affective),na.rm=TRUE)

```

```

1008 print(femalemean)
1009 femalesd <- sd(as.numeric(Female$affective), na.rm = TRUE)
1010 print(femalesd)
1011
1012
1013
1014 femalemean <- mean(as.numeric(Female$cognitive),na.rm=TRUE)
1015 print(femalemean)
1016 femalesd <- sd(as.numeric(Female$cognitive), na.rm = TRUE)
1017 print(femalesd)
1018
1019
1020 femalemean <- mean(as.numeric(Female$intention),na.rm=TRUE)
1021 print(femalemean)
1022 femalesd <- sd(as.numeric(Female$intention), na.rm = TRUE)
1023 print(femalesd)
1024
1025
1026
1027
1028 Other<-V1[V1$Q14=="3",]
1029
1030
1031 othermean <- mean(as.numeric(Other$affective),na.rm=TRUE)
1032 print(othermean)
1033 othersd <- sd(as.numeric(Other$affective), na.rm = TRUE)
1034 print(othersd)
1035
1036
1037
1038 othermean <- mean(as.numeric(Other$cognitive),na.rm=TRUE)
1039 print(othermean)
1040 othersd <- sd(as.numeric(Other$cognitive), na.rm = TRUE)
1041 print(othersd)
1042
1043
1044 othermean <- mean(as.numeric(Other$intention),na.rm=TRUE)
1045 print(othermean)
1046 othersd <- sd(as.numeric(Other$intention), na.rm = TRUE)
1047 print(othersd)
1048
1049
1050
1051
1052
1053 age1<-V1[V1$Q13=="2",]
1054
1055
1056 age1mean <- mean(as.numeric(age1$affective),na.rm=TRUE)
1057 print(age1mean)
1058 age1sd <- sd(as.numeric(age1$affective), na.rm = TRUE)
1059 print(age1sd)
1060
1061
1062
1063 age1mean <- mean(as.numeric(age1$cognitive),na.rm=TRUE)
1064 print(age1mean)
1065 age1sd <- sd(as.numeric(age1$cognitive), na.rm = TRUE)
1066 print(age1sd)
1067
1068
1069 age1mean <- mean(as.numeric(age1$intention),na.rm=TRUE)
1070 print(age1mean)
1071 age1sd <- sd(as.numeric(age1$intention), na.rm = TRUE)
1072 print(age1sd)
1073
1074
1075 age2<-V1[V1$Q13=="3",]
1076
1077
1078 age2mean <- mean(as.numeric(age2$affective),na.rm=TRUE)
1079 print(age2mean)
1080 age2sd <- sd(as.numeric(age2$affective), na.rm = TRUE)

```

```

1081 print(age2sd)
1082
1083
1084
1085 age2mean <- mean(as.numeric(age2$cognitive),na.rm=TRUE)
1086 print(age2mean)
1087 age2sd <- sd(as.numeric(age2$cognitive), na.rm = TRUE)
1088 print(age2sd)
1089
1090
1091 age2mean <- mean(as.numeric(age2$intention),na.rm=TRUE)
1092 print(age2mean)
1093 age2sd <- sd(as.numeric(age2$intention), na.rm = TRUE)
1094 print(age2sd)
1095
1096
1097
1098
1099 ed1<-V1[V1$Q15=="1",]
1100
1101
1102 ed1mean <- mean(as.numeric(ed1$affective),na.rm=TRUE)
1103 print(ed1mean)
1104 ed1sd <- sd(as.numeric(ed1$affective), na.rm = TRUE)
1105 print(ed1sd)
1106
1107
1108
1109 ed1mean <- mean(as.numeric(ed1$cognitive),na.rm=TRUE)
1110 print(ed1mean)
1111 ed1sd <- sd(as.numeric(ed1$cognitive), na.rm = TRUE)
1112 print(ed1sd)
1113
1114
1115 ed1mean <- mean(as.numeric(ed1$intention),na.rm=TRUE)
1116 print(ed1mean)
1117 ed1sd <- sd(as.numeric(ed1$intention), na.rm = TRUE)
1118 print(ed1sd)
1119
1120
1121
1122
1123 ed2<-V1[V1$Q15=="2",]
1124
1125
1126 ed2mean <- mean(as.numeric(ed2$affective),na.rm=TRUE)
1127 print(ed2mean)
1128 ed2sd <- sd(as.numeric(ed2$affective), na.rm = TRUE)
1129 print(ed2sd)
1130
1131
1132
1133 ed2mean <- mean(as.numeric(ed2$cognitive),na.rm=TRUE)
1134 print(ed2mean)
1135 ed2sd <- sd(as.numeric(ed2$cognitive), na.rm = TRUE)
1136 print(ed2sd)
1137
1138
1139 ed2mean <- mean(as.numeric(ed2$intention),na.rm=TRUE)
1140 print(ed2mean)
1141 ed2sd <- sd(as.numeric(ed2$intention), na.rm = TRUE)
1142 print(ed2sd)
1143
1144
1145
1146 ed3<-V1[V1$Q15=="3",]
1147
1148
1149 ed3mean <- mean(as.numeric(ed3$affective),na.rm=TRUE)
1150 print(ed3mean)
1151 ed3sd <- sd(as.numeric(ed3$affective), na.rm = TRUE)
1152 print(ed3sd)
1153

```

```

1154
1155
1156 ed3mean <- mean(as.numeric(ed3$cognitive),na.rm=TRUE)
1157 print(ed3mean)
1158 ed3sd <- sd(as.numeric(ed3$cognitive), na.rm = TRUE)
1159 print(ed3sd)
1160
1161
1162 ed3mean <- mean(as.numeric(ed3$intention),na.rm=TRUE)
1163 print(ed3mean)
1164 ed3sd <- sd(as.numeric(ed3$intention), na.rm = TRUE)
1165 print(ed3sd)
1166
1167
1168
1169 ed4<-V1[V1$Q15=="4",]
1170
1171
1172 ed4mean <- mean(as.numeric(ed4$affective),na.rm=TRUE)
1173 print(ed4mean)
1174 ed4sd <- sd(as.numeric(ed4$affective), na.rm = TRUE)
1175 print(ed4sd)
1176
1177
1178
1179 ed4mean <- mean(as.numeric(ed4$cognitive),na.rm=TRUE)
1180 print(ed4mean)
1181 ed4sd <- sd(as.numeric(ed4$cognitive), na.rm = TRUE)
1182 print(ed4sd)
1183
1184
1185 ed4mean <- mean(as.numeric(ed4$intention),na.rm=TRUE)
1186 print(ed4mean)
1187 ed4sd <- sd(as.numeric(ed4$intention), na.rm = TRUE)
1188 print(ed4sd)
1189
1190
1191
1192 ed5<-V1[V1$Q15=="5",]
1193
1194
1195 ed5mean <- mean(as.numeric(ed5$affective),na.rm=TRUE)
1196 print(ed5mean)
1197 ed5sd <- sd(as.numeric(ed5$affective), na.rm = TRUE)
1198 print(ed5sd)
1199
1200
1201
1202 ed5mean <- mean(as.numeric(ed5$cognitive),na.rm=TRUE)
1203 print(ed5mean)
1204 ed5sd <- sd(as.numeric(ed5$cognitive), na.rm = TRUE)
1205 print(ed5sd)
1206
1207
1208 ed5mean <- mean(as.numeric(ed5$intention),na.rm=TRUE)
1209 print(ed5mean)
1210 ed5sd <- sd(as.numeric(ed5$intention), na.rm = TRUE)
1211 print(ed5sd)
1212
1213
1214
1215
1216
1217
1218 ## V2
1219
1220 Male<-V2[V2$Q14=="1",]
1221
1222 malemean <- mean(as.numeric(Male$affective),na.rm=TRUE)
1223 print(malemean)
1224 malesd <- sd(as.numeric(Male$affective), na.rm = TRUE)
1225 print(malesd)
1226

```

```

1227
1228
1229 malemean <- mean(as.numeric(Male$cognitive),na.rm=TRUE)
1230 print(malemean)
1231 malesd <- sd(as.numeric(Male$cognitive), na.rm = TRUE)
1232 print(malesd)
1233
1234
1235 malemean <- mean(as.numeric(Male$intention),na.rm=TRUE)
1236 print(malemean)
1237 malesd <- sd(as.numeric(Male$intention), na.rm = TRUE)
1238 print(malesd)
1239
1240
1241
1242 Female<-V2[V2$Q14=="2",]
1243
1244
1245 femalemean <- mean(as.numeric(Female$affective),na.rm=TRUE)
1246 print(femalemean)
1247 femalesd <- sd(as.numeric(Female$affective), na.rm = TRUE)
1248 print(femalesd)
1249
1250
1251
1252 femalemean <- mean(as.numeric(Female$cognitive),na.rm=TRUE)
1253 print(femalemean)
1254 femalesd <- sd(as.numeric(Female$cognitive), na.rm = TRUE)
1255 print(femalesd)
1256
1257
1258 femalemean <- mean(as.numeric(Female$intention),na.rm=TRUE)
1259 print(femalemean)
1260 femalesd <- sd(as.numeric(Female$intention), na.rm = TRUE)
1261 print(femalesd)
1262
1263
1264
1265
1266 Other<-V2[V2$Q14=="3",]
1267
1268
1269 othermean <- mean(as.numeric(Other$affective),na.rm=TRUE)
1270 print(othermean)
1271 othersd <- sd(as.numeric(Other$affective), na.rm = TRUE)
1272 print(othersd)
1273
1274
1275
1276 othermean <- mean(as.numeric(Other$cognitive),na.rm=TRUE)
1277 print(othermean)
1278 othersd <- sd(as.numeric(Other$cognitive), na.rm = TRUE)
1279 print(othersd)
1280
1281
1282 othermean <- mean(as.numeric(Other$intention),na.rm=TRUE)
1283 print(othermean)
1284 othersd <- sd(as.numeric(Other$intention), na.rm = TRUE)
1285 print(othersd)
1286
1287
1288
1289
1290
1291 age1<-V2[V2$Q13=="2",]
1292
1293
1294 age1mean <- mean(as.numeric(age1$affective),na.rm=TRUE)
1295 print(age1mean)
1296 age1sd <- sd(as.numeric(age1$affective), na.rm = TRUE)
1297 print(age1sd)
1298
1299

```

```

1300
1301 age1mean <- mean(as.numeric(age1$cognitive),na.rm=TRUE)
1302 print(age1mean)
1303 age1sd <- sd(as.numeric(age1$cognitive), na.rm = TRUE)
1304 print(age1sd)
1305
1306
1307 age1mean <- mean(as.numeric(age1$intention),na.rm=TRUE)
1308 print(age1mean)
1309 age1sd <- sd(as.numeric(age1$intention), na.rm = TRUE)
1310 print(age1sd)
1311
1312
1313 age2<-V2[V2$Q13=="3",]
1314
1315
1316 age2mean <- mean(as.numeric(age2$affective),na.rm=TRUE)
1317 print(age2mean)
1318 age2sd <- sd(as.numeric(age2$affective), na.rm = TRUE)
1319 print(age2sd)
1320
1321
1322
1323 age2mean <- mean(as.numeric(age2$cognitive),na.rm=TRUE)
1324 print(age2mean)
1325 age2sd <- sd(as.numeric(age2$cognitive), na.rm = TRUE)
1326 print(age2sd)
1327
1328
1329 age2mean <- mean(as.numeric(age2$intention),na.rm=TRUE)
1330 print(age2mean)
1331 age2sd <- sd(as.numeric(age2$intention), na.rm = TRUE)
1332 print(age2sd)
1333
1334
1335
1336
1337 ed1<-V2[V2$Q15=="1",]
1338
1339
1340 ed1mean <- mean(as.numeric(ed1$affective),na.rm=TRUE)
1341 print(ed1mean)
1342 ed1sd <- sd(as.numeric(ed1$affective), na.rm = TRUE)
1343 print(ed1sd)
1344
1345
1346
1347 ed1mean <- mean(as.numeric(ed1$cognitive),na.rm=TRUE)
1348 print(ed1mean)
1349 ed1sd <- sd(as.numeric(ed1$cognitive), na.rm = TRUE)
1350 print(ed1sd)
1351
1352
1353 ed1mean <- mean(as.numeric(ed1$intention),na.rm=TRUE)
1354 print(ed1mean)
1355 ed1sd <- sd(as.numeric(ed1$intention), na.rm = TRUE)
1356 print(ed1sd)
1357
1358
1359
1360
1361 ed2<-V2[V2$Q15=="2",]
1362
1363
1364 ed2mean <- mean(as.numeric(ed2$affective),na.rm=TRUE)
1365 print(ed2mean)
1366 ed2sd <- sd(as.numeric(ed2$affective), na.rm = TRUE)
1367 print(ed2sd)
1368
1369
1370
1371 ed2mean <- mean(as.numeric(ed2$cognitive),na.rm=TRUE)
1372 print(ed2mean)

```

```

1373 ed2sd <- sd(as.numeric(ed2$cognitive), na.rm = TRUE)
1374 print(ed2sd)
1375
1376
1377 ed2mean <- mean(as.numeric(ed2$intention), na.rm=TRUE)
1378 print(ed2mean)
1379 ed2sd <- sd(as.numeric(ed2$intention), na.rm = TRUE)
1380 print(ed2sd)
1381
1382
1383
1384 ed3<-V2[V2$Q15=="3",]
1385
1386
1387 ed3mean <- mean(as.numeric(ed3$affective), na.rm=TRUE)
1388 print(ed3mean)
1389 ed3sd <- sd(as.numeric(ed3$affective), na.rm = TRUE)
1390 print(ed3sd)
1391
1392
1393
1394 ed3mean <- mean(as.numeric(ed3$cognitive), na.rm=TRUE)
1395 print(ed3mean)
1396 ed3sd <- sd(as.numeric(ed3$cognitive), na.rm = TRUE)
1397 print(ed3sd)
1398
1399
1400 ed3mean <- mean(as.numeric(ed3$intention), na.rm=TRUE)
1401 print(ed3mean)
1402 ed3sd <- sd(as.numeric(ed3$intention), na.rm = TRUE)
1403 print(ed3sd)
1404
1405
1406
1407 ed4<-V2[V2$Q15=="4",]
1408
1409
1410 ed4mean <- mean(as.numeric(ed4$affective), na.rm=TRUE)
1411 print(ed4mean)
1412 ed4sd <- sd(as.numeric(ed4$affective), na.rm = TRUE)
1413 print(ed4sd)
1414
1415
1416
1417 ed4mean <- mean(as.numeric(ed4$cognitive), na.rm=TRUE)
1418 print(ed4mean)
1419 ed4sd <- sd(as.numeric(ed4$cognitive), na.rm = TRUE)
1420 print(ed4sd)
1421
1422
1423 ed4mean <- mean(as.numeric(ed4$intention), na.rm=TRUE)
1424 print(ed4mean)
1425 ed4sd <- sd(as.numeric(ed4$intention), na.rm = TRUE)
1426 print(ed4sd)
1427
1428
1429
1430 ed5<-V2[V2$Q15=="5",]
1431
1432
1433 ed5mean <- mean(as.numeric(ed5$affective), na.rm=TRUE)
1434 print(ed5mean)
1435 ed5sd <- sd(as.numeric(ed5$affective), na.rm = TRUE)
1436 print(ed5sd)
1437
1438
1439
1440 ed5mean <- mean(as.numeric(ed5$cognitive), na.rm=TRUE)
1441 print(ed5mean)
1442 ed5sd <- sd(as.numeric(ed5$cognitive), na.rm = TRUE)
1443 print(ed5sd)
1444
1445

```

```

1446 ed5mean <- mean(as.numeric(ed5$intention),na.rm=TRUE)
1447 print(ed5mean)
1448 ed5sd <- sd(as.numeric(ed5$intention), na.rm = TRUE)
1449 print(ed5sd)
1450
1451
1452
1453
1454 ## V3
1455
1456 Male<-V3[V3$Q14=="1",]
1457
1458 malemean <- mean(as.numeric(Male$affective),na.rm=TRUE)
1459 print(malemean)
1460 malesd <- sd(as.numeric(Male$affective), na.rm = TRUE)
1461 print(malesd)
1462
1463
1464
1465 malemean <- mean(as.numeric(Male$cognitive),na.rm=TRUE)
1466 print(malemean)
1467 malesd <- sd(as.numeric(Male$cognitive), na.rm = TRUE)
1468 print(malesd)
1469
1470
1471 malemean <- mean(as.numeric(Male$intention),na.rm=TRUE)
1472 print(malemean)
1473 malesd <- sd(as.numeric(Male$intention), na.rm = TRUE)
1474 print(malesd)
1475
1476
1477
1478 Female<-V3[V3$Q14=="2",]
1479
1480
1481 femalemean <- mean(as.numeric(Female$affective),na.rm=TRUE)
1482 print(femalemean)
1483 femalesd <- sd(as.numeric(Female$affective), na.rm = TRUE)
1484 print(femalesd)
1485
1486
1487
1488 femalemean <- mean(as.numeric(Female$cognitive),na.rm=TRUE)
1489 print(femalemean)
1490 femalesd <- sd(as.numeric(Female$cognitive), na.rm = TRUE)
1491 print(femalesd)
1492
1493
1494 femalemean <- mean(as.numeric(Female$intention),na.rm=TRUE)
1495 print(femalemean)
1496 femalesd <- sd(as.numeric(Female$intention), na.rm = TRUE)
1497 print(femalesd)
1498
1499
1500
1501
1502 Other<-V3[V3$Q14=="3",]
1503
1504
1505 othermean <- mean(as.numeric(Other$affective),na.rm=TRUE)
1506 print(othermean)
1507 othersd <- sd(as.numeric(Other$affective), na.rm = TRUE)
1508 print(othersd)
1509
1510
1511
1512 othermean <- mean(as.numeric(Other$cognitive),na.rm=TRUE)
1513 print(othermean)
1514 othersd <- sd(as.numeric(Other$cognitive), na.rm = TRUE)
1515 print(othersd)
1516
1517
1518 othermean <- mean(as.numeric(Other$intention),na.rm=TRUE)

```



```

1519 print(othermean)
1520 othersd <- sd(as.numeric(Other$intention), na.rm = TRUE)
1521 print(othersd)
1522
1523
1524
1525
1526
1527 age1<-V3[V3$Q13=="2",]
1528
1529
1530 age1mean <- mean(as.numeric(age1$affective),na.rm=TRUE)
1531 print(age1mean)
1532 age1sd <- sd(as.numeric(age1$affective), na.rm = TRUE)
1533 print(age1sd)
1534
1535
1536
1537 age1mean <- mean(as.numeric(age1$cognitive),na.rm=TRUE)
1538 print(age1mean)
1539 age1sd <- sd(as.numeric(age1$cognitive), na.rm = TRUE)
1540 print(age1sd)
1541
1542
1543 age1mean <- mean(as.numeric(age1$intention),na.rm=TRUE)
1544 print(age1mean)
1545 age1sd <- sd(as.numeric(age1$intention), na.rm = TRUE)
1546 print(age1sd)
1547
1548
1549 age2<-V3[V3$Q13=="3",]
1550
1551
1552 age2mean <- mean(as.numeric(age2$affective),na.rm=TRUE)
1553 print(age2mean)
1554 age2sd <- sd(as.numeric(age2$affective), na.rm = TRUE)
1555 print(age2sd)
1556
1557
1558
1559 age2mean <- mean(as.numeric(age2$cognitive),na.rm=TRUE)
1560 print(age2mean)
1561 age2sd <- sd(as.numeric(age2$cognitive), na.rm = TRUE)
1562 print(age2sd)
1563
1564
1565 age2mean <- mean(as.numeric(age2$intention),na.rm=TRUE)
1566 print(age2mean)
1567 age2sd <- sd(as.numeric(age2$intention), na.rm = TRUE)
1568 print(age2sd)
1569
1570
1571
1572
1573 ed1<-V3[V3$Q15=="1",]
1574
1575
1576 ed1mean <- mean(as.numeric(ed1$affective),na.rm=TRUE)
1577 print(ed1mean)
1578 ed1sd <- sd(as.numeric(ed1$affective), na.rm = TRUE)
1579 print(ed1sd)
1580
1581
1582
1583 ed1mean <- mean(as.numeric(ed1$cognitive),na.rm=TRUE)
1584 print(ed1mean)
1585 ed1sd <- sd(as.numeric(ed1$cognitive), na.rm = TRUE)
1586 print(ed1sd)
1587
1588
1589 ed1mean <- mean(as.numeric(ed1$intention),na.rm=TRUE)
1590 print(ed1mean)
1591 ed1sd <- sd(as.numeric(ed1$intention), na.rm = TRUE)

```

```

1592 print(ed1sd)
1593
1594
1595
1596
1597 ed2<-V3[V3$Q15=="2",]
1598
1599
1600 ed2mean <- mean(as.numeric(ed2$affective),na.rm=TRUE)
1601 print(ed2mean)
1602 ed2sd <- sd(as.numeric(ed2$affective), na.rm = TRUE)
1603 print(ed2sd)
1604
1605
1606
1607 ed2mean <- mean(as.numeric(ed2$cognitive),na.rm=TRUE)
1608 print(ed2mean)
1609 ed2sd <- sd(as.numeric(ed2$cognitive), na.rm = TRUE)
1610 print(ed2sd)
1611
1612
1613 ed2mean <- mean(as.numeric(ed2$intention),na.rm=TRUE)
1614 print(ed2mean)
1615 ed2sd <- sd(as.numeric(ed2$intention), na.rm = TRUE)
1616 print(ed2sd)
1617
1618
1619
1620 ed3<-V3[V3$Q15=="3",]
1621
1622
1623 ed3mean <- mean(as.numeric(ed3$affective),na.rm=TRUE)
1624 print(ed3mean)
1625 ed3sd <- sd(as.numeric(ed3$affective), na.rm = TRUE)
1626 print(ed3sd)
1627
1628
1629
1630 ed3mean <- mean(as.numeric(ed3$cognitive),na.rm=TRUE)
1631 print(ed3mean)
1632 ed3sd <- sd(as.numeric(ed3$cognitive), na.rm = TRUE)
1633 print(ed3sd)
1634
1635
1636 ed3mean <- mean(as.numeric(ed3$intention),na.rm=TRUE)
1637 print(ed3mean)
1638 ed3sd <- sd(as.numeric(ed3$intention), na.rm = TRUE)
1639 print(ed3sd)
1640
1641
1642
1643 ed4<-V3[V3$Q15=="4",]
1644
1645
1646 ed4mean <- mean(as.numeric(ed4$affective),na.rm=TRUE)
1647 print(ed4mean)
1648 ed4sd <- sd(as.numeric(ed4$affective), na.rm = TRUE)
1649 print(ed4sd)
1650
1651
1652
1653 ed4mean <- mean(as.numeric(ed4$cognitive),na.rm=TRUE)
1654 print(ed4mean)
1655 ed4sd <- sd(as.numeric(ed4$cognitive), na.rm = TRUE)
1656 print(ed4sd)
1657
1658
1659 ed4mean <- mean(as.numeric(ed4$intention),na.rm=TRUE)
1660 print(ed4mean)
1661 ed4sd <- sd(as.numeric(ed4$intention), na.rm = TRUE)
1662 print(ed4sd)
1663
1664

```

```

1665
1666 ed5<-V3[V3$Q15=="5",]
1667
1668
1669 ed5mean <- mean(as.numeric(ed5$affective),na.rm=TRUE)
1670 print(ed5mean)
1671 ed5sd <- sd(as.numeric(ed5$affective), na.rm = TRUE)
1672 print(ed5sd)
1673
1674
1675
1676 ed5mean <- mean(as.numeric(ed5$cognitive),na.rm=TRUE)
1677 print(ed5mean)
1678 ed5sd <- sd(as.numeric(ed5$cognitive), na.rm = TRUE)
1679 print(ed5sd)
1680
1681
1682 ed5mean <- mean(as.numeric(ed5$intention),na.rm=TRUE)
1683 print(ed5mean)
1684 ed5sd <- sd(as.numeric(ed5$intention), na.rm = TRUE)
1685 print(ed5sd)
1686
1687
1688
1689
1690
1691 AllAgeMean <- mean(as.numeric(constructs$Q13),na.rm=TRUE)
1692 print(AllAgeMean)
1693 AllAgesd <- sd(as.numeric(constructs$Q13), na.rm = TRUE)
1694 print(AllAgesd)
1695 AllAgeMedian <- median(as.numeric(constructs$Q13), na.rm = TRUE)
1696 print(AllAgeMedian)
1697
1698
1699
1700 AllGenderMean <- mean(as.numeric(constructs$Q14),na.rm=TRUE)
1701 print(AllGenderMean)
1702 AllGendersd <- sd(as.numeric(constructs$Q14), na.rm = TRUE)
1703 print(AllGendersd)
1704 AllGenderMedian <- median(as.numeric(constructs$Q14), na.rm = TRUE)
1705 print(AllGenderMedian)
1706
1707
1708 AllEducationMean <- mean(as.numeric(constructs$Q15),na.rm=TRUE)
1709 print(AllEducationMean)
1710 AllEducationsd <- sd(as.numeric(constructs$Q15), na.rm = TRUE)
1711 print(AllEducationsd)
1712 AllEducationMedian <- median(as.numeric(constructs$Q15), na.rm = TRUE)
1713 print(AllEducationMedian)
1714
1715
1716
1717
1718 table <- table(constructs$Country_of_Residence)
1719
1720
1721 print(table)
1722 # Use barplot to visualize
1723 # a frequency table in a graphic
1724 barplot(table)
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735 ### Stratify check V2
1736 ## Formula: 400-NA
1737 constructs %>%

```

```

1738     count(constructs$V2F1SQ1) # 43
1739
1740 constructs %>%
1741     count(constructs$V2F1FQ1) # 51
1742
1743 constructs %>%
1744     count(constructs$V2M1SQ1) # 51
1745
1746 constructs %>%
1747     count(constructs$V2M1FQ1) # 54
1748
1749
1750
1751
1752
1753 ### Gender voice preference of opposite gender ###
1754
1755
1756 Female<-constructs[constructs$V3Female=="2",]
1757 Female<-constructs[!is.na(constructs$V3Female), ]
1758
1759
1760 Male<-constructs[constructs$V3Male=="2",]
1761 Male<-constructs[!is.na(constructs$V3Male), ]
1762
1763
1764
1765
1766
1767
1768 Male_Male<-Male[Male$Q14=="1",] # 39 Je bent man en kiest voor mannen stem
1769 Female_Male<-Female[Female$Q14=="1",] # 59 Je bent man en kiest voor vrouwen stem
1770
1771
1772 Male_Female<-Male[Male$Q14=="2",] # 39 Je bent vrouw en kiest voor mannen stem
1773 Female_Female<-Female[Female$Q14=="2",] # 56 Je bent een vrouw en kiest voor vrouwen stem
1774
1775 # In totaal 193 mensen, waarvan 98 man, 95 vrouw.
1776
1777
1778 # 39/95 = 0.41105 = 41.1% van de vrouwen kiest voor een mannen stem.
1779 # 56/95 = 0.58947 = 58.9% van de vrouwen kiest voor een vrouwen stem.
1780 # Het is niet waar voor vrouwen.
1781
1782
1783 # 39/98 = 0.39796 = 39.8% van de mannen kiest voor een mannen stem.
1784 # 59/98 = 0.60204 = 60.2% van de mannen kiest voor een vrouwen stem.
1785 # Het is waar voor mannen.
1786
1787
1788
1789
1790 # (56+39)/193 = 0.4922279 = 49.2% kiest voor zelfde geslacht.
1791 # (39+59)/193 = 0.5077720 = 50.1% kiest voor opposite geslacht.
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803 # report the smallest test statistics (w value)
1804
1805 # Cognitive Trust
1806
1807 wilcox.test(V1$cognitive, V2$cognitive, paired = FALSE)
1808 wilcox.test(V2$cognitive, V1$cognitive, paired = FALSE)
1809
1810

```

```

1811 wilcox.test(V1$cognitive, V3$cognitive, paired = FALSE)
1812 wilcox.test(V3$cognitive, V1$cognitive, paired = FALSE) # 0.03295
1813
1814 wilcox.test(V2$cognitive, V3$cognitive, paired = FALSE)
1815 wilcox.test(V3$cognitive, V2$cognitive, paired = FALSE)
1816
1817
1818 mean(V1$cognitive)
1819 mean(V2$cognitive)
1820 mean(V3$cognitive)
1821
1822
1823 sd(V1$cognitive)
1824 sd(V2$cognitive)
1825 sd(V3$cognitive)
1826
1827
1828 # Cohen's D: V1V3 | 0.197068
1829
1830
1831 totaleffect <- clm(as.factor(Q11b_1)~AllConcat, data = constructs, link="logit")
1832
1833 summary(totaleffect)
1834 confint(totaleffect)
1835
1836 stargazer(totaleffect, type="text")
1837
1838 # Print the odds ratio
1839 exp(totaleffect$coefficients[-1])
1840 exp(confint(totaleffect))
1841
1842
1843
1844
1845
1846
1847
1848
1849 ## Affective + Cognitive trust
1850 wilcox.test(V1$combined, V2$combined, paired = FALSE)
1851 wilcox.test(V2$combined, V1$combined, paired = FALSE)
1852
1853
1854 wilcox.test(V1$combined, V3$combined, paired = FALSE)
1855 wilcox.test(V3$combined, V1$combined, paired = FALSE)
1856
1857 wilcox.test(V2$combined, V3$combined, paired = FALSE)
1858 wilcox.test(V3$combined, V2$combined, paired = FALSE)
1859
1860
1861
1862
1863
1864
1865
1866
1867 # Anthropomorphism
1868 wilcox.test(V1$Q7_1, V2$Q7_1, paired = FALSE)
1869 wilcox.test(V2$Q7_1, V1$Q7_1, paired = FALSE)
1870
1871
1872 wilcox.test(V1$Q7_1, V3$Q7_1, paired = FALSE)
1873 wilcox.test(V3$Q7_1, V1$Q7_1, paired = FALSE)
1874
1875
1876 wilcox.test(V2$Q7_1, V3$Q7_1, paired = FALSE)
1877 wilcox.test(V3$Q7_1, V2$Q7_1, paired = FALSE)
1878
1879
1880 mean(V1$Q7_1)
1881 mean(V2$Q7_1)
1882 mean(V3$Q7_1)
1883

```

```

1884
1885
1886 sd(V1$Q7_1)
1887 sd(V2$Q7_1)
1888 sd(V3$Q7_1)
1889
1890
1891 median(V1$Q7_1)
1892 median(V2$Q7_1)
1893 median(V3$Q7_1)
1894
1895
1896 # Cohen's D: V1V2 | 0.071628
1897 # Cohen's D: V1V3 | 0.023491
1898
1899
1900
1901
1902
1903 # Benevolence = Affective Trust
1904 wilcox.test(V1$Q8_1, V2$Q8_1, paired = FALSE)
1905 wilcox.test(V2$Q8_1, V1$Q8_1, paired = FALSE)
1906
1907
1908 wilcox.test(V1$Q8_1, V3$Q8_1, paired = FALSE)
1909 wilcox.test(V3$Q8_1, V1$Q8_1, paired = FALSE)
1910
1911
1912 wilcox.test(V2$Q8_1, V3$Q8_1, paired = FALSE)
1913 wilcox.test(V3$Q8_1, V2$Q8_1, paired = FALSE)
1914
1915
1916 mean(V1$Q8_1)
1917 mean(V2$Q8_1)
1918 mean(V3$Q8_1)
1919
1920
1921 sd(V1$Q8_1)
1922 sd(V2$Q8_1)
1923 sd(V3$Q8_1)
1924
1925
1926 median(V1$Q8_1)
1927 median(V2$Q8_1)
1928 median(V3$Q8_1)
1929
1930 # Cohen's D: V1V2 | 0.055746
1931 # Cohen's D: V1V3 | 0.217876
1932
1933
1934
1935 # Ability
1936 wilcox.test(V1$Q9_1, V2$Q9_1, paired = FALSE)
1937 wilcox.test(V2$Q9_1, V1$Q9_1, paired = FALSE)
1938
1939
1940 wilcox.test(V1$Q9_1, V3$Q9_1, paired = FALSE) # 0.02997
1941 wilcox.test(V3$Q9_1, V1$Q9_1, paired = FALSE)
1942
1943 wilcox.test(V2$Q9_1, V3$Q9_1, paired = FALSE)
1944 wilcox.test(V3$Q9_1, V2$Q9_1, paired = FALSE)
1945
1946
1947 mean(V1$Q9_1)
1948 mean(V2$Q9_1)
1949 mean(V3$Q9_1)
1950
1951
1952 sd(V1$Q9_1)
1953 sd(V2$Q9_1)
1954 sd(V3$Q9_1)
1955
1956 median(V1$Q9_1)

```

```

1957 median(V2$Q9_1)
1958 median(V3$Q9_1)
1959 # Cohen's D: V1V2 | 0.167847
1960 # Cohen's D: V1V3 | 0.200090
1961
1962
1963
1964 # Predictability
1965 wilcox.test(V1$Q11_1, V2$Q11_1, paired = FALSE)
1966 wilcox.test(V2$Q11_1, V1$Q11_1, paired = FALSE)
1967
1968
1969 wilcox.test(V1$Q11_1, V3$Q11_1, paired = FALSE)
1970 wilcox.test(V3$Q11_1, V1$Q11_1, paired = FALSE)
1971
1972
1973 wilcox.test(V2$Q11_1, V3$Q11_1, paired = FALSE)
1974 wilcox.test(V3$Q11_1, V2$Q11_1, paired = FALSE)
1975
1976
1977 mean(V1$Q11_1)
1978 mean(V2$Q11_1)
1979 mean(V3$Q11_1)
1980
1981
1982 sd(V1$Q11_1)
1983 sd(V2$Q11_1)
1984 sd(V3$Q11_1)
1985
1986 median(V1$Q11_1)
1987 median(V2$Q11_1)
1988 median(V3$Q11_1)
1989 # Cohen's D: V1V2 | 0.049781
1990 # Cohen's D: V1V3 | 0.228524
1991
1992
1993
1994 # Overall Trust
1995 wilcox.test(V1$Q11b_1, V2$Q11b_1, paired = FALSE)
1996 wilcox.test(V2$Q11b_1, V1$Q11b_1, paired = FALSE)
1997
1998
1999 wilcox.test(V1$Q11b_1, V3$Q11b_1, paired = FALSE)
2000 wilcox.test(V3$Q11b_1, V1$Q11b_1, paired = FALSE)
2001
2002
2003
2004 wilcox.test(V2$Q11b_1, V3$Q11b_1, paired = FALSE)
2005 wilcox.test(V3$Q11b_1, V2$Q11b_1, paired = FALSE)
2006
2007
2008 mean(V1$Q11b_1)
2009 mean(V2$Q11b_1)
2010 mean(V3$Q11b_1)
2011
2012
2013
2014
2015 sd(V1$Q11b_1)
2016 sd(V2$Q11b_1)
2017 sd(V3$Q11b_1)
2018
2019
2020 median(V1$Q11b_1)
2021 median(V2$Q11b_1)
2022 median(V3$Q11b_1)
2023
2024 # Cohen's D: V1V2 | 0.139411
2025 # Cohen's D: V1V3 | 0.263336
2026
2027
2028 # Intention to use
2029 wilcox.test(V1$Q12_1, V2$Q12_1, paired = FALSE)

```

```

2030 wilcox.test(V2$Q12_1, V1$Q12_1, paired = FALSE)
2031
2032 wilcox.test(V1$Q12_1, V3$Q12_1, paired = FALSE)
2033 wilcox.test(V3$Q12_1, V1$Q12_1, paired = FALSE)
2034
2035
2036 wilcox.test(V2$Q12_1, V3$Q12_1, paired = FALSE)
2037 wilcox.test(V3$Q12_1, V2$Q12_1, paired = FALSE)
2038
2039
2040 mean(V1$Q12_1)
2041 mean(V2$Q12_1)
2042 mean(V3$Q12_1)
2043
2044
2045
2046 sd(V1$Q12_1)
2047 sd(V2$Q12_1)
2048 sd(V3$Q12_1)
2049
2050 median(V1$Q12_1)
2051 median(V2$Q12_1)
2052 median(V3$Q12_1)
2053
2054 # Cohen's D: V1V2 | 0.015965
2055 # Cohen's D: V1V3 | 0.213661
2056
2057
2058 # Voice or not?
2059
2060 V1<-V1[!is.na(V1$Q12b_1), ]
2061 V2<-V2[!is.na(V2$Q12b_1), ]
2062 V3<-V3[!is.na(V3$Q12b_1), ]
2063
2064
2065
2066 wilcox.test(V1$Q12b_1, V2$Q12b_1, paired = FALSE)
2067 wilcox.test(V2$Q12b_1, V1$Q12b_1, paired = FALSE)
2068
2069
2070 wilcox.test(V1$Q12b_1, V3$Q12b_1, paired = FALSE)
2071 wilcox.test(V3$Q12b_1, V1$Q12b_1, paired = FALSE)
2072
2073 wilcox.test(V2$Q12b_1, V3$Q12b_1, paired = FALSE)
2074 wilcox.test(V3$Q12b_1, V2$Q12b_1, paired = FALSE)
2075
2076
2077 mean(V1$Q12b_1)
2078 mean(V2$Q12b_1)
2079 mean(V3$Q12b_1)
2080
2081
2082 sd(V1$Q12b_1)
2083 sd(V2$Q12b_1)
2084 sd(V3$Q12b_1)
2085
2086 median(V1$Q12b_1)
2087 median(V2$Q12b_1)
2088 median(V3$Q12b_1)
2089
2090 # Cohen's D: V1V3 | 0.958178
2091
2092
2093
2094
2095 totaleffect <- clm(as.factor(Q12b_1)~AllConcat, data = constructs, link="logit")
2096
2097 summary(totaleffect)
2098 confint(totaleffect)
2099
2100 stargazer(totaleffect, type="text")
2101
2102 # Print the odds ratio

```



```

2103 exp(totaleffect$coefficients[-1])
2104 exp(confint(totaleffect))
2105
2106
2107
2108
2109
2110 ##### Interaction Test #####
2111
2112 # install.packages("ordinal")
2113 library(ordinal)
2114
2115 # install.packages("rcompanion")
2116 library(rcompanion)
2117
2118 library(MASS) # Turn of when using dplyr package
2119 # install.packages("brant")
2120 library(brant)
2121
2122 # install.packages("AER")
2123 library(AER)
2124
2125 #install.packages("rockchalk")
2126 library(rockchalk)
2127
2128
2129
2130
2131 ## (personalized) Voice / Text --> Affective Trust --> Intention to Use
2132 # Depending on which dataset you are using, you could get different results. V1V2 or V1V3
2133
2134
2135 # Step 1
2136 totaleffect <- clm(as.factor(Q12_1)~AllConcat, data = constructs, link="logit")
2137
2138 summary(totaleffect)
2139 confint(totaleffect)
2140
2141 stargazer(totaleffect, type="text")
2142
2143 # Print the odds ratio
2144 exp(totaleffect$coefficients[-1])
2145 exp(confint(totaleffect))
2146
2147
2148 # Step 2
2149 fit.mediator <- clm(as.factor(affective)~AllConcat, data = constructs, link="logit")
2150
2151 summary(fit.mediator)
2152 confint(fit.mediator)
2153
2154 stargazer(fit.mediator, type="text")
2155
2156 # Print the odds ratio
2157 exp(fit.mediator$coefficients[-1])
2158 exp(confint(fit.mediator))
2159
2160
2161 # Step 3
2162 fit.dv <- clm(as.factor(Q12_1)~AllConcat+affective, data = constructs, link="logit")
2163
2164 summary(fit.dv)
2165 confint(fit.dv)
2166
2167 stargazer(fit.dv, type="text")
2168
2169 # Print the odds ratio
2170 exp(fit.dv$coefficients[-1])
2171 exp(confint(fit.dv))
2172
2173
2174
2175

```

```

2176 ## (personalized) Voice / Text --> Cognitive Trust --> Intention to Use
2177
2178
2179
2180 # Step 2
2181 fit.mediator <- clm(as.factor(cognitive)~AllConcat, data = constructs, link="logit")
2182
2183 summary(fit.mediator)
2184 confint(fit.mediator)
2185
2186 stargazer(fit.mediator, type="text")
2187
2188 # Print the odds ratio
2189 exp(fit.mediator$coefficients[-1])
2190 exp(confint(fit.mediator))
2191
2192
2193 # Step 3
2194 fit.dv <- clm(as.factor(Q12_1)~AllConcat+cognitive, data = constructs, link="logit")
2195
2196 summary(fit.dv)
2197 confint(fit.dv)
2198
2199 stargazer(fit.dv, type="text")
2200
2201 # Print the odds ratio
2202 exp(fit.dv$coefficients[-1])
2203 exp(confint(fit.dv))
2204
2205
2206 ## Overall trust
2207
2208
2209 # Step 2
2210 fit.mediator <- clm(as.factor(Q11b_1)~AllConcat, data = constructs, link="logit")
2211
2212 summary(fit.mediator)
2213 confint(fit.mediator)
2214
2215 stargazer(fit.mediator, type="text")
2216
2217 # Print the odds ratio
2218 exp(fit.mediator$coefficients[-1])
2219 exp(confint(fit.mediator))
2220
2221
2222 # Step 3
2223 fit.dv <- clm(as.factor(Q12_1)~AllConcat+Q11b_1, data = constructs, link="logit")
2224
2225 summary(fit.dv)
2226 confint(fit.dv)
2227
2228 stargazer(fit.dv, type="text")
2229
2230 # Print the odds ratio
2231 exp(fit.dv$coefficients[-1])
2232 exp(confint(fit.dv))
2233
2234
2235
2236 ## Anthropomorphism
2237
2238
2239
2240
2241
2242 # Step 1
2243 totaleffect <- clm(as.factor(affective)~AllConcat, data = constructs, link="logit")
2244
2245 summary(totaleffect)
2246 confint(totaleffect)
2247
2248 stargazer(totaleffect, type="text")

```

```

2249
2250 # Print the odds ratio
2251 exp(totaleffect$coefficients[-1])
2252 exp(confint(totaleffect))
2253
2254
2255 # Step 2
2256 fit.mediator <- clm(as.factor(Q11_1)~AllConcat, data = constructs, link="logit")
2257
2258 summary(fit.mediator)
2259 confint(fit.mediator)
2260
2261 stargazer(fit.mediator, type="text")
2262
2263 # Print the odds ratio
2264 exp(fit.mediator$coefficients[-1])
2265 exp(confint(fit.mediator))
2266
2267
2268 # Step 3
2269 fit.dv <- clm(as.factor(affective)~AllConcat+Q11_1, data = constructs, link="logit")
2270
2271 summary(fit.dv)
2272 confint(fit.dv)
2273
2274 stargazer(fit.dv, type="text")
2275
2276 # Print the odds ratio
2277 exp(fit.dv$coefficients[-1])
2278 exp(confint(fit.dv))
2279
2280
2281
2282 ##### Which voice attribute is more trust and intention to use #####
2283 ##as.factor(...) <- this could be: affective, cognitive or Q12_1
2284
2285 model1 <- clm(as.factor(affective)~Gender*Speed, data = V2, link="logit")
2286 summary(model1)
2287 confint(model1)
2288
2289 stargazer(model1, type="latex")
2290
2291 # Print the odds ratio
2292 exp(model1$coefficients[-1])
2293 exp(confint(model1))
2294
2295 # Print to check if the proportional regression assumption holds
2296 #model2 <- polr(affective~Gender+Speed + Gender*Speed, data = V2, Hess= TRUE)
2297 #brant(model2)
2298
2299
2300 V2 %>%
2301   dplyr::select(affective, Gender, Speed) %>%
2302   ggplot() +
2303   geom_boxplot(aes(x = Gender, y = affective, fill = Speed)) +
2304   theme_minimal()
2305
2306 V2 %>%
2307   dplyr::select(affective, Gender, Speed) %>%
2308   group_by(Gender, Speed) %>%
2309   dplyr::summarize(m = median(affective)) %>%
2310   ungroup() %>%
2311   ggplot() +
2312   geom_point(aes(x = Gender, y = m, color = Speed, shape = Speed)) +
2313   geom_line(aes(x = Gender, y = m, group = Speed, color = Speed)) +
2314   theme_minimal() +
2315   labs(y = "Levels of Affective Trust")
2316
2317
2318
2319
2320
2321 model1 <- clm(as.factor(cognitive)~Gender*Speed, data = V2, link="logit")

```

```

2322 summary(model1)
2323 confint(model1)
2324
2325 stargazer(model1, type="latex")
2326
2327 # Print the odds ratio
2328 exp(model1$coefficients[-1])
2329 exp(confint(model1))
2330
2331 # Print to check if the proportional regression assumption holds
2332 #model2 <- polr(cognitive~Gender+Speed + Gender*Speed, data = V2, Hess= TRUE)
2333 #brant(model2)
2334
2335
2336 V2 %>%
2337   dplyr::select(cognitive, Gender, Speed) %>%
2338   ggplot() +
2339     geom_boxplot(aes(x = Gender, y = cognitive, fill = Speed)) +
2340     theme_minimal()
2341
2342 V2 %>%
2343   dplyr::select(cognitive, Gender, Speed) %>%
2344   group_by(Gender, Speed) %>%
2345   dplyr::summarize(m = median(cognitive)) %>%
2346   ungroup() %>%
2347   ggplot() +
2348     geom_point(aes(x = Gender, y = m, color = Speed, shape = Speed)) +
2349     geom_line(aes(x = Gender, y = m, group = Speed, color = Speed)) +
2350     theme_minimal() +
2351     labs(y = "Levels of Cognitive Trust")
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370 ## Data preparation
2371
2372 #V1reduced=V1[100:200,]
2373
2374
2375
2376 # Mann-Whitney U test - GENDER
2377 V2Female<-V2[V2$Female=="Female",]
2378 V2Male<-V2[V2$Male=="Male",]
2379
2380
2381
2382 ## Affective Trust
2383
2384 # Female
2385 wilcox.test(V2Female$affective, V1$affective, paired = FALSE)
2386 wilcox.test(V1$affective, V2Female$affective, paired = FALSE)
2387 mean(V1$affective)
2388 mean(V2Female$affective)
2389
2390
2391 # Male
2392 wilcox.test(V2Male$affective, V1$affective, paired = FALSE)
2393 wilcox.test(V1$affective, V2Male$affective, paired = FALSE)
2394 mean(V1$affective)

```

```

2395 mean(V2Male$affective)
2396
2397
2398 ## Cognitive Trust
2399
2400 # Female
2401 wilcox.test(V2Female$cognitive, V1$cognitive, paired = FALSE)
2402 wilcox.test(V1$cognitive, V2Female$cognitive, paired = FALSE)
2403 mean(V1$cognitive)
2404 mean(V2Female$cognitive)
2405
2406
2407
2408 totaleffect <- clm(as.factor(cognitive)~Female, data = constructs, link="logit")
2409
2410 summary(totaleffect)
2411 confint(totaleffect)
2412
2413 stargazer(totaleffect, type="text")
2414
2415 # Print the odds ratio
2416 exp(totaleffect$coefficients[-1])
2417 exp(confint(totaleffect))
2418
2419
2420
2421
2422
2423
2424
2425 # Male
2426 wilcox.test(V2Male$cognitive, V1$cognitive, paired = FALSE)
2427 wilcox.test(V1$cognitive, V2Male$cognitive, paired = FALSE)
2428 mean(V1$cognitive)
2429 mean(V2Male$cognitive)
2430
2431
2432
2433
2434
2435
2436 # Mann-Whitney U test - SPEED
2437 V2Slow<-V2[V2$Slow=="Slow",]
2438 V2Fast<-V2[V2$Fast=="Fast",]
2439
2440
2441
2442 ## Affective Trust
2443
2444 # Slow
2445 wilcox.test(V2Slow$affective, V1$affective, paired = FALSE)
2446 wilcox.test(V1$affective, V2Slow$affective, paired = FALSE)
2447 mean(V1$affective)
2448 mean(V2Slow$affective)
2449
2450
2451 # Fast
2452 wilcox.test(V2Fast$affective, V1$affective, paired = FALSE)
2453 wilcox.test(V1$affective, V2Fast$affective, paired = FALSE)
2454 mean(V1$affective)
2455 mean(V2Fast$affective)
2456
2457
2458 ## Cognitive Trust
2459
2460 # Slow
2461 wilcox.test(V2Slow$cognitive, V1$cognitive, paired = FALSE)
2462 wilcox.test(V1$cognitive, V2Slow$cognitive, paired = FALSE)
2463 mean(V1$cognitive)
2464 mean(V2Slow$cognitive)
2465
2466
2467 # Fast

```

```

2468 wilcox.test(V2Fast$cognitive, V1$cognitive, paired = FALSE)
2469 wilcox.test(V1$cognitive, V2Fast$cognitive, paired = FALSE)
2470 mean(V1$cognitive)
2471 mean(V2Fast$cognitive)
2472
2473
2474
2475 #####
2476
2477 ### Country Classification
2478
2479
2480
2481
2482
2483 ## V1
2484
2485 V1Developed<-V1[V1$Classification=="Developed",]
2486
2487 V1Developing<-V1[V1$Classification=="Developing",]
2488
2489
2490 ## Affective Trust
2491
2492 # Developed versus developing
2493 wilcox.test(V1Developed$affective, V1Developing$affective, paired = FALSE)
2494 wilcox.test(V1Developing$affective, V1Developed$affective, paired = FALSE)
2495 mean(V1Developing$affective)
2496 mean(V1Developed$affective)
2497
2498
2499
2500
2501 ## Cognitive Trust
2502
2503 # Developed versus developing
2504 wilcox.test(V1Developed$cognitive, V1Developing$cognitive, paired = FALSE)
2505 wilcox.test(V1Developing$cognitive, V1Developed$cognitive, paired = FALSE)
2506 mean(V1Developing$cognitive)
2507 mean(V1Developed$cognitive)
2508
2509
2510 ## Intention to Use
2511 # Developed versus developing
2512 wilcox.test(V1Developed$intention, V1Developing$intention, paired = FALSE)
2513 wilcox.test(V1Developing$intention, V1Developed$intention, paired = FALSE)
2514 mean(V1Developing$intention)
2515 mean(V1Developed$intention)
2516
2517
2518
2519 # Mann-Whitney U test - Country Class
2520 V2Developed<-V2[V2$Classification=="Developed",]
2521
2522 V2Developing<-V2[V2$Classification=="Developing",]
2523
2524 #V2Developed_reduced=V2Developed[1:62,]
2525
2526 #V1reduced=V1[1:62,]
2527
2528
2529
2530 ## V2
2531
2532
2533 ## Affective Trust
2534
2535 # Developed
2536 wilcox.test(V2Developed$affective, V1$affective, paired = FALSE)
2537 wilcox.test(V1$affective, V2Developed$affective, paired = FALSE)
2538 mean(V1$affective)
2539 mean(V2Developed$affective)
2540

```

```

2541
2542 # Developing
2543 wilcox.test(V2Developing$affektive, V1$affektive, paired = FALSE)
2544 wilcox.test(V1$affektive, V2Developing$affektive, paired = FALSE)
2545 mean(V1$affektive)
2546 mean(V2Developing$affektive)
2547
2548
2549
2550
2551 # Developed versus developing
2552 wilcox.test(V2Developed$affektive, V2Developing$affektive, paired = FALSE)
2553 wilcox.test(V2Developing$affektive, V2Developed$affektive, paired = FALSE)
2554 mean(V2Developing$affektive)
2555 mean(V2Developed$affektive)
2556
2557
2558
2559
2560
2561 ## Cognitive Trust
2562
2563 # Developed
2564 wilcox.test(V2Developed$cognitive, V1$cognitive, paired = FALSE)
2565 wilcox.test(V1$cognitive, V2Developed$cognitive, paired = FALSE)
2566 mean(V1$cognitive)
2567 mean(V2Developed$cognitive)
2568
2569
2570
2571
2572
2573
2574 totaleffect <- clm(as.factor(cognitive)~Classification, data = V2, link="logit")
2575
2576 summary(totaleffect)
2577 confint(totaleffect)
2578
2579 stargazer(totaleffect, type="text")
2580
2581 # Print the odds ratio
2582 exp(totaleffect$coefficients[-1])
2583 exp(confint(totaleffect))
2584
2585
2586
2587
2588
2589
2590
2591
2592 # Developing
2593 wilcox.test(V2Developing$cognitive, V1$cognitive, paired = FALSE)
2594 wilcox.test(V1$cognitive, V2Developing$cognitive, paired = FALSE)
2595 mean(V1$cognitive)
2596 mean(V2Developing$cognitive)
2597
2598
2599
2600 # Developed versus developing
2601 wilcox.test(V2Developed$cognitive, V2Developing$cognitive, paired = FALSE)
2602 wilcox.test(V2Developing$cognitive, V2Developed$cognitive, paired = FALSE)
2603 mean(V2Developing$cognitive)
2604 mean(V2Developed$cognitive)
2605
2606
2607
2608
2609 ## Intention to Use
2610
2611 # Developed
2612 wilcox.test(V2Developed$intention, V1$intention, paired = FALSE)
2613 wilcox.test(V1$intention, V2Developed$intention, paired = FALSE)

```

```

2614 mean(V1$intention)
2615 mean(V2Developed$intention)
2616
2617
2618 # Developing
2619 wilcox.test(V2Developing$intention, V1$intention, paired = FALSE)
2620 wilcox.test(V1$intention, V2Developing$intention, paired = FALSE)
2621 mean(V1$intention)
2622 mean(V2Developing$intention)
2623
2624
2625
2626 # Developed versus developing
2627 wilcox.test(V2Developed$intention, V2Developing$intention, paired = FALSE)
2628 wilcox.test(V2Developing$intention, V2Developed$intention, paired = FALSE)
2629 mean(V2Developing$intention)
2630 mean(V2Developed$intention)
2631
2632
2633
2634
2635 totaleffect <- clm(as.factor(Q12b_1)~Classification, data = V2, link="logit")
2636
2637 summary(totaleffect)
2638 confint(totaleffect)
2639
2640 stargazer(totaleffect, type="text")
2641
2642 # Print the odds ratio
2643 exp(totaleffect$coefficients[-1])
2644 exp(confint(totaleffect))
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658 V3Developed<-V3[V3$Classification=="Developed",]
2659
2660 V3Developing<-V3[V3$Classification=="Developing",]
2661
2662
2663 ## V3
2664
2665
2666 ## Affective Trust
2667
2668 # Developed
2669 wilcox.test(V3Developed$affective, V1$affective, paired = FALSE)
2670 wilcox.test(V1$affective, V3Developed$affective, paired = FALSE)
2671 mean(V1$affective)
2672 mean(V3Developed$affective)
2673
2674
2675 # Developing
2676 wilcox.test(V3Developing$affective, V1$affective, paired = FALSE)
2677 wilcox.test(V1$affective, V3Developing$affective, paired = FALSE)
2678 mean(V1$affective)
2679 mean(V3Developing$affective)
2680
2681
2682 # Developed versus developing
2683 wilcox.test(V3Developed$affective, V3Developing$affective, paired = FALSE)
2684 wilcox.test(V3Developing$affective, V3Developed$affective, paired = FALSE)
2685 mean(V3Developing$affective)
2686 mean(V3Developed$affective)

```



```

2687
2688
2689
2690 ## Cognitive Trust
2691
2692 # Developed
2693 wilcox.test(V3Developed$cognitive, V1$cognitive, paired = FALSE)
2694 wilcox.test(V1$cognitive, V3Developed$cognitive, paired = FALSE)
2695 mean(V1$cognitive)
2696 mean(V3Developed$cognitive)
2697
2698
2699 # Developing
2700 wilcox.test(V3Developing$cognitive, V1$cognitive, paired = FALSE)
2701 wilcox.test(V1$cognitive, V3Developing$cognitive, paired = FALSE)
2702 mean(V1$cognitive)
2703 mean(V3Developing$cognitive)
2704
2705
2706
2707 # Developed versus developing
2708 wilcox.test(V3Developed$cognitive, V3Developing$cognitive, paired = FALSE)
2709 wilcox.test(V3Developing$cognitive, V3Developed$cognitive, paired = FALSE)
2710 mean(V3Developing$cognitive)
2711 mean(V3Developed$cognitive)
2712
2713
2714
2715
2716 ## Intention to Use
2717
2718 # Developed
2719 wilcox.test(V3Developed$intention, V1$intention, paired = FALSE)
2720 wilcox.test(V1$intention, V3Developed$intention, paired = FALSE)
2721 mean(V1$intention)
2722 mean(V3Developed$intention)
2723
2724
2725 # Developing
2726 wilcox.test(V3Developing$intention, V1$intention, paired = FALSE)
2727 wilcox.test(V1$intention, V3Developing$intention, paired = FALSE)
2728 mean(V1$intention)
2729 mean(V3Developing$intention)
2730
2731
2732
2733
2734 # Developed versus developing
2735 wilcox.test(V3Developed$intention, V3Developing$intention, paired = FALSE)
2736 wilcox.test(V3Developing$intention, V3Developed$intention, paired = FALSE)
2737 mean(V3Developing$intention)
2738 mean(V3Developed$intention)
2739
2740
2741
2742
2743
2744
2745
2746 #####
2747
2748 ## demographic difference in gender
2749
2750
2751
2752 ## V1
2753
2754 V1F<-V1[V1$Q14=="2",]
2755
2756 V1M<-V1[V1$Q14=="1",]
2757
2758
2759 ## Affective Trust

```

```

2760 # Female VS Male
2761 wilcox.test(V1F$affective, V1M$affective, paired = FALSE)
2762 wilcox.test(V1M$affective, V1F$affective, paired = FALSE)
2763 mean(V1M$affective)
2764 mean(V1F$affective)
2765
2766
2767 ## Cognitive Trust
2768 # Female VS Male
2769 wilcox.test(V1F$cognitive, V1M$cognitive, paired = FALSE)
2770 wilcox.test(V1M$cognitive, V1F$cognitive, paired = FALSE)
2771 mean(V1M$cognitive)
2772 mean(V1F$cognitive)
2773
2774
2775 ## Intention to Use
2776 # Female VS Male
2777 wilcox.test(V1F$intention, V1M$intention, paired = FALSE)
2778 wilcox.test(V1M$intention, V1F$intention, paired = FALSE)
2779 mean(V1M$intention)
2780 mean(V1F$intention)
2781
2782
2783
2784
2785
2786 ## V2
2787
2788
2789 # Mann-Whitney U test - Gender
2790 V2F<-V2[V2$Q14=="2",]
2791
2792 V2M<-V2[V2$Q14=="1",]
2793
2794
2795 ## Affective Trust
2796
2797
2798 # Female
2799 wilcox.test(V2F$affective, V1$affective, paired = FALSE)
2800 wilcox.test(V1$affective, V2F$affective, paired = FALSE)
2801 mean(V1$affective)
2802 mean(V2F$affective)
2803
2804
2805 # Male
2806 wilcox.test(V2M$affective, V1$affective, paired = FALSE)
2807 wilcox.test(V1$affective, V2M$affective, paired = FALSE)
2808 mean(V1$affective)
2809 mean(V2M$affective)
2810
2811
2812 # Female VS Male
2813 wilcox.test(V2F$affective, V2M$affective, paired = FALSE)
2814 wilcox.test(V2M$affective, V2F$affective, paired = FALSE)
2815 mean(V2M$affective)
2816 mean(V2F$affective)
2817
2818
2819
2820
2821 ## Cognitive Trust
2822
2823 # Female
2824 wilcox.test(V2F$cognitive, V1$cognitive, paired = FALSE)
2825 wilcox.test(V1$cognitive, V2F$cognitive, paired = FALSE)
2826 mean(V1$cognitive)
2827 mean(V2F$cognitive)
2828
2829
2830 # Male
2831 wilcox.test(V2M$cognitive, V1$cognitive, paired = FALSE)
2832 wilcox.test(V1$cognitive, V2M$cognitive, paired = FALSE)

```

```

2833 mean(V1$cognitive)
2834 mean(V2M$cognitive)
2835
2836
2837 # Female VS Male
2838 wilcox.test(V2F$cognitive, V2M$cognitive, paired = FALSE)
2839 wilcox.test(V2M$cognitive, V2F$cognitive, paired = FALSE)
2840 mean(V2M$cognitive)
2841 mean(V2F$cognitive)
2842
2843
2844
2845 ## Intention to Use
2846
2847 # Female
2848 wilcox.test(V2F$intention, V1$intention, paired = FALSE)
2849 wilcox.test(V1$intention, V2F$intention, paired = FALSE)
2850 mean(V1$intention)
2851 mean(V2F$intention)
2852
2853
2854 # Male
2855 wilcox.test(V2M$intention, V1$intention, paired = FALSE)
2856 wilcox.test(V1$intention, V2M$intention, paired = FALSE)
2857 mean(V1$intention)
2858 mean(V2M$intention)
2859
2860
2861 # Female VS Male
2862 wilcox.test(V2F$intention, V2M$intention, paired = FALSE)
2863 wilcox.test(V2M$intention, V2F$intention, paired = FALSE)
2864 mean(V2M$intention)
2865 mean(V2F$intention)
2866
2867
2868
2869 ## V3
2870
2871
2872 # Mann-Whitney U test - Gender
2873 V3F<-V3[V3$Q14=="2",]
2874
2875 V3M<-V3[V3$Q14=="1",]
2876
2877
2878 ## Affective Trust
2879
2880
2881 # Female
2882 wilcox.test(V3F$affective, V1$affective, paired = FALSE)
2883 wilcox.test(V1$affective, V3F$affective, paired = FALSE)
2884 mean(V1$affective)
2885 mean(V3F$affective)
2886
2887
2888 # Male
2889 wilcox.test(V3M$affective, V1$affective, paired = FALSE)
2890 wilcox.test(V1$affective, V3M$affective, paired = FALSE)
2891 mean(V1$affective)
2892 mean(V3M$affective)
2893
2894
2895 # Female VS Male
2896 wilcox.test(V3F$affective, V3M$affective, paired = FALSE)
2897 wilcox.test(V3M$affective, V3F$affective, paired = FALSE)
2898 mean(V3M$affective)
2899 mean(V3F$affective)
2900
2901
2902
2903
2904
2905

```

```

2906 ## Cognitive Trust
2907
2908 # Female
2909 wilcox.test(V3F$cognitive, V1$cognitive, paired = FALSE)
2910 wilcox.test(V1$cognitive, V3F$cognitive, paired = FALSE)
2911 mean(V1$cognitive)
2912 mean(V3F$cognitive)
2913
2914
2915 # Male
2916 wilcox.test(V3M$cognitive, V1$cognitive, paired = FALSE)
2917 wilcox.test(V1$cognitive, V3M$cognitive, paired = FALSE)
2918 mean(V1$cognitive)
2919 mean(V3M$cognitive)
2920
2921
2922 # Female VS Male
2923 wilcox.test(V3F$cognitive, V3M$cognitive, paired = FALSE)
2924 wilcox.test(V3M$cognitive, V3F$cognitive, paired = FALSE)
2925 mean(V3M$cognitive)
2926 mean(V3F$cognitive)
2927
2928
2929
2930
2931 ## Intention to Use
2932
2933 # Female
2934 wilcox.test(V3F$intention, V1$intention, paired = FALSE)
2935 wilcox.test(V1$intention, V3F$intention, paired = FALSE)
2936 mean(V1$intention)
2937 mean(V3F$intention)
2938
2939
2940 # Male
2941 wilcox.test(V3M$intention, V1$intention, paired = FALSE)
2942 wilcox.test(V1$intention, V3M$intention, paired = FALSE)
2943 mean(V1$intention)
2944 mean(V3M$intention)
2945
2946
2947
2948 # Female VS Male
2949 wilcox.test(V3F$intention, V3M$intention, paired = FALSE)
2950 wilcox.test(V3M$intention, V3F$intention, paired = FALSE)
2951 mean(V3M$intention)
2952 mean(V3F$intention)

```

Listing 1: R code for data processing used in this study.

