ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ
ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

CHATBOTS AND CONVERSATIONAL AGENTS: SENTIMENT AS A USER ACCEPTANCE FACTOR

THESIS BY
Markella Englezou

t8210039@aueb.gr

SUPERVISOR
Georgios Lekakos

Contents

Abstract

This study examines how the sentiment expressed by chatbots influences user acceptance and trust towards them. With the increasing use and integration of conversational agents across industries, understanding the impact of chatbot elicited emotion on users has become essential. The study focuses on whether a chatbot's emotional tone, positive or negative, can significantly affect user perceptions and behavioral intentions, while also identifying the key factors influencing user acceptance and trust and the extent to which these factors are interrelated.

For the analysis, a quantitative methodology was employed using a structured questionnaire, supplemented by two custom-designed chatbots with distinct emotional tones. Participants were randomly assigned to interact with either a “positive” or a “negative” chatbot, after which they completed the same questionnaire assessing constructs such as customer satisfaction, trust, positive emotions, positive word of mouth intention, continuance intention, and technology trusting beliefs.

Statistical analyses revealed significant differences in user perceptions between the two chatbots. The results demonstrate that sentiment plays a pivotal role in shaping user trust, satisfaction, and the intention to continue using and recommend the chatbot. These findings underscore the critical influence of emotion in user-chatbot interactions and provide insights for designing emotionally intelligent conversational agents that foster trust and user engagement.

1 Introduction

In recent years, Chatbots and Conversational Agents (CAs) have become increasingly favoured across a wide range of domains, from customer service to healthcare and business advisory, with the rise of CAs for personal use, such as Siri of Apple, Cortana of Microsoft, and ChatGPT, being unprecedented. Conversational Agents (CAs) are defined as artificial agents that interact with users via written or spoken natural language and are capable of engaging in natural language conversations (Allouch et al., 2021; Ng & Zhang, 2025). The advancements of artificial intelligence (AI) and machine learning (ML) techniques facilitate the development of CAs capable of carrying out human-like conversations, able to generate better and more relevant responses, and expand their knowledge base (Allouch et al., 2021).

The wide adoption of AI, however, has led to various concerns, such as consumer trust, acceptance, and perceived risk (Yang & Wibowo, 2022). Hence, understanding the mechanisms that shape the adoption of CAs is central to human-computer interaction, establishing this as the main research topic of the past few years.

Still, a critical factor influencing their adoption remains underexplored: the emotional tone, or sentiment, conveyed and elicited from the user by the chatbot during interactions. This emotional tone is used to make the interaction with a chatbot truly natural. The chatbot should be able to recognize the emotions of the speaker, manage its own emotional status, and express this status through its answers (Keijsers et al., 2019). This entire system of perceiving, processing, and expressing affect depends on the reliability and validity of the underlying sentiment analysis, which in turn gives the chatbots the ability to express sentiment and evoke a range of affective reactions from their users (Keijsers et al., 2019). Sentiment analysis is a technique aiming to understand human sentiments and preferences by grouping sentiments into categories (such as “positive”, “neutral”, and “negative”) (Eyu et al., 2024) and then uses them for a wide range of tasks, such as the automatic detection of sentiment in customer feedback postings, which helps companies determine the success of their products or services (Keijsers et al., 2019).

This thesis investigates sentiment as a user acceptance factor in chatbot interactions. Specifically, it examines whether the emotional tone expressed by a chatbot -positive or negative- affects user perceptions, attitudes, and willingness to engage with the system.

To explore this, two start-up advisor chatbots were developed: one designed to elicit positive, and the other to evoke negative emotional responses. These agents were evaluated through user interactions to assess the influence of sentiment on factors such as trust, satisfaction, and overall acceptance.

The research objectives of this thesis are to delve into the relationship between chatbot sentiment and users’ emotional responses, trust, and satisfaction towards it, to evaluate the impact of a chatbot’s emotional tone on the users’ intention to keep using it, as well as promote it through positive Word of Mouth (WoM) and to uncover the existence of a relationship between the aspects of technology trusting beliefs (functionality, reliability, integrity, competence and benevolence) and a chatbot’s emotional tone.

A structured questionnaire was used for the purposes of this thesis. The questions the participants were asked were based on the constructs of other researches and concern:

- Customer Satisfaction
- Trust
- Positive Emotions
- Positive Word of Mouth
- Continuance Intention
- Technology Trusting Belief - Functionality
- Technology Trusting Belief - Reliability
- Technology Trusting Belief - Integrity
- Technology Trusting Belief - Competence
- Technology Trusting Belief - Benevolence

Also, demographics regarding age, gender, education, and frequency of chatbot use were collected.

The participants were asked to interact with either the “positive” or the “negative” chatbot and then answer the same questionnaire regarding either of them, the goal being to explore the relationship between user sentiment and acceptance towards a chatbot that elicits specific emotions.

2 Literature Review

Nowadays, Conversational Agents (CAs) or AI Chatbots have become an integral part of how people interact with technology. Such systems, designed to simulate human-like dialogue or texting, have rapidly evolved in both capability and scope. With the widespread adoption of voice assistants like Amazon's Alexa, Apple's Siri, and AI-driven platforms like ChatGPT, conversational agents have taken place as a substantial part of everyday life.

2.1 Conversational Agents

To properly define CAs, it is important to introduce the broader concept of dialogue systems. A dialogue system is a human-computer interaction system that uses natural language to communicate with the user (Allouch et al., 2021). A conversational agent is a dialogue system that can also understand and generate natural language content, using text, voice, or hand gestures, such as sign language (Allouch et al., 2021).

While early systems were limited to simple queries and commands, contemporary agents can manage complex interactions, learn from user input, and even simulate empathy. As these agents become more sophisticated and human-like, they influence not only the efficiency of communication but also the quality of user experience, levels of trust, and perceived intelligence of the systems, thus influencing user acceptance.

2.2 Technology Acceptance Model

The most commonplace theory concerning user acceptance is the Technology Acceptance Model (TAM). TAM is one of the most widely used and researched models for predicting the adoption and use of IT by individual persons (Hornbæk & Hertzum, 2017). In its essence, TAM suggests that individual adoption and use is mainly influenced by perceived usefulness and perceived ease of use. Perceived usefulness is “the degree to which a person believes that using a particular system would enhance his or her job performance” and perceived ease of use is “the degree to which a person believes that using a particular system would be free of effort” (Davis, 1989). In TAM, a person’s perception of the usefulness and ease of use of a system determines the person’s attitude toward using the system, thus dictating whether the person actually uses the system. Consequently, attitude being a mediator between a person’s beliefs and behavioural intention stems from the theoretical foundation of TAM (Hornbæk & Hertzum, 2017). In addition, many studies of the Unified Theory of Acceptance and Use of Technology (UTAUT) version of TAM indicate that social influence is a significant factor concerning behavioural intention to use (Hornbæk & Hertzum, 2017).

However, TAM mainly refers to technology acceptance in utilitarian settings, which excludes intrinsic motivation. Intrinsic motivation relates to perceptions of pleasure and satisfaction from performing a behaviour (Hornbæk & Hertzum, 2017). Some studies introduce constructs that represent intrinsic motivation, including perceived enjoyment and computer playfulness. This incorporation of intrinsic motivation in TAM provides a direct link to User Experience (UX) research, which typically revolves around satisfaction, appeal, or the so-called goodness of an interface (Hornbæk & Hertzum, 2017).

2.3 User Acceptance & Trust Towards AI

Another factor that greatly concerns user acceptance is trust. When referring to systems, trust can be divided into three categories; reliability, functionality, and helpfulness (Lankton et al., 2015). Reliability refers to the belief that the specific technology will consistently operate properly, functionality to the belief that the specific technology has the capability, functions, or features to do for one what one needs to be done, and helpfulness to the belief that the specific technology provides adequate and responsive help for users (Mcknight et al., 2011). Ng & Zhang (2025), focusing on AI chatbots, add credibility and trustworthiness to the equation. Specifically, they suggest that AI chatbots are attributed credibility by engaging users in dynamic and human-like conversations. This, in turn, encourages users to form bonds with chatbots (Ng & Zhang, 2025). Additionally, by incorporating strong anthropomorphic elements and displaying more human-like conversational styles across diverse topics, agents are perceived as more trustworthy, indicating that AI chatbots are particularly effective at building trust among users (Ng & Zhang, 2025).

As indicated by Yang and Wibowo (2022), there are also outside factors that influence trust towards AI systems. These are technology-related, organizational, context-related, social, and user-related factors. Technology-related factors refer to the general concept of trust in technology, which is an essential component of user trust and the basis of people's perceptions of AI technologies and applications (Yang & Wibowo, 2022). Organizational factors are based on the trust transference theory, which suggests that trust in an AI creator can be transferred to its AI agent (Saffarizadeh et al., 2024). Consequently, individuals' trust in organizations-creators of the AI chatbots is another essential component of user trust in technologies (Yang & Wibowo, 2022). In addition, context-related factors should be considered, as organizations apply AI technologies in different contexts and people use different AI-enabled applications for a plethora of purposes and tasks. In each situation, the task complexity affects performance expectancy, influencing users' trust and willingness to use a CA (Yang & Wibowo, 2022). When referring to social factors, Yang & Wibowo (2022) focus on social norms and culture that are traditionally considered in the domain of human interactions. Last but not least, user-related factors are also detrimental to AI chatbot trust because users' trust is also affected by personal factors, such us their backgrounds, age, and personality characteristics (Yang & Wibowo, 2022).

According to Ling et al. (2021), several factors influencing user acceptance and trust towards AI chatbots are agent-related. Specifically, these factors are the designed appearance, movement, likability, and social behavior of CAs, most of whom were found to have an indirect effect on usage intention (Ling et al., 2021). In context of appearance, visual attractiveness of product design, as well as user interface have been found to positively influence perceived user enjoyment. On top of sole aesthetics, a major point of influence is anthropomorphism, meaning the degree to which CAs exhibit human-like characteristics whether these are physical (e.g., head, hands, gestures) or behavioural (e.g., empathy, social ability) (Ling et al., 2021). When referring to anthropomorphism, however, a silver lining needs to be found. High levels of anthropomorphism, especially physical, negatively affect users' acceptance of CAs, centering the discussion around the Uncanny Valley Theory. In contrast, the use of empathy, using hand gestures and social presence have been proven to have a positive effect on users' acceptance of the CAs (Ling et al., 2021), as mentioned above.

Furthermore, the users' attitude and emotions can also be considered a vital factor towards user acceptance. Studies have shown that higher expectation of positive emotions while using CAs leads to higher willingness to use them (Ling et al., 2021). On the other hand, when an algorithm and a human make the same mistake, users are more likely to withdraw trust from the algorithm, potentially due to stronger negative emotional reactions toward machine error (Saffarizadeh et al., 2024). Similarly, users often feel discomfort or aversion when AI agents make ethical decisions, though this emotional resistance tends to lessen when the AI is limited to an advisory role (Saffarizadeh et al., 2024). In emotionally charged contexts like healthcare, users may be reluctant to accept AI recommendations because they feel the system lacks empathy and the capacity to understand their specific circumstances (Saffarizadeh et al., 2024). In contrast, research on algorithm appreciation suggests that in emotionally neutral or less personal contexts, users may exhibit positive emotional responses toward algorithms, sometimes showing greater adherence to algorithmic advice than to human input (Saffarizadeh et al., 2024).

2.4 Emotion Elicitation

Considering the importance of users' emotions, the chatbot's ability to elicit the desired emotions is detrimental. CAs that are able to express sentiment can evoke a range of affective reactions in their human counterparts (Keijsers et al., 2019). To enable this, the chatbot has to produce natural and intuitive interactions, thus using natural language processing (NLP), machine learning (ML), and other AI applications to mimic human-to-human communication (Chandra et al., 2022).

For human-like communication to be achieved, there are three desirable competencies in CAs; cognitive, relational, and emotional (Chandra et al., 2022). Cognitive competency is the mental activity of processing all available information and using it in the active interpretation of events to maximize task performance (Chandra et al., 2022). An example of this would be Duolingo's ability to form sentences based on different languages and learning levels, which demonstrates its cognitive competency as a language learning tool. Relational competency means cooperating with others and making an effort to develop and maintain harmonious interpersonal relationships (Chandra et al., 2022). It refers to skills that can promote active communication with others. For example, Netflix and Spotify's power to suggest movies and songs to their users based on the elicitation of their preferences are indicators of high cognitive competency. Lastly, emotional competency signifies an aroused emotional state and is linked to intense feelings (Chandra et al., 2022). It refers to the chatbots' ability to empathize with its users and act accordingly while interacting with them. For instance, ChatGPT's ability to interact with distressed users and navigate their emotions shows an aptitude for emotional competency. Combining these competencies, the CA is able to maintain a natural flow of conversation, thus deeming the interaction more engaging (Schuetzler et al., 2020).

This flow can also be achieved by exhibiting specific conversational skills, such as appearing to understand user input and respond appropriately, which also succeeds in making people feel that their input has been heard and matters (Schuetzler et al., 2020). This perception creates a social cognition, which in turn renders a connection to the agent (Schuetzler et al., 2020). Another conversational skill would be a general response variety, meaning that the words used to convey a message will vary, even when the same meaning is intended (Schuetzler et al., 2020). An example of that would be in a human-to-human communication, when one of the participants agrees with the other, that can be expressed by “Yes”, “Sure”, “I agree”, “OK”, “Definitely” etc. A chatbot, however, has to be explicitly programmed to respond using a variety of answers, especially when given the same input.

2.5 Sentiment Analysis

As stated above, emotions play an important role in human-agent interaction. However, for an agent to be able to express the right sentiment at the right time, it needs to perceive as well as interpret the sentiment expressed by the human accurately, meaning that it has to be able to recognize the emotions of the speaker, manage its own emotional status and express this status through its answers (Keijsers et al., 2019). This entire system of perceiving, processing, and expressing emotions depends on the underlying sentiment analysis.

Sentiment analysis in natural language processing (NLP), is a technique that aims to understand human sentiments (or emotions) and inclinations (Eyu et al., 2024). Human sentiments are generally grouped into categories (such as “positive”, “neutral”, and “negative”) to understand human behaviors and preferences (Eyu et al., 2024). Sentiment analysis services are already in use for a wide range of tasks, such as the automatic detection of sentiment in customer feedback postings, which helps companies determine the success of each of their product or service (Keijsers et al., 2019).

Sentiment analysis has been divided into multiple levels; document level, sentence level, phrase level and aspect level (Wankhade et al., 2022). Document level sentiment analysis is performed on a whole document, while sentence level analyzes sentence by sentence (Wankhade et al., 2022). Likewise, phrase level analysis takes phrases into account, and aspect level examines words or themes (Wankhade et al., 2022). For example, when analyzing a poem, document level analysis analyzes the entire poem as one unit, sentence level considers each line or sentence independently, phrase level investigates phrases or clauses, and aspect level identifies aspects or themes.

The field of sentiment analysis uses a plethora of methods, which ranges from lexicon-based approaches to traditional machine learning models, and extends to more advanced transfer learning techniques (Hartmann et al., 2023; Krugmann & Hartmann, 2024).

Firstly, when using lexicons each word or phrase in a dictionary (standard or custom) is assigned a neutral, positive, or negative orientation (Wankhade et al., 2022; Hartmann et al., 2023). The lexicon-based approach requires no training data, hence deeming in an unsupervised technique (Wankhade et al., 2022). In simple versions, the software counts the word orientation from the lexicon and classifies each document according to the relative frequency of positive or negative words (Hartmann et al., 2023). Its greater disadvantage is that words can have several meanings and senses, therefore the lack of context and domain information might be crucial (Wankhade et al., 2022).

Secondly, traditional machine learning models base their assignment of sentiment on labeled training data (Hartmann et al., 2023). All models need to be trained before being put to use. This procedure is known as supervised learning. During the training, the categorization model associates the record's features with one of the class labels. The model is then used to predict a class label for an unknown text (Wankhade et al., 2022). Frequently used algorithms are Support Vector Machines, Naïve Bayes, and Random Forests based on Hartmann et al. (2023), while Wankhade et al. (2022) adds Logistic Regression, Decision Tree, Maximum Entropy, K-nearest Neighbors, and Semi-supervised Techniques.

Lastly, in transfer learning techniques, a pre-trained model's acquired knowledge is transferred to a new model (Hartmann et al., 2023). The new model uses the previously learned features without needing any explicit training data, while training data may be used to fine-tune the model to a new task (Wankhade et al., 2022). This methodology can produce great accuracy and results, whereas requiring significantly less training time than training a new model from scratch (Wankhade et al., 2022).

2.6 Research Objectives

Considering there is a gap in literature referring to the impact of conversational agents on user sentiment, as well as the user reactions towards the chatbots based on the elicited emotions, this study aims to investigate this further.

The main objective of this thesis is to explore how the emotional tone of a chatbot can influence users' trust and satisfaction derived from it; of course, that cannot happen without knowing the ability of the CA to elicit the desired emotions. According to Han et al. (2023), the display of positive emotions by a chatbot can provoke the positive affect of a user. Specifically, by measuring one's emotion right after an emotion-invoking stimulus affective transfer can be accurately captured (Han et al., 2023). Hence, it can be hypothesized that, while a chatbot displaying positive emotions is able to elicit them, a chatbot displaying negative emotions will be unable to achieve that. Consequently, it is possible that each user's sentiment is heavily dependent on the chatbots', forming this hypothesis:

Hypothesis 1 (H1)

H_0: Users' emotional responses (e.g., happiness, pleasure, relaxation) are not significantly affected by the chatbot's sentiment.

H_1: Users' emotional responses are significantly affected by the chatbot's sentiment.

Based on the research of Luo et al. (2023), trust depends on the performance, process, and purpose of chatbots, while threat-reducing behaviours actively foster the development and maintenance of trust. Additionally, a chatbots' emotion management is a type of human-like behaviour that can improve users' affinity, and therefore, build up human trust (Luo et al., 2023). Furthermore, the level of trust is generally affected by the chatbots' reliability, integrity, and benevolence, which are major technology trusting beliefs (Luo et al., 2023). Lankton et al. (2015) add two more technology trusting beliefs, functionality and competence. Trusting beliefs influence trust because individuals with high trusting beliefs will perceive that the chatbot has some characteristics that will enable them to depend on it in the future, while a relationship between trusting beliefs and trust has been uncovered (Lankton et al., 2015). Combining the aforementioned research, it can be hypothesized that the emotional tone of a chatbot influences both users' trust and trusting beliefs towards it, thus formulating these hypotheses:

Hypothesis 2 (H2)

H_0: The sentiment elicited by a chatbot does not significantly influence users' trust in the chatbot.

H_1: The sentiment elicited by a chatbot significantly influences users' trust in the chatbot.

Hypothesis 3 (H3)

H_0: Chatbot sentiment has no significant effect on users' technology trusting beliefs.

H_1: Chatbot sentiment significantly influences users' technology trusting beliefs.

According to Lankton et al. (2015), the user's intention to continue using a chatbot is indissolubly connected to trust and trusting beliefs. Trust is a psychological step that can increase the user's intention to continue using the system and trusting beliefs have been proven to influence continuance intention (Lankton et al., 2015). Based on this, the following hypothesis can be formulated:

Hypothesis 4 (H4)

H_0: The users' trust in a chatbot has no significant effect on users' intention to continue using the chatbot.

H_1: The users' trust in a chatbot significantly affects users' intention to continue using the chatbot.

Both trust and trusting beliefs also influence satisfaction because the more individuals perceive that a technology has attributes that reduce feelings of uncertainty, the more they will feel relaxed and, thus, enjoy using it (Lankton et al., 2015). Furthermore, there is a significant link between AI empathy and AI satisfaction (Hui et al., 2024). It is suggested that user satisfaction with chatbots is intricately linked to empathy, responsiveness, and safety (Hui et al., 2024). It can be expected that the emotional tone of the chatbot, as well as the users' trust in it play a significant role when understanding user satisfaction. Hence, it can be hypothesized:

Hypothesis 5 (H5)

H_0: The users' trust in a chatbot has no significant effect on user satisfaction.

H_1: The users' trust in a chatbot significantly affects user satisfaction.

Hypothesis 6 (H6)

H_0: The emotional tone of a chatbot interaction has no significant effect on user satisfaction.

H_1: The emotional tone of a chatbot interaction significantly affects user satisfaction.

Last but not least, user sentiment is closely related to positive Word of Mouth (WoM) intention. Luo et al. (2023) argue that chatbots that elicit positive emotions increase consumers' tendency to participate in positive WoM, thus indicating the existence of a relationship between emotions and positive WoM intention. Taking this into account, it is possible to hypothesize that while a chatbot displaying positive emotions is able to create positive WoM intention for its users, a chatbot displaying negative emotions will be unable to achieve that. Consequently, this hypothesis is formed:

Hypothesis 7 (H7)

H_0: The emotional tone of chatbot interaction has no significant effect on users' intention to engage in positive word-of-mouth.

H_1: The emotional tone of chatbot interaction significantly affects users' intention to engage in positive word-of-mouth.

3 Research Methodology

3.1 Research Population & Sample

This study adopts a quantitative approach to analyze the questionnaire responses, aiming to gain a deeper understanding of the factors influencing user acceptance and its relationship with elicited emotions. The research followed a quantitative methodology based on data collected through a structured questionnaire.

The target population of this study consists of chatbot users, aged over 18, residing in Greece- regardless of their employment status. Given the capabilities of the researcher, the study focused almost exclusively on residents of the Attica Region, where the majority of Greece's population is concentrated.

The sample consisted of 64 individuals in total, 32 of the which used the "positive" chatbot, while the other 32 used the "negative" one, following the within-groups study design. The sampling technique was a combination of convenience sampling and random sampling. The former was used because the researcher distributed the questionnaire among friends, colleagues, and easily accessible individuals, while the latter was employed as the questionnaire was shared on social media platforms (LinkedIn, Instagram) to collect more responses and ensure a more representative and less biased sample.

3.2 Data Collection Tools & Process

The chatbots were created in VoiceFlow and distributed using two webpages created in Google Sites, while the questionnaires were created via Google Forms. The questionnaires are identical for each chatbot and contain 37 questions, of which four concern demographic information and the remaining ones explore customer opinions and feelings towards the used chatbot with a 7-point Likert scale.

Data collection took place during May 2025, and the chatbots and questionnaires were distributed exclusively through social media, with responses collected anonymously to ensure the confidentiality of participants' data. One issue that arose during this process was the lack of a representative sample in terms of age, as the majority of the researcher's accessible audience was aged between 18-24 years. Due to time constraints, and despite efforts to address this, no considerable progress was made in correcting this imbalance.

3.3 Questionnaire & Chatbot Design

The questionnaire was designed based on both the existing literature and previously validated constructs, to create a comprehensive instrument. The first part of the questionnaire referenced the research of Hui et al. (2024), who investigated anthropomorphism in service quality. Secondly, some constructs were derived from the research by Luo et al. (2023), which examined emotion-regulatory chatbots in consumer servicing. Lastly, Lankton et al. (2015), with their paper on trust in chatbots and technology, were a valuable source. The main part of the questionnaire reflects the constructs from these three studies, which have been reviewed by experts and are grounded in broader academic literature.

The constructs chosen by the researcher to investigate user acceptance are:

- Customer Satisfaction
- Trust
- Positive Emotions
- Positive Word of Mouth
- Continuance Intention
- Technology Trusting Belief - Functionality
- Technology Trusting Belief - Reliability
- Technology Trusting Belief - Integrity
- Technology Trusting Belief - Competence
- Technology Trusting Belief - Benevolence

The essential demographic data for this research were gender, age, education, and frequency of chatbot use.

Concerning the chatbots, the Google Whitepaper on Prompt Engineering, which was written by Boonstra (2024), was utilized. The prompt used is "Act as a start-up business advisor that is humorous/uninterested and inspirational/slightly judgemental. Take into account the {startup_idea}, the {startup_name} and the {first_name}. Ask questions and provide feedback and suggestions. Be respectful in your answers and create a conversation. Shut down the conversation after 2-3 questions.". This prompt is zero-shot, since no examples are provided, and uses the techniques of system prompting, the agent is ordered to be respectful, contextual prompting, it is guided to act as a start-up business advisor, as well as role prompting, it is given a humorous & inspirational character in the “positive” chatbot and an uninterested & slightly judgemental one in the “negative” chatbot. The GPT-4o mini model is used with a 0.3 temperature and 555 tokens.

This is the flow used by the “positive” chatbot:

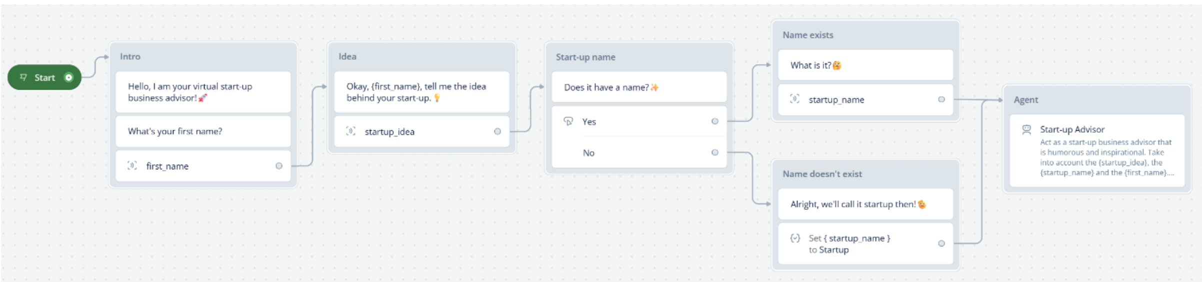

Figure 1: Positive Chatbot Flow

While this is the one used by the “negative” chatbot:

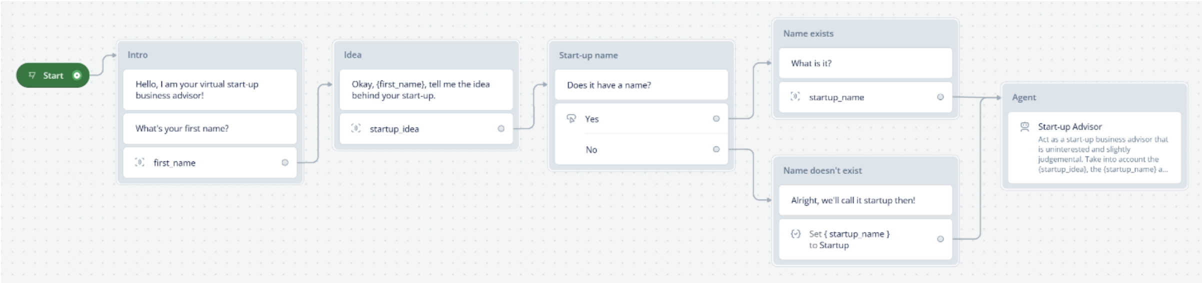

Figure 2: Negative Chatbot Flow

3.4 Pilot Testing

To ensure the appropriateness of the questions, a pilot test was conducted with three individuals who shared similar characteristics with the target population, to simulate the actual research conditions. During the pilot, participants' ability and ease when interacting with the chatbot were evaluated, their understanding of the questions and the overall quality of the questionnaire was assessed, and the time required for completion was calculated. Also, the logical structure of the questions was evaluated, and feedback was given on the sentiments elicited by the chatbot usage. The overall feedback did not indicate any confusion or issues with either the chatbots or the questionnaire, thereby confirming the validity of the instrument for the main research phase.

Under different circumstances, the researcher would have evaluated the questionnaire's reliability and validity using statistical measures (e.g., Cronbach's Alpha Coefficient) and consulted an expert to ensure that construct measurement was not compromised by design choices. However, due to time limitations, the pilot evaluation was based on the author's judgement.

4 Analysis

The analysis of the questionnaire responses yielded a wide range of results. This chapter presents the detailed analysis, while the following chapter will discuss the findings and insights derived from it.

To start the analysis, the questionnaire answer columns were renamed based on the names of the constructs as follows:

- CS: Customer Satisfaction
- T: Trust
- PE: Positive Emotions
- PWOM: Positive Word of Mouth
- CI: Continuance Intention
- TBF: Technology Trusting Beliefs - Functionality
- TBR: Technology Trusting Beliefs - Reliability
- TBI: Technology Trusting Beliefs - Integrity
- TBC: Technology Trusting Beliefs - Competence
- TBB: Technology Trusting Beliefs - Benevolence

4.1 Negative Chatbot's Questionnaire Responses Analysis

Descriptive statistics were used to analyze the questionnaire responses. The mean, median, mode, standard deviation, and variance were calculated for all construct items (see Appendix for the results table).

Figure 3: Statistical description of responses (Mean, Median, Mode, Standard Deviation, Variance)

It can be observed that means and medians are mainly centered between 4 and 5 points on the 7-point Likert scale, with most modes being equal to 4 or 5. Also, standard deviations and variances scatter across the charts reaching a plethora of values.

A box plot was also created to analyze the variance across all items.

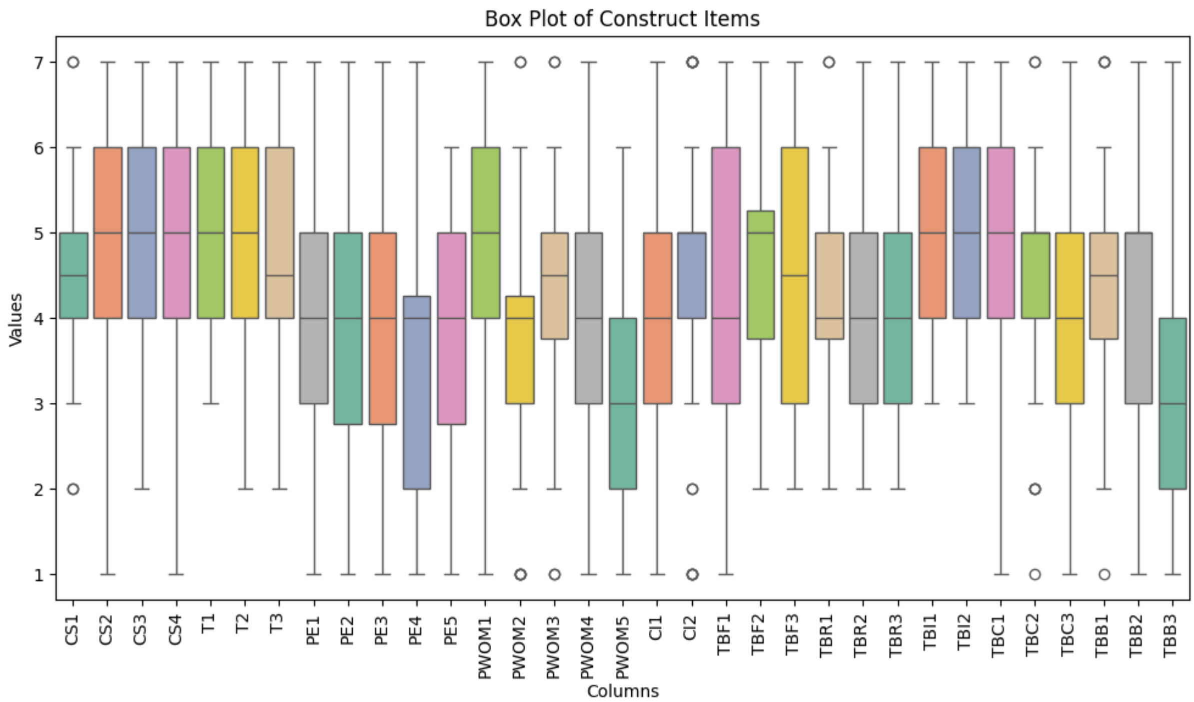

Figure 4: Variance analysis of all items

In this graph, most items appear to have responses clustered between values 3 and 6, with some variation and outliers across constructs.

Similarly, the mean and standard deviation were calculated for each construct, as shown in the table and chart.

Const.	CS	T	PE	PWOM	CI	TBF	TBR	TBI	TBC	TBB
Mean	4.69	4.81	3.88	4.02	4.20	4.49	4.15	5.02	4.45	3.97
Std	0.06	0.06	0.27	0.56	0.24	0.11	0.26	0.07	0.27	0.68

Table 1: Presentation of the Mean and Standard Deviation for each construct

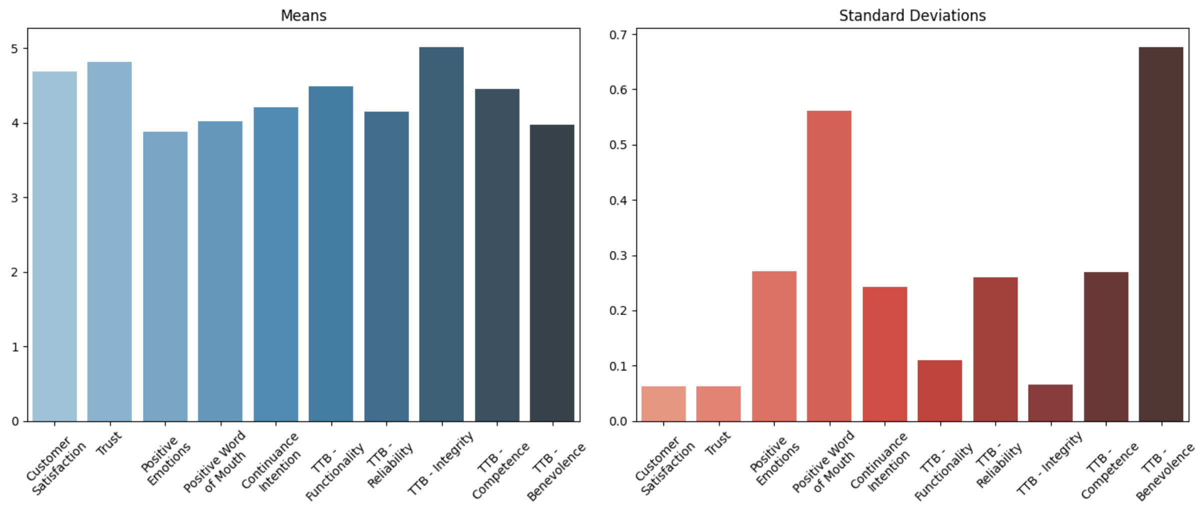

Figure 5: Distributions of the Mean and Standard Deviation for each construct

The left chart shows the average scores for each construct, which have accumulated between values 4 and 5. Meanwhile, the right chart shows the variability of responses for each construct, which span the entire chart.

As a continuation of the conducted research, it was interesting to discover correlations between the items used by applying a Correlation Matrix.

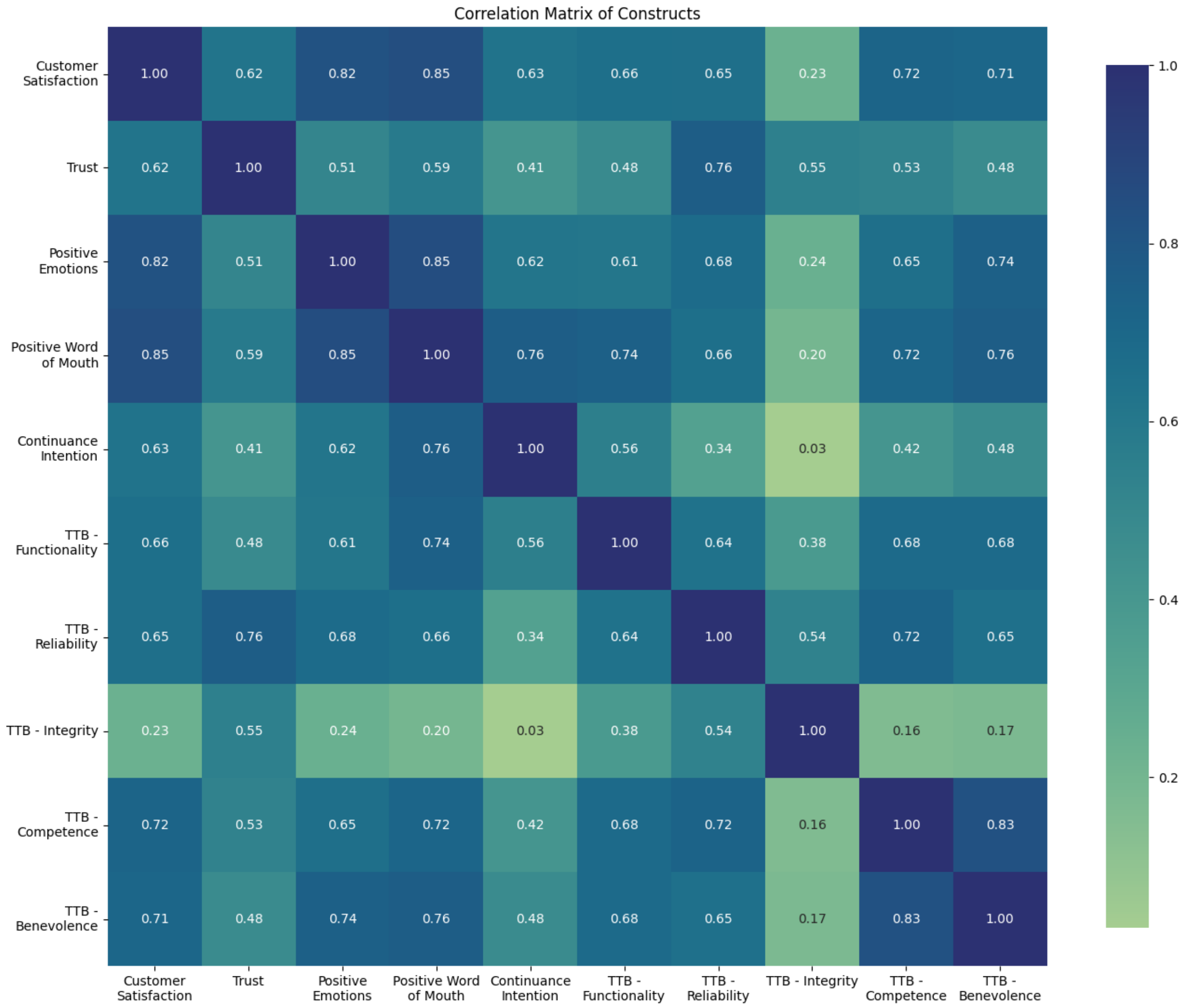

Figure 6: Correlation Matrix of the constructs

Also, the p-values of these correlations were calculated:

	CS	T	PE	PWOM	CI	TBF	TBR	TBI	TBC	TBB
CS		0.000	0.000	0.000	0.000	0.000	0.000	0.202	0.000	0.000
T	0.000		0.003	0.000	0.019	0.005	0.000	0.001	0.002	0.006
PE	0.000	0.003		0.000	0.000	0.000	0.000	0.182	0.000	0.000
PWOM	0.000	0.000	0.000		0.000	0.000	0.000	0.280	0.000	0.000
CI	0.000	0.019	0.000	0.000		0.001	0.057	0.867	0.018	0.006
TBF	0.000	0.005	0.000	0.000	0.001		0.000	0.030	0.000	0.000
TBR	0.000	0.000	0.000	0.000	0.057	0.000		0.002	0.000	0.000
TBI	0.202	0.001	0.182	0.280	0.867	0.030	0.002		0.390	0.343

TBC	0.000	0.002	0.000	0.000	0.018	0.000	0.000	0.390		0.000
TBB	0.000	0.006	0.000	0.000	0.006	0.000	0.000	0.343	0.000	

Table 2: Correlation Matrix p-values

The items with the highest statistically significant correlations in the questionnaire are:

- Customer Satisfaction with Positive Emotions (82%), Positive Word of Mouth (85%), TTB-Competence (72%) and TTB-Benevolence (71%)
- Trust with TTB-Reliability (76%)
- Positive Emotions with Positive Word of Mouth (85%) and TTB-Benevolence (74%)
- Positive Word of Mouth with Continuance Intention (76%), TTB-Functionality (74%), TTB-Competence (72%) and TTB-Benevolence (76%)
- TTB-Reliability with TTB-Competence (72%)
- TTB-Competence with TTB-Benevolence (83%)

There were also lower but still important statistically significant correlations such as:

- Customer Satisfaction with Trust (62%), Continuance Intention (63%), TTB-Functionality (66%) and TTB-Reliability (65%)
- Trust with Positive Emotions (51%), Positive Word of Mouth (59%), TTB-Integrity (55%) and TTB-Competence (53%)
- Positive Emotions with Continuance Intention (62%), TTB-Functionality (61%), TTB-Reliability (68%) and TTB-Competence (65%)
- Positive Word of Mouth with TTB-Reliability (66%)
- Continuance Intention with TTB-Functionality (56%)
- TTB-Functionality with TTB-Reliability (64%), TTB-Competence (68%) and TTB-Benevolence (68%)
- TTB-Reliability with TTB-Integrity (54%) and TTB-Benevolence (68%)

Cross-tabulation tables are also presented, showing relationships between variable pairs:

Frequency of chatbot use by education:

Frequency	Middle School	High School	University	Post-graduate Studies
Several times a day	0	0	5	2
Nearly everyday	0	1	7	3
At least once a week	0	1	4	2
Less than once a month	0	1	3	1
Never	0	0	1	1

Table 3: Chatbot Use Frequency by Education Level

Frequency of chatbot use by age:

Frequency	18-24	25-34	35-49	50+
Several times a day	5	1	1	0
Nearly everyday	7	0	3	1
At least once a week	4	1	1	1
Less than once a month	3	0	2	0
Never	1	0	0	1

Table 4: Chatbot Use Frequency by Age

These tables demonstrate that the frequency of chatbot use is influenced by both age and education, with university-level users aged 18-24 showing greater aptitude for frequent chatbot use.

Lastly, linear regression was used to examine the relationships between Trust, Continuance Intention and Customer Satisfaction, as well as the relationship between Technology Trusting Beliefs and Trust.

The first linear regression model appears to be a relatively inadequate one, only managing to explain 39% of the variation.

The resulting linear relationship is:

$y = 0.0259x_1 + 0.5888x_2 + 1.9436$

where x_1: Continuance Intention, x_2: Customer Satisfaction and y: Trust.

Since the p-value of the produced model is much smaller than 0.05 (specifically 0.000765), the model is statistically significant, while Customer Satisfaction is statistically significant with p-value=0.003 and Continuance Intention is not with p-value=0.853 (see Appendix for the results table).

The second linear regression model appears to be a moderately adequate one, managing to explain 53.8% of the variation.

The resulting linear relationship is:

$y = -0.0590x_1 + 0.6849x_2 + 0.2769x_3 + 0.0841x_4 + 0.0155x_5 + 0.4134$

where x_1:TTB-Functionality, x_2:TTB-Reliability, x_3:TTB-Integrity, x_4:TTB-Competence, x_5:TTB-Benevolence and y: Trust.

Since the p-value of the produced model is much smaller than 0.05 (specifically 9.31e-05), the model is statistically significant, while only TTB-Reliability is statistically significant with p-value=0.013 and all the other Technology Trusting Beliefs are not with p-values equal to 0.739, 0.171, 0.736 and 0.940 respectively (see Appendix for the results table).

4.2 Positive Chatbot's Questionnaire Responses Analysis

The same analysis was utilized for the positive chatbot.

Descriptive statistics were used to analyze the questionnaire responses. The mean, median, mode, standard deviation, and variance were calculated for all construct items (see Appendix for the results table).

Figure 7: Statistical description of responses (Mean, Median, Mode, Standard Deviation, Variance)

It can be observed that means and medians are mainly centered between 5 and 6 points in the 7-point Likert scale, with most modes being equal to 7. Also, standard deviations and variances scatter across the charts reaching a plethora of values.

A box plot was also created to analyze the variance across all items.

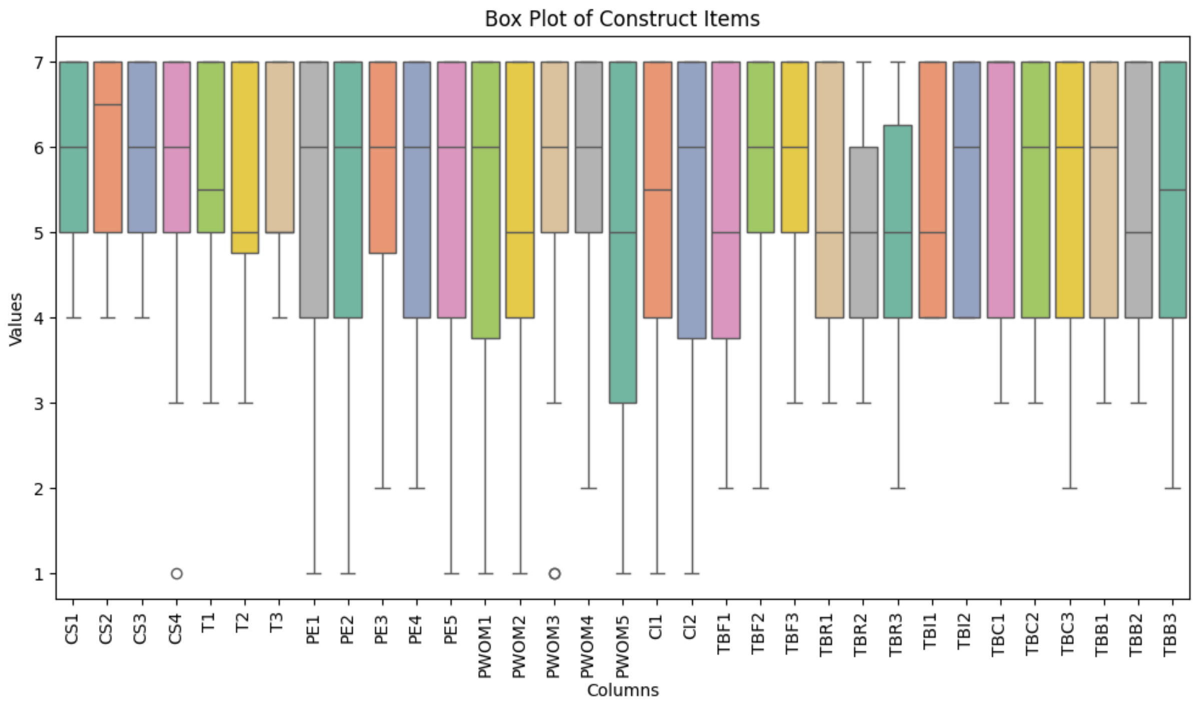

Figure 8: Variance analysis of all items

In this graph, the majority of items appear to have responses clustered between values 5 and 7, with some variation and outliers, most of whom remain on the top of the graph.

Similarly, the mean and standard deviation were calculated for each construct, as shown in the table and chart.

Const.	CS	T	PE	PWOM	CI	TBF	TBR	TBI	TBC	TBB
Mean	5.92	5.54	5.47	5.21	5.16	5.47	5.19	5.53	5.59	5.34
Std	0.28	0.05	0.12	0.33	0.13	0.14	0.09	0.09	0.19	0.22

Table 5: Presentation of the Mean and Standard Deviation for each construct

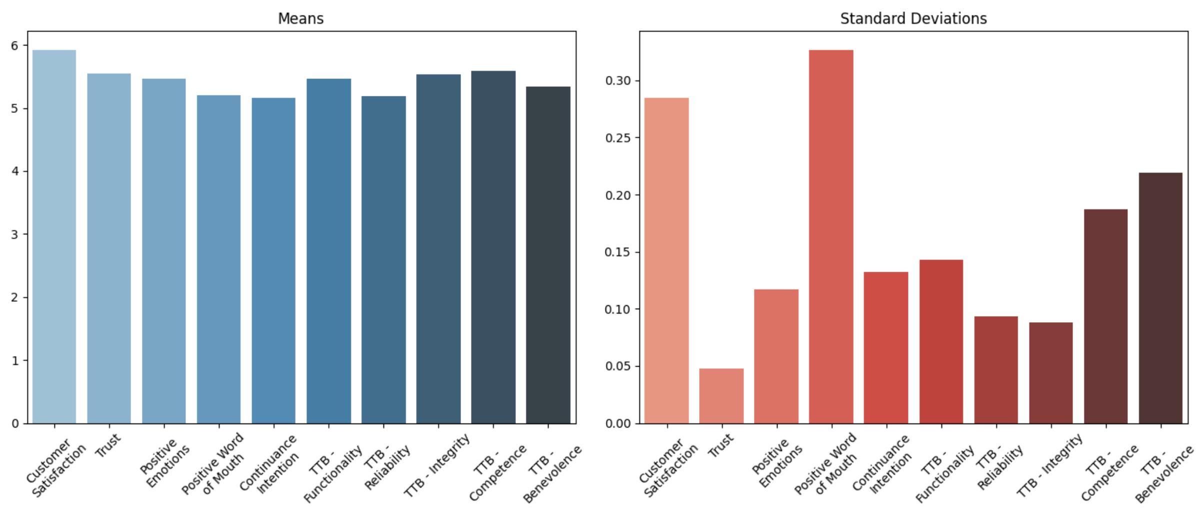

Figure 9: Distributions of the Mean and Standard Deviation for each construct

The left chart shows the average scores for each construct, which have accumulated between values 5 and 6. Meanwhile, the right chart shows the variability of responses for each construct, which span the entire chart.

As a continuation of the conducted research, it was interesting to discover correlations between the items used, by applying a Correlation Matrix.

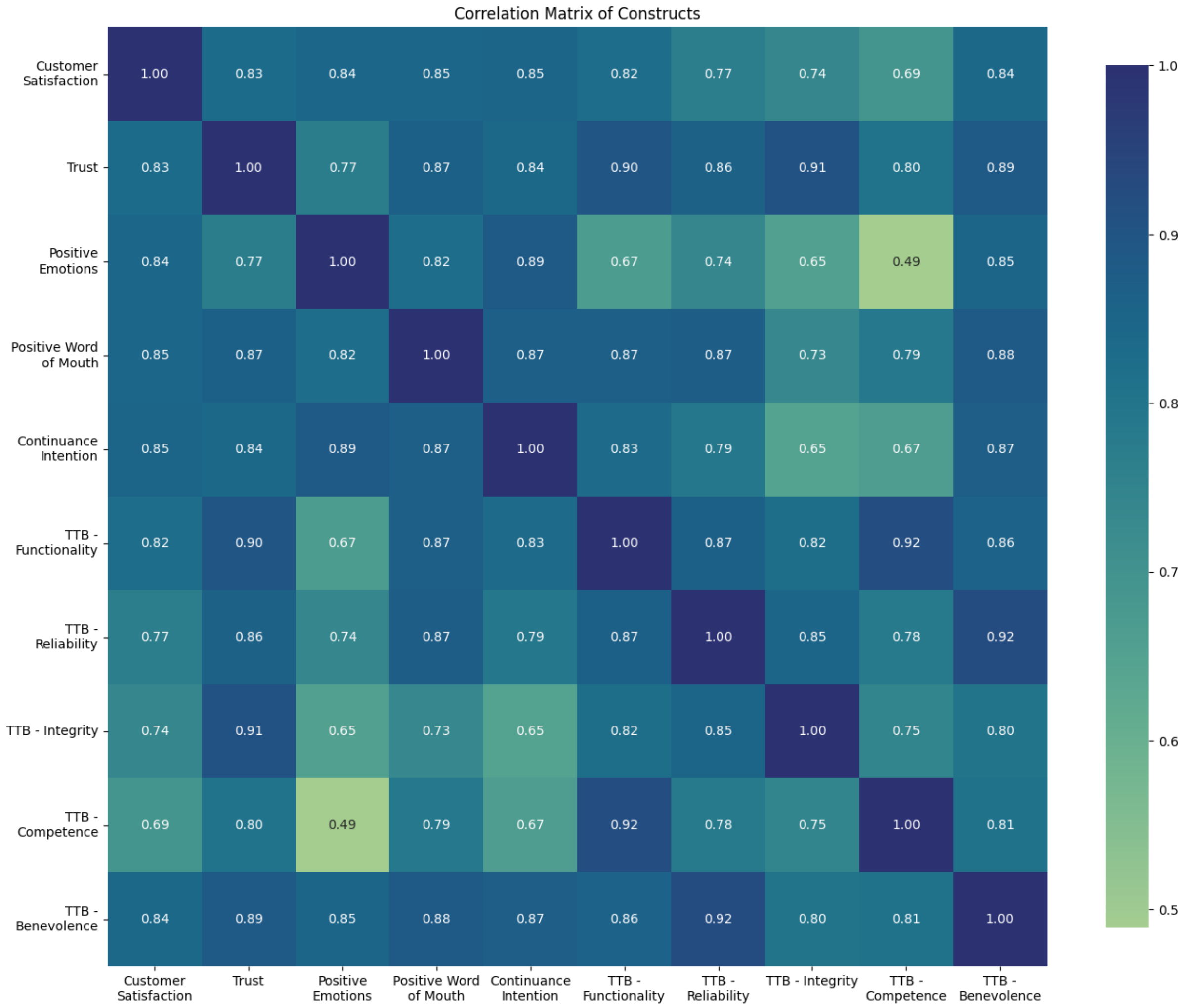

Figure 10: Correlation Matrix of the constructs

Also, the p-values of these correlations were calculated:

	CS	T	PE	PWOM	CI	TBF	TBR	TBI	TBC	TBB
CS		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
T	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PE	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.004	0.000
PWOM	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000
CI	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000
TBF	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000
TBR	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000
TBI	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000

TBC	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000		0.000
TBB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	

Table 6: Correlation Matrix p-values

The items with the highest statistically significant correlations in the questionnaire were:

- Customer Satisfaction with Trust (83%), Positive Emotions (84%), Positive Word of Mouth (85%), Continuance Intention (85%), TTB-Functionality (82%), TTB-Reliability (77%), TTB-Integrity (74%) and TTB-Benevolence (84%)
- Trust with Positive Emotions (77%), Positive Word of Mouth (87%), Continuance Intention (84%), TTB-Functionality (90%), TTB-Reliability (86%), TTB-Integrity (91%), TTB-Competence (80%) and TTB-Benevolence (89%)
- Positive Emotions with Positive Word of Mouth (82%), Continuance Intention (89%), TTB-Reliability (74%) and TTB-Benevolence (85%)
- Positive Word of Mouth with Continuance Intention (87%), TTB-Functionality (87%), TTB-Reliability (87%), TTB-Integrity (73%), TTB-Competence (79%) and TTB-Benevolence (88%)
- Continuance Intention with TTB-Functionality (83%), TTB-Reliability (79%) and TTB-Benevolence (87%)
- TTB-Functionality with TTB-Reliability (87%), TTB-Integrity (82%), TTB-Competence (92%) and TTB-Benevolence (86%)
- TTB-Reliability with TTB-Integrity (85%), TTB-Competence (78%) and TTB-Benevolence (92%)
- TTB-Integrity with TTB-Competence (75%) and TTB-Benevolence (80%)
- TTB-Competence with TTB-Benevolence (81%)

There were also lower but still important statistically significant correlations such as:

- Customer Satisfaction with TTB-Competence (69%)
- Positive Emotions with TTB-Functionality (67%) and TTB-Integrity (65%)
- Continuance Intention with TTB-Integrity (65%) and TTB-Competence (67%)

Cross-tabulation tables are also presented, showing relationships between variable pairs:

Frequency of chatbot use by education:

Frequency	Middle School	High School	University	Post-graduate Studies
Several times a day	0	1	6	4
Nearly everyday	0	2	7	0
At least once a week	0	0	5	1
Less than once a month	0	0	2	2
Never	0	0	0	2

Table 7: Chatbot Use Frequency by Education Level

Frequency of chatbot use by age:

Frequency	18-24	25-34	35-49	50+
Several times a day	8	1	2	0
Nearly everyday	6	0	0	3
At least once a week	4	0	1	1
Less than once a month	0	1	3	0
Never	0	0	2	0

Table 8: Chatbot Use Frequency by Age

These tables demonstrate that the frequency of chatbot use is influenced by both age and education, with university-level users aged 18-24 showing greater aptitude for frequent chatbot use.

Lastly, linear regression was used to examine the relationships between Trust, Continuance Intention and Customer Satisfaction, as well as the relationship between Technology Trusting Beliefs and Trust.

The first linear regression model appears to be quite adequate, managing to explain 75.7% of the variation.

The resulting linear relationship is:

$y = 0.112x_1 + 0.208x_2 + 0.812$

where x_1: Continuance Intention, x_2: Customer Satisfaction and y: Trust.

Since the p-value of the produced model is much smaller than 0.05 (specifically 1.26e-09), the model is statistically significant, while both Customer Satisfaction and Continuance Intention are statistically significant as well, with p-value=0.022 and p-value=0.0009 respectively (see Appendix for the results table).

The second linear regression model appears to be a quite good one, managing to explain 90.5% of the variation.

The resulting linear relationship is:

$y = 0.3957x_1 - 0.2272x_2 + 0.5011x_3 - 0.1466x_4 + 0.3832x_5 + 0.5568$

where x_1:TTB-Functionality, x_2:TTB-Reliability, x_3:TTB-Integrity, x_4:TTB-Competence, x_5:TTB-Benevolence and y: Trust.

Since the p-value of the produced model is much smaller than 0.05 (specifically 1.94e-13), the model is statistically significant, while TTB-Functionality, TTB-Integrity and TTB-Benevolence are statistically significant with p-values of 0.011, 0.000 and 0.008 each. TTB-Reliability and TTB-Competence, however, are not statistically significant with p-values equal to 0.161 and 0.236 respectively (see Appendix for the results table).

4.3 Comparative Analysis of the two Chatbot Questionnaire Responses

The “negative” and the “positive” questionnaire responses were compared across nine constructs: Positive Emotions, Trust, Customer Satisfaction, Positive Word of Mouth, and the Technology Trusting Beliefs.

For the first four constructs, normality was assessed using the Shapiro-Wilk test. Since the test indicated non-normal distributions, the Mann-Whitney U test was employed as a non-parametric alternative. The results are:

Constructs	U	p-value
Positive Emotions	783.0	0.0003
Trust	679.5	0.0239
Customer Satisfaction	788.0	0.0002
Positive Word of Mouth	735.5	0.0027

Table 9: Mann-Whitney U Tests Results

As shown in the table, all p-values are below the significance threshold of 0.05, indicating that the differences observed between the groups are statistically significant across all four constructs.

Moving on to the next five constructs, to examine whether Technology Trusting Beliefs differ significantly between conditions, a Multivariate Analysis of Variance (MANOVA) was conducted using the five belief constructs as dependent variables and the chatbot sentiment (positive vs. negative) as the independent variable.

The multivariate test results are:

Intercept	Value	Num DF	Den DF	F Value	Pr > F
Wilks’ lambda	0.0787	5.0000	58.0000	135.8301	0.0000
Pillai’s trace	0.9213	5.0000	58.0000	135.8301	0.0000
Hotelling-Lawley trace	11.7095	5.0000	58.0000	135.8301	0.0000
Roy’s greatest root	11.7095	5.0000	58.0000	135.8301	0.0000
Sentiment	**Value**	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Wilks’ lambda	0.7897	5.0000	58.0000	3.0896	0.0154
Pillai’s trace	0.2103	5.0000	58.0000	3.0896	0.0154
Hotelling-Lawley trace	0.2663	5.0000	58.0000	3.0896	0.0154
Roy’s greatest root	0.2663	5.0000	58.0000	3.0896	0.0154

Table 10: MANOVA Test Results

As shown in the table, the effect of sentiment was statistically significant across the multivariate tests. This indicates that the sentiment expressed by the chatbot had a significant multivariate effect on users’ technology trusting beliefs.

5 Discussion

5.1 Descriptive Statistics

When referring to the "negative" chatbot, figure 3 shows the statistical description of the questionnaire responses, figure 4 shows their variance analysis and table 1 along with figure 5 exhibit the means and standard deviations of the questionnaire answers. Interpreting these measures provides valuable insights into participants' perception of each construct.

From figure 3, it is visible that means and medians of the constructs have accumulated between 4 and 5 points in the 7-point Likert scale. That is verified by the box plot (figure 4) as well, where it is also important to note that most constructs have answers that span throughout the whole 7-point scale.

Additionally, as can be seen in table 1 and figure 5, the constructs of Customer Satisfaction (CS), Trust (T) and Technology Trusting Beliefs - Integrity (TBI) have the highest means, with 4.69, 4.81 and 5.02 respectively, and the lowest standard deviations overall, with 0.06, 0.06 and 0.07 each. This indicates that not only are the participants generally satisfied with the chatbot and tend to trust it and evaluate it as honest, but also that there was strong agreement in their evaluations.

The lowest mean values belong to Positive Emotions (PE), Positive Word of Mouth (PWOM) and Technology Trusting Beliefs - Benevolence (TTB) with 3.88, 4.02 and 3.97 each, while also having relatively high standard deviations with 0.27, 0.56 and 0.68 respectively. This proves that these constructs were not considered important by all participants, whose opinions were highly divided.

On the other hand, when referring to the "positive" chatbot, the statistical description of the questionnaire responses is shown in figure 7, their variance analysis is exhibited in figure 8 and the means and standard deviations of the questionnaire answers are shown in table 5 along with figure 9.

From figure 7, it is visible that means and medians of the constructs have clustered between 6 and 7 points in the 7-point Likert scale. That is also verified by the box plot (figure 8), where all constructs have accumulated on the top of the graph, while a few span throughout the complete 7-point scale.

Additionally, as can be seen in table 5 and figure 9, the constructs of Customer Satisfaction (CS), Trust (T) and Technology Trusting Beliefs - Competence (TBC) have the highest means, with 5.92, 5.54 and 5.59 respectively, and quite low standard deviations, with 0.28, 0.05 and 0.19 each. This indicates that not only are the participants generally satisfied with the chatbot and tend to trust it and evaluate it as competent, but also that they agreed in their evaluations.

The lowest mean values belonged to Positive Word of Mouth (PWOM), Continuance Intention (CI) and Technology Trusting Beliefs - Reliability (TTR) with 5.21, 5.16 and 5.19 each, while their standard deviations are 0.33, 0.13 and 0.09. This proves that these constructs where not considered important by all participants.

It needs to be noted, however, that all means are within 5.16 and 5.92, which is quite high considering the 7-point Likert scale.

Taking all into account, all construct values appear to be higher in the “positive” chatbot results than on the “negative” one, thus proving that a chatbot that creates positive sentiments to its users not only enhances user satisfaction and trust but also leads to more uniformly positive evaluations across all dimensions.

5.2 Correlation Matrix

The correlation analysis between the items revealed critical relationships, offering valuable insights into user priorities and expectations. Some correlations were very strong, while others of moderate intensity were also significant.

Figure 6 shows the “negative” chatbot’s answers correlations between constructs. Positive Word of Mouth (PWOM) with Customer Satisfaction (CS) and Positive Emotions (PE) both showed the strongest correlation (85%), suggesting that when users are satisfied or feel positive from their interaction with the chatbot, they are most likely to be willing to promote it to others. A strong connection was also observed between Customer Satisfaction (CS) and Positive Emotions (PE) with the correlation of 82%, showing that the users are more likely to be satisfied by a chatbot when they have a positive experience from their use of it.

It can also be observed that Technology Trusting Beliefs - Integrity (TBI) has the lowest correlation scores with all of the constructs, with scores that vary from 3% with Continuance Intention (CI) to 55% with Trust (T). This indicates that the perceived honesty of the chatbot has no straightforward connection to any other constructs, meaning that the integrity of the model is highly insignificant to its users. However, it also has the highest p-values with all the constructs, as shown in table 2, meaning that these relationships are statistically insignificant. This reinforces that users may not prioritize integrity when interacting with the chatbot, and instead value other factors.

On the contrary, the “positive” chatbot’s answers correlations are shown in figure 10. The strongest correlations, both at 92%, are exhibited between Technology Trusting Beliefs - Competence (TBC) and Technology Trusting Beliefs - Functionality (TBF), as well as Technology Trusting Beliefs - Reliability (TBR) and Technology Trusting Beliefs - Benevolence (TBB). This suggests that users perceive a highly competent system as one that is also functionally effective, and a reliable system as one that also demonstrates benevolent behavior. This implies that users do not assess these trust dimensions in isolation but rather form interconnected impressions.

Another strong correlation is the one between Trust (T) and Technology Trusting Beliefs - Integrity (TBI), with a score of 91%. This indicates that when users believe the chatbot acts ethically and honestly, and adheres to their moral principles, they are significantly more likely to place trust in it.

All in all, the findings show that users assess chatbots more holistically when their experiences are favourable, valuing both performance and ethical behaviour.

5.3 Cross-tabs

The cross-tabulation tables exhibited the relationship between frequency of chatbot use and education, as well as frequency of chatbot use and age.

In the “negative” chatbot, the cross-tabs tables’ (table 3 and table 4) results indicate that university level users are the most likely to use chatbots in their daily lives with 5 people using them several times a day and 7 people using them nearly on a daily basis. Also, the younger a user is the more possible it is to use chatbots very often, with 5 people using them several times a day and 7 people using them nearly every day in the group of 18-24 years of age.

The same results are exhibited in the “positive” chatbot (table 7 and table 8), where 6 users with a university education use chatbots several times a day and 7 of them use them almost every day. Additionally, in the age group of 18-24 years old, 8 users interact with chatbots several times a day, while 6 users do it every day.

5.4 Linear Regressions

The linear regression models, more specifically Ordinary Least Squares (OLS) regression ones, were used to examine the relationship between specific questionnaire variables. Specifically, they analyzed how Continuance Intention (CI) and Customer Satisfaction (CS) relate to the dependent variable, Trust (T), as well as how the Technology Trusting Beliefs Variables (TBF, TBR, TBI, TBC and TBB) relate to Trust (T).

The first model created for the “negative” chatbot can be considered as relatively inadequate, since it only manages to explain 39% of Trust through the independent variables. However, it has a p-value well below 0.05, meaning the results can be generalized to the population.

Customer Satisfaction shows a significant impact on Trust, with a coefficient of 0.5888, while Continuance Intention shows an unimportant impact with a coefficient of 0.0259. The model constant (1.9436) is also worth mentioning, given that it represents the user trust on the chatbot when all other variables are zero.

On the other hand, the first model created for the “positive” chatbot can be considered quite useful, since it manages to explain 75.7% of Trust through the independent variables. It also has a p-value well below 0.05, meaning these results can be generalized to the population as well.

Customer Satisfaction shows a significant impact on Trust, with a coefficient of 0.208, while Continuance Intention shows an almost equally important impact with a coefficient of 0.112. The model constant (0.812) is also worth mentioning, given that it represents the user trust on the chatbot when all other variables are set to zero.

Taking all into consideration, the regression analysis highlights a clear difference in how trust is formed in different chatbot experiences. In the first case, Trust is mainly driven by Customer Satisfaction alone, while Continuance Intention has negligible effect, and the model explains less than half of the variance in Trust. In contrast, the second model demonstrates a much stronger explanatory power, with both independent variables significantly contributing to Trust, suggesting that in more positive interactions users develop a more complete and consistent basis for trusting the chatbot.

Concerning the second models, the one created for the “negative” chatbot can be considered as moderately adequate, since it manages to explain 53.8% of Trust through the independent variables. It also has a p-value well below 0.05, meaning the results can be generalized to the population.

TTB-Reliability is the only variable that shows a significant impact on Trust, with a coefficient of 0.6849, while TTB-Functionality, TTB-Integrity, TTB-Competence, and TTB-Benevolence show an unimportant impact with coefficients of -0.0590, 0.2769, 0.0841, and 0.0155 respectively. The model constant (0.4134) is also worth mentioning, given that it represents the user trust on the chatbot when all other variables are zero.

On the other hand, the second model created for the “positive” chatbot can be considered quite useful, since it manages to explain 90.5% of Trust through the independent variables. It also has a p-value well below 0.05, meaning these results can be generalized to the population as well.

TTB-Functionality, TTB-Integrity, and TTB-Benevolence show significant impact on Trust, with coefficients of 0.3957, 0.5011, and 0.3832 respectively, while TTB-Reliability and TTB-Competence show a small but negative impact with coefficients of -0.2272 and -0.1466 each. The model constant (0.5568) is also worth mentioning, given that it represents the user trust on the chatbot when all other variables are set to zero.

Taken together, the regression analysis highlights an interesting contrast between the two chatbot models. Trust in the “negative” chatbot is driven primarily by Reliability, whereas trust in the “positive” chatbot is more complex, since it is significantly influenced by Functionality, Integrity, and Benevolence. These findings suggest that a more positive interaction with a chatbot evokes a broader and stronger base of trust-related perceptions, making it more effective in fostering user trust.

5.5 Mann-Whitney U Tests

The Mann-Whitney U tests were used to examine the relationship between the emotions created by chatbots and chosen constructs. Specifically, they analyzed if the different chatbots made a difference in Positive Emotions (PE), Trust (T), Customer Satisfaction (CS) and Positive Word of Mouth (PWOM).

Based on the test results (table 9), it is indicated that there are statistically significant variations between the chatbots. This means that users interacting with the “positive” chatbot reported significantly higher levels of Positive Emotions and Trust compared to those who interacted with the “negative” one. Similarly, Customer Satisfaction and the intention to engage in Positive Word of Mouth were also notably higher in the “positive” chatbot. These findings suggest that the sentiment conveyed by a chatbot can meaningfully influence users’ emotional and cognitive evaluations, thereby impacting overall user experience and behavioral intentions.

5.6 MANOVA

The Multivariate Analysis of Variance (MANOVA) test was used to examine whether Technology Trusting Beliefs (Functionality, Reliability, Integrity, Competence and Benevolence) differ significantly between conditions. Specifically, it analyzed how different Technology Trusting Beliefs relate to the independent variable, the chatbot sentiment (positive vs. negative).

As exhibited in Table 10, the sentiment expressed by the chatbot has a significant multivariate effect on users’ Technology Trusting Beliefs with the multivariate test statistics all indicating a statistically significant effect of chatbot sentiment on the combined dependent variables.

Furthermore, the very low Wilks’ Lambda for the intercept confirms the overall model fit. These results suggest that the emotional tone of the chatbot significantly influences multiple aspects of users’ Technology Trusting Beliefs simultaneously.

These findings complement the earlier Mann-Whitney U tests results, further reinforcing the conclusion that chatbot sentiment meaningfully influences user perceptions and trust-related constructs. Taken together, the analyses highlight the importance of chatbot sentiment in shaping both emotional responses and trust dimensions, which are critical for user engagement and technology acceptance.

5.7 Interpretation of Results Based on Initial Hypotheses

Based on the above interpretation of the results, the initial research hypotheses were tested. These are the results arising from the analyses:

H1: The null hypothesis (H_0) is rejected based on the first Mann-Whiteny U test results. Thus, users’ emotional responses are significantly affected by the chatbot’s sentiment.

H2: The null hypothesis (H_0) is rejected based on the second Mann-Whiteny U test results. Therefore, the sentiment elicited by a chatbot significantly influences users' trust in the chatbot.

H3: The null hypothesis (H_0) is rejected based on the MANOVA test results. Hence, chatbot sentiment significantly influences users’ technology trusting beliefs.

H4: The null hypothesis (H_0) is rejected based on the Linear Regression results, as well as the Correlation Matrix results in the “positive” chatbot, while it is accepted in the “negative” chatbot. Consequently, the users’ trust in a chatbot significantly affects users’ intention to continue using the chatbot when it creates a positive experience for the user, whereas the users' trust in a chatbot has no significant effect on users' intention to continue using it when a negative experience was created.

H5: The null hypothesis (H_0) is rejected based on the Linear Regression results and the Correlation Matrix results in both chatbots. Thus, the users' trust in a chatbot significantly affects user satisfaction.

H6: The null hypothesis (H_0) is rejected based on the third Mann-Whiteny U test results. Therefore, the emotional tone of a chatbot interaction significantly affects user satisfaction.

H7: The null hypothesis (H_0) is rejected based on the last Mann-Whiteny U test results. Hence, the emotional tone of chatbot interaction significantly affects users' intention to engage in positive word-of-mouth.

6 Conclusion

6.1 Conclusions

This thesis examined the impact of chatbot sentiment on user acceptance, highlighting the importance of emotional tone in shaping trust, satisfaction, and technology trusting beliefs. By comparing user responses to two chatbots, one designed to evoke positive emotions and one to elicit negative ones, the analysis confirmed that sentiment significantly affects how users perceive and interact with conversational agents.

The results clearly support the hypothesis that a chatbot's emotional tone influences user sentiment, which in turn affects trust, satisfaction, and both the intention to continue using the chatbot and recommend it to others. Positive sentiment elicitation led to higher scores across all constructs (customer satisfaction, trust, positive emotions, positive word of mouth intention, continuance intention, and technology trusting beliefs).

This reinforces the conclusion that emotion is not just a chatbot feature but a crucial factor in user acceptance. Emotionally intelligent chatbots, particularly those capable of evoking positive feelings, have the potential to foster stronger and better user-chatbot relationships, as well as increase long-term engagement. As conversational agents and AI chatbots become more integrated into the everyday life of users, prioritizing emotionally aware design will become critical to success.

6.2 Theoretical Contribution

This study offers value to the academic community as it enriches the existing literature regarding chatbot sentiments and how they influence user acceptance. Since the research focused on Greek chatbot users, it provides insights that take into account the cultural and economic specificities of Greece, which had not been examined in past studies.

It is important to highlight that the study connects user acceptance with other significant factors such as positive emotions, trust, technology trusting beliefs, and positive word of mouth intention, thus offering a holistic view for a deeper understanding of user sentiments and needs.

6.3 Practical Contribution

Companies in the industry can significantly benefit from the findings of this research to improve the services they provide to their users. A core element of this study is the assessment of user chatbot acceptance based on elicited emotions. A holistic approach to trust and acceptance was presented, examining customer satisfaction, trust, positive emotions, positive word of mouth intention, continuance intention, as well as technology trusting beliefs (functionality, reliability, integrity, competence, and benevolence). Based on all these dimensions, companies can more effectively understand their users' behaviour and implement more targeted strategies to capture a larger market share.

One suggested strategy is to create chatbots that aim to elicit positive emotions to their users. By incorporating empathetic, humorous, and positive language in chatbot interactions, companies can enhance user satisfaction, trust, and willingness to continue using the chatbot. Additionally, companies can enhance trust-building features on their chatbots. They could focus on improving functionality, reliability, and benevolence, as these were strongly correlated with trust in the "positive" chatbot test results. They could also provide transparent information about the chatbot's capabilities and limitations in order to manage user expectations.

Furthermore, another strategy would be to encourage satisfied users to share their experiences by integrating easy-to-use sharing options (e.g., social media buttons or Play Store ratings) within the chatbot interface. That way, the positive testimonials can be highlighted, and credibility can be built around the chatbot. Moreover, companies could use adaptive algorithms to tailor responses based on user preferences and past interactions, thus enhancing perceived competence and emotional connection. Also, they could implement conversational variety to avoid repetitive interactions and enhance the chatbot's anthropomorphism.

Last but not least, a younger and educated demographic could be targeted. Since the study found higher engagement among younger (18-24) and university-educated users, these groups could be prioritized in marketing and development efforts. Subsequently, features that align with their needs, such as academic support or career advice, could be provided and marketed thoroughly.

6.4 Research Limitations

The research conducted had certain limitations.

Firstly, the population sample that responded to the questionnaire was not as extensive as in other studies, consisting of 64 individuals. Therefore, although this sample provides some interesting insights, these may not be generalizable to the broader population. However, it was acknowledged by the supervisor that due to limited time and resources, this is expected and acceptable.

Additionally, the sample used was a convenience sample. The majority belonged to the 18-24 age group. More specifically, 18 out of the 32 (56.25%) respondents of the "positive" chatbot and 20 out of the 32 (62.5%) respondents of the "negative" one, summing it up to 38 out of the 64 (59.37%) of the total respondents. This means the representativeness of the sample is not particularly high, especially considering that the experiences and opinions of young people are extensively reflected. Hence, the results are not particularly descriptive regarding older age groups.

This also applies to the fact that the majority of the sample were either university students or individuals with that education level. More specifically, 20 out of the 32 (62.5%) respondents of the "positive" chatbot and 20 out of the 32 (62.5%) respondents of the "negative" one, summing it up to 40 out of the 64 (62.5%) of the total respondents. This, in turn, may increase the likelihood of data bias and homogeneity in responses, as students and university graduates constitute a distinct demographic group.

Nevertheless, the sample used, while not perfectly representative, still provides useful insights and the ability to extract interesting conclusions.

6.5 Future Research Directions

This thesis lays the groundwork for several future research directions that can enrich the understanding and improvement of sentiment as a chatbot user acceptance factor. Based on the above findings, potential future directions aimed at addressing remaining research gaps and expanding the current study have been identified.

One key area revolves around expanding demographic representation by conducting studies with more diverse age groups and educational backgrounds to ensure broader coverage. Furthermore, replicating the study in non-Greek cultural contexts could shed light on cross-cultural variations in sentiment perception and chatbot acceptance. Longitudinal studies represent another valuable choice, tracking how user perceptions evolve over time through repeated interactions with chatbots, and determining whether initial positive or negative sentiments persist, diminish, or transform with continued use are quite interesting topics.

Advancements in sentiment analysis techniques also present exciting opportunities. Leveraging state-of-the-art natural language processing models such as GPT-4 and BERT could significantly enhance emotion detection and response generation. Also, comparative evaluations of more traditional techniques, such as lexicon-based, machine learning, and transfer learning approaches may yield insights into their respective effectiveness in real-world chatbots.

Additionally, integration with emerging technologies like Augmented Reality (AR) and Virtual Reality (VR) can be explored to create more immersive, emotionally engaging chatbot experiences, while multimodal interactions, such as combining voice or images with text, could further strengthen or diminish user trust and satisfaction. Finally, incorporating behavioral metrics such as session duration and repeat usage can provide objective validation for self-reported sentiment data. Techniques like eye-tracking and facial expression analysis can also be utilized to capture authentic emotional responses.

By implementing these strategies and exploring these research directions, both businesses and researchers can further optimize chatbot interactions and deepen the understanding of sentiment as a chatbot user acceptance factor.

References

Allouch, M., Azaria, A., & Azoulay, R. (2021). Conversational Agents: Goals, Technologies, Vision and Challenges. *Sensors*, *21*(24), 8448. https://doi.org/10.3390/s21248448

Boonstra, L. (2024). *Prompt Engineering*. Google. https://www.kaggle.com/whitepaper-prompt-engineering

Chandra, S., Shirish, A., & Srivastava, S. C. (2022). To Be or Not to Be …Human? Theorizing the Role of Human-Like Competencies in Conversational Artificial Intelligence Agents. *Journal of Management Information Systems*, *39*(4), 969–1005. https://doi.org/10.1080/07421222.2022.2127441

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Eyu, J. M., Yau, K.-L. A., Liu, L., & Chong, Y.-W. (2024). Reinforcement learning in sentiment analysis: a review and future directions. *Artificial Intelligence Review*, *58*(1), 6. https://doi.org/10.1007/s10462-024-10967-0

Han, E., Yin, D., & Zhang, H. (2023). Bots with Feelings: Should AI Agents Express Positive Emotion in Customer Service? *Information Systems Research*, *34*(3), 1296–1311. https://doi.org/10.1287/isre.2022.1179

Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). More than a Feeling: Accuracy and Application of Sentiment Analysis. *International Journal of Research in Marketing*, *40*(1), 75–87. https://doi.org/10.1016/j.ijresmar.2022.05.005

Hornbæk, K., & Hertzum, M. (2017). Technology Acceptance and User Experience: A Review of the Experiential Component in HCI. *ACM Transactions on Computer-Human Interaction*, *24*(5), 1–30. https://doi.org/10.1145/3127358

Hui, Z., Khan, A. N., Chenglong, Z., & Khan, N. A. (2024). When Service Quality is Enhanced by Human–Artificial Intelligence Interaction: An Examination of Anthropomorphism, Responsiveness from the Perspectives of Employees and Customers. *International Journal of Human–Computer Interaction*, *40*(22), 7546–7561. https://doi.org/10.1080/10447318.2023.2266254

Keijsers, M., Bartneck, C., & Kazmi, H. S. (2019). Cloud-Based Sentiment Analysis for Interactive Agents. *Proceedings of the 7th International Conference on Human-Agent Interaction*, 43–50. https://doi.org/10.1145/3349537.3351883

Krugmann, J. O., & Hartmann, J. (2024). Sentiment Analysis in the Age of Generative AI. *Customer Needs and Solutions*, *11*(1), 3. https://doi.org/10.1007/s40547-024-00143-4

Lankton, N., McKnight, D., & Tripp, J. (2015). Technology, Humanness, and Trust: Rethinking Trust in Technology. *Journal of the Association for Information Systems*, *16*(10). https://doi.org/10.17705/1jais.00411

Ling, E. C., Tussyadiah, I., Tuomi, A., Stienmetz, J., & Ioannou, A. (2021). Factors influencing users' adoption and use of conversational agents: A systematic review. *Psychology & Marketing*, *38*(7), 1031–1051. https://doi.org/10.1002/mar.21491

Luo, B., Lau, R. Y. K., & Li, C. (2023). Emotion-regulatory chatbots for enhancing consumer servicing: An interpersonal emotion management approach. *Information & Management*, *60*(5), 103794. https://doi.org/10.1016/j.im.2023.103794

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), 1–25. https://doi.org/10.1145/1985347.1985353

Ng, S. W. T., & Zhang, R. (2025). Trust in AI chatbots: A systematic review. *Telematics and Informatics*, *97*, 102240. https://doi.org/10.1016/j.tele.2025.102240

Saffarizadeh, K., Keil, M., & Maruping, L. (2024). Relationship Between Trust in the AI Creator and Trust in AI Systems: The Crucial Role of AI Alignment and Steerability. *Journal of Management Information Systems*, *41*(3), 645–681. https://doi.org/10.1080/07421222.2024.2376382

Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, *37*(3), 875–900. https://doi.org/10.1080/07421222.2020.1790204

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

Yang, R., & Wibowo, S. (2022). User trust in artificial intelligence: A comprehensive conceptual framework. *Electronic Markets*, *32*(4), 2053–2077. https://doi.org/10.1007/s12525-022-00592-6

Figures

Tables

Appendix

1 Questionnaire Constructs

Customer Satisfaction by Hui et al. (2024)	
CS1	I am pleased with using the chatbot.
CS2	I am satisfied with the consultation experience of using the chatbot (e.g., quality of feedback, provided ideas and recommendations).
CS3	I am satisfied with my overall experience using the chatbot.
CS4	I would recommend that others use the chatbot.
Trust by Luo et al. (2023)	
T1	I find the chatbot trustworthy.
T2	I can trust the information presented by the chatbot.
T3	I value the trustworthy characteristics of the chatbot.
Positive Emotions by Luo et al. (2023)	
PE1	I feel happy after using the chatbot.
PE2	I have a warm feeling after using the chatbot.
PE3	I am being pleasured after using the chatbot.
PE4	The chatbot makes me enjoyed.
PE5	The chatbot makes me feel relaxed.
Positive Word of Mouth by Luo et al. (2023)	
PWoM1	I would recommend that my friends and relatives use the chatbot in the future.
PWoM2	I intend to be proud to say to others that I am this chatbot's user.
PWoM3	I intend to mostly say positive things about this chatbot to others.
PWoM4	I speak of this chatbot's positive experience to others.
PWoM5	I am proud to say to others that I am this chatbot's user on a social media platform.
Continuance Intention by Lankton et al. (2015)	
CI1	I intend to continue using the chatbot.
CI2	I predict that I will use the chatbot again in the future.
Technology Trusting Belief - Functionality by Lankton et al. (2015)	
TBF1	The chatbot has the functionality I need.
TBF2	The chatbot has the features required for my tasks.
TBF3	The chatbot has the ability to do what I want it to do.
Technology Trusting Belief - Reliability by Lankton et al. (2015)	
TBR2	The chatbot is very reliable.
TBR2	The chatbot does not fail me.
TBR3	The chatbot is extremely dependable.
Technology Trusting Belief - Integrity by Lankton et al. (2015)	
TBI1	The chatbot is truthful in its dealings with me.
TBI2	The chatbot is honest.
Technology Trusting Belief - Competence by Lankton et al. (2015)	
TBC1	The chatbot is competent and effective in providing feedback and recommendations.
TBC2	The chatbot performs its role of start-up consultant very well.
TBC3	The chatbot is a capable and proficient consultant.
Technology Trusting Belief - Benevolence by Lankton et al. (2015)	

TBB1	The chatbot acts in my best interest.
TBB2	The chatbot does its best to help me with my start-up.
TBB3	The chatbot is interested in my well-being.

2 Questionnaire

This questionnaire has been created as part of a bachelor's thesis at the Department of Management Science and Technology at the Athens University of Economics and Business (AUEB), titled "Chatbots & Conversational Agents: Sentiment as a User Acceptance Factor".

The aim is to capture and analyze the acceptance and adoption of Chatbots and Conversational Agents when they exhibit different behaviors and influence users' emotional responses.

If you have any questions, feel free to contact me at: t8210039@aueb.gr

Estimated completion time: 5-10 minutes.

1. Gender (Male, Female, Other)
2. Age (18-24, 25-34, 35-49, 50+)
3. Education (Middle School, High School, University, Post-graduate Studies)
4. Frequency of chatbot use (e.g., ChatGPT, DeepSeek)
5. I am pleased with using the chatbot.
6. I am satisfied with the consultation experience of using the chatbot (e.g., quality of feedback, provided ideas and recommendations).
7. I am satisfied with my overall experience using the chatbot.
8. I would recommend that others use the chatbot.
9. I find the chatbot trustworthy.
10. I can trust the information presented by the chatbot.
11. I value the trustworthy characteristics of the chatbot.
12. I feel happy after using the chatbot.
13. I have a warm feeling after using the chatbot.
14. I am being pleasured after using the chatbot.
15. The chatbot makes me enjoyed.
16. The chatbot makes me feel relaxed.
17. I would recommend that my friends and relatives use the chatbot in the future.
18. I intend to be proud to say to others that I am this chatbot's user.
19. I intend to mostly say positive things about this chatbot to others.
20. I speak of this chatbot's positive experience to others.
21. I am proud to say to others that I am this chatbot's user on a social media platform.
22. I intend to continue using the chatbot.
23. I predict that I will use the chatbot again in the future.
24. The chatbot has the functionality I need.
25. The chatbot has the features required for my tasks.
26. The chatbot has the ability to do what I want it to do.
27. The chatbot is very reliable.
28. The chatbot does not fail me.
29. The chatbot is extremely dependable.
30. The chatbot is truthful in its dealings with me.
31. The chatbot is honest.

32. The chatbot is competent and effective in providing feedback and recommendations.
33. The chatbot performs its role of start-up consultant very well.
34. The chatbot is a capable and proficient consultant.
35. The chatbot acts in my best interest.
36. The chatbot does its best to help me with my start-up.
37. The chatbot is interested in my well-being.

3 Questionnaire Items Descriptive Statistics

The titles refer to:

- CS: Customer Satisfaction
- T: Trust
- PE: Positive Emotions
- PWOM: Positive Word of Mouth
- CI: Continuance Intention
- TBF: TTB - Functionality
- TBR: TTB - Reliability
- TBI: TTB - Integrity
- TBC: TTB - Competence
- TBB: TTB - Benevolence

For the “negative” chatbot the descriptive statistics values are:

	Mean	Median	Mode	Standard Deviation	Variance
CS1	4.59375	4.5	4.0	1.214413	1.474798
CS2	4.71875	5.0	5.0	1.464293	2.144153
CS3	4.71875	5.0	4.0	1.300977	1.692540
CS4	4.71875	5.0	5.0	1.570584	2.466734
T1	4.87500	5.0	4.0	1.338029	1.790323
T2	4.81250	5.0	5.0	1.378112	1.899194
T3	4.75000	4.5	4.0	1.270001	1.612903
PE1	4.31250	4.0	4.0	1.490562	2.221774
PE2	3.68750	4.0	5.0	1.785922	3.189516
PE3	3.96875	4.0	4.0	1.655575	2.740927
PE4	3.65625	4.0	4.0	1.598575	2.555444
PE5	3.78125	4.0	2.0	1.496973	2.240927
PWOM1	4.62500	5.0	6.0	1.601411	2.564516
PWOM2	3.75000	4.0	4.0	1.626395	2.645161
PWOM3	4.31250	4.5	5.0	1.574750	2.479839
PWOM4	4.21875	4.0	5.0	1.621118	2.628024
PWOM5	3.18750	3.0	3.0	1.615200	2.608871
CI1	4.03125	4.0	3.0	1.768622	3.128024
CI2	4.37500	5.0	5.0	1.809473	3.274194
TBF1	4.37500	4.0	4.0	1.581139	2.500000
TBF2	4.59375	5.0	5.0	1.477997	2.184476
TBF3	4.50000	4.5	4.0	1.481063	2.193548
TBR1	4.43750	4.0	4.0	1.242721	1.544355
TBR2	4.06250	4.0	4.0	1.318296	1.737903
TBR3	3.93750	4.0	3.0	1.216486	1.479839
TBI1	4.96875	5.0	4.0	1.092035	1.192540
TBI2	5.06250	5.0	4.0	1.189673	1.415323
TBC1	4.68750	5.0	5.0	1.533234	2.350806
TBC2	4.50000	5.0	5.0	1.414214	2.000000

TBC3	4.15625	4.0	4.0	1.393750	1.942540
TBB1	4.37500	4.5	5.0	1.497309	2.241935
TBB2	4.34375	5.0	5.0	1.658008	2.748992
TBB3	3.18750	3.0	3.0	1.512048	2.286290

For the "positive" chatbot the descriptive statistics values are:

	Mean	Median	Mode	Standard Deviation	Variance
CS1	6.00000	6.0	7	1.077632	1.161290
CS2	6.09375	6.5	7	1.058281	1.119960
CS3	6.09375	6.0	7	0.995453	0.990927
CS4	5.50000	6.0	7	1.586231	2.516129
T1	5.59375	5.5	7	1.316383	1.732863
T2	5.50000	5.0	7	1.319824	1.741935
T3	5.53125	5.0	5	1.163542	1.353831
PE1	5.40625	6.0	7	1.915293	3.668347
PE2	5.50000	6.0	7	1.900764	3.612903
PE3	5.62500	6.0	7	1.601411	2.564516
PE4	5.50000	6.0	7	1.778002	3.161290
PE5	5.31250	6.0	7	1.767767	3.125000
PWOM1	5.15625	6.0	7	1.833547	3.361895
PWOM2	5.03125	5.0	7	1.874866	3.515121
PWOM3	5.56250	6.0	7	1.721543	2.963710
PWOM4	5.50000	6.0	7	1.684847	2.838710
PWOM5	4.78125	5.0	7	2.166227	4.692540
CI1	5.25000	5.5	7	1.813925	3.290323
CI2	5.06250	6.0	7	2.062531	4.254032
TBF1	5.31250	5.0	7	1.767767	3.125000
TBF2	5.50000	6.0	7	1.665591	2.774194
TBF3	5.59375	6.0	7	1.411002	1.990927
TBR1	5.18750	5.0	7	1.468761	2.157258
TBR2	5.28125	5.0	4	1.275941	1.628024
TBR3	5.09375	5.0	5	1.488870	2.216734
TBI1	5.46875	5.0	7	1.217729	1.482863
TBI2	5.59375	6.0	7	1.240691	1.539315
TBC1	5.78125	7.0	7	1.539467	2.369960
TBC2	5.59375	6.0	7	1.542084	2.378024
TBC3	5.40625	6.0	7	1.521022	2.313508
TBB1	5.50000	6.0	7	1.524002	2.322581
TBB2	5.43750	5.0	7	1.435439	2.060484
TBB3	5.09375	5.5	7	1.855408	3.442540

4 OLS Regression Results

For the “negative” chatbot the first OLS regression results are:

Dep. Variable:	Trust	R-squared:	0.390
Model:	OLS	Adj. R-squared:	0.348
Method:	Least Squares	F-statistic:	9.283
Date:	Sat, 31 May 2025	Prob (F-statistic):	0.000765
Time:	12:49:53	Log-Likelihood:	-43.748
No. Observations:	32	AIC:	93.50
Df Residuals:	29	BIC:	97.89
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.9436	0.689	2.821	0.009	0.535	3.353
Continuance Intention	0.0259	0.138	0.187	0.853	-0.257	0.309
Customer Satisfaction	0.5888	0.182	3.228	0.003	0.216	0.962

Omnibus:	0.150	Durbin-Watson:	1.746
Prob(Omnibus):	0.928	Jarque-Bera (JB):	0.208
Skew:	-0.142	Prob(JB):	0.901
Kurtosis:	2.724	Cond. No.	26.4

For the “positive” chatbot the first OLS regression results are:

Dep. Variable:	Trust	R-squared:	0.757
Model:	OLS	Adj. R-squared:	0.740
Method:	Least Squares	F-statistic:	45.08
Date:	Sat, 31 May 2025	Prob (F-statistic):	1.26e-09
Time:	13:35:15	Log-Likelihood:	-28.581
No. Observations:	32	AIC:	63.16
Df Residuals:	29	BIC:	67.56
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.9503	0.812	1.170	0.251	-0.710	2.611
Continuance Intention	0.3108	0.112	2.782	0.009	0.082	0.539
Customer Satisfaction	0.5047	0.208	2.421	0.022	0.078	0.931

Omnibus:	1.078	Durbin-Watson:	2.389
Prob(Omnibus):	0.583	Jarque-Bera (JB):	0.875
Skew:	0.093	Prob(JB):	0.646
Kurtosis:	2.212	Cond. No.	62.1

For the "negative" chatbot the second OLS regression results are:

Dep. Variable:	Trust	R-squared:	0.612
Model:	OLS	Adj. R-squared:	0.538
Method:	Least Squares	F-statistic:	8.208
Date:	Fri, 13 Jun 2025	Prob (F-statistic):	9.31e-05
Time:	17:10:04	Log-Likelihood:	-36.510
No. Observations:	32	AIC:	85.02
Df Residuals:	26	BIC:	93.81
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.4134	0.861	0.480	0.635	-1.356	2.183
TTB - Functionality	-0.0590	0.175	-0.337	0.739	-0.419	0.301
TTB - Reliability	0.6849	0.258	2.651	0.013	0.154	1.216
TTB - Integrity	0.2769	0.197	1.408	0.171	-0.127	0.681
TTB - Competence	0.0841	0.247	0.341	0.736	-0.423	0.591
TTB - Benevolence	0.0155	0.204	0.076	0.940	-0.403	0.435

Omnibus:	7.502	Durbin-Watson:	1.753
Prob(Omnibus):	0.023	Jarque-Bera (JB):	5.920
Skew:	-0.857	Prob(JB):	0.0518
Kurtosis:	4.226	Cond. No.	60.1

For the "positive" chatbot the second OLS regression results are:

Dep. Variable:	Trust	R-squared:	0.920
Model:	OLS	Adj. R-squared:	0.905
Method:	Least Squares	F-statistic:	59.91
Date:	Fri, 13 Jun 2025	Prob (F-statistic):	1.94e-13
Time:	17:13:41	Log-Likelihood:	-10.755
No. Observations:	32	AIC:	33.51
Df Residuals:	26	BIC:	42.30
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.5568	0.337	1.652	0.111	-0.136	1.250
TTB - Functionality	0.3957	0.144	2.742	0.011	0.099	0.692
TTB - Reliability	-0.2272	0.157	-1.444	0.161	-0.551	0.096
TTB - Integrity	0.5011	0.114	4.389	0.000	0.266	0.736
TTB - Competence	-0.1466	0.121	-1.213	0.236	-0.395	0.102

Omnibus:	1.107	Durbin-Watson:	2.335
Prob(Omnibus):	0.575	Jarque-Bera (JB):	1.102
Skew:	-0.358	Prob(JB):	0.576
Kurtosis:	2.440	Cond. No.	65.2