HELLENIC OPEN UNIVERSITY

School of Social Sciences

Master in Business Administration (MBA)

Postgraduate Dissertation

Sentiment Analysis and User Emotional Disposition During Service Chatbot Interaction in Online Service Environments

Dimitra Mitsi

Supervisor: Georgios Oikonomidis

Patras, Greece, June 2026

Sentiment Analysis and User Emotional Disposition During Service Chatbot Interaction in Online Service Environments

Dimitra Mitsi

Supervising Committee

Supervisor:	Co-Supervisor:
Georgios Oikonomidis	Dimitrios Christopoulos

Patras, Greece, June 2026

I would like to express my sincere gratitude to my supervisor for the guidance and feedback provided throughout this dissertation.

I would also like to thank my family and my close friends for their continuous encouragement and support throughout this journey.

Abstract

This present study explores how chatbot communication styles influence user experience in customer service environments. More specifically, it examines whether an empathetic, more human-like chatbot results in higher user experience than a professional, more neutral chatbot. The study concentrates on four key variables: Emotional response, Satisfaction , Perceived service quality , and Trust. To address the purpose of this study a quantitative experimental design was used. Two chatbot versions were designed for the study, with the only difference being the tone: one had an empathetic (human-like), while the other a professional (neutral) tone, keeping all the other elements the same. Then participants were randomly assigned to interact with one of the two chatbots in a banking scenario, where they were asked to imagine that they had lost their card. After completing the conversation, participants were asked to complete a questionnaire. In total, 107 participants comprised the sample. IBM SPSS Statistics was used for data analysis. The results showed that the empathetic chatbot received higher scores across all key variables. In particular, participants who interacted with the empathetic chatbot reported more positive Emotional responses. Also, they reported higher Satisfaction , better perception of service quality, and higher levels of Trust than those who interacted with the professional chatbot. Moreover, the strongest predictors of Trust were revealed to be Emotional response and Perceived service quality . After further investigation through mediation analysis, the findings suggested that Emotional response is a key mechanism in explaining how Empathy enhances both Satisfaction and Trust. Overall, the findings show that chatbot communication style influences Trust and the overall user experience of interaction.

Keywords

AI chatbots, Chatbot Empathy, Emotional response, Satisfaction , Perceived service quality , Trust

Ανάλυση Συναισθήματος και Συναισθηματική Προδιάθεση Χρηστών σε Αλληλεπιδράσεις με Chatbots Εξυπηρέτησης Πελατών

Δήμητρα Μήτση

Περίληψη

Η παρούσα μελέτη εξετάζει πώς τα διαφορτεικά στυλ επικοινωνίας που μπορούν να χρησιμοποιηθούν στα chatbots επηρεάζουν την γενικότερη εμπειρία του χρήστη σε διαδικτυακά περιβάλλοντα εξυπηρέτησης πελατών. Αναλυτικότερα, η μελέτη εξετάζει αν ένα ενσυναισθητικό (ανθρωποκεντρικό) chatbot προσφέρει καλύτερη εμπειρία χρήστη σε σχέση με ένα πιο επαγγελματικό (ουδέτερο) chatbot. Οι τέσσερις βασικές μεταβλητές στις οποίες επικεντρώνεται η συγκεκριμένη μελέτη είναι η συναισθηματική απόκριση, η ικανοποίηση, η αντιλαμβανόμενη ποιότητα υπηρεσίας και η εμπιστοσύνη. Για να ερευνηθεί η συγκεκριμένη σχέση χρησιμοποιηθήκε ποσοτικός πειραματικός σχεδιασμός. Πιο συγκεκριμένα, σχεδιάστηκαν δύο εκδοχές chatbot με μοναδική διαφορά τον τόνο επικοινωνίας: η μία με ευσυναισθηματικό τόνο, ενώ η άλλη πιο επαγγελματικό τόνο, διατηρώντας όλα τα υπόλοιπα στοιχεία ίδια. Οι συμμετέχοντας επιλέχθηκαν τυχαία ώστε να αλληλεπιδράσουν με ένα από τα δύο chatbots σε ένα τραπεζικό σενάριο, στο οποίο τους ζητήθηκε να φανταστούν ότι είχαν χάσει την τραπεζική τους κάρτας. Αφού ολοκλήρωσαν την συνομιλίας τους, κλήθηκαν να συμπληρώσουν ένα ερωτηματολόγιο. Συνολικά, συγκεντρώθηκαν απαντήσεις από 107 συμμετέχοντες. Για την ανάλυση των δεδομένων χρησιμοποιήθηκε το IBM SPSS Statistics. Τα αποτελέσματα της έρευνας έδειξαν ότι το ενσθναισθηματικό chatbot είχε υψηλότερες βαθμολογίες σε όλες τις εξεταζόμενες μεταβλητές. Συγκεκριμένα, παρατηρήθηκε ότι όσοι συμμετέχοντες αλληλεπιδράσανε με το ενσθναισθηματικό chatbot αναφέρανε θετικότερη συναισθηματική ανταπόκριση και αντίληψη της ποιότητας καθώς και υψηλότερη ικανοποίηση και εμπιστόσυνη σε σχέση με εκείνους που συνομιλήσανε με το επαγγελαμτικό chatbot. Επιπρόσθετα, διαπιστώθηκε ότι η συναισθηματική ανταπόκριση και η αντιλαμβανόμενη ποιότητα υπηρεσίας είχαν ισχυρότερο προβλεπτικό ρόλο στην εμπιστοσύνη. Τα ευρήματα της ανάλυσης διαμεσολάβησης έδειξαν, επίσης, ότι η συναισθηματκή απόκριση είναι ο βασικός μηχανισμός μέσω του οποίου η ενσυναίσθηση ενισχύει τόσο τον εμπιστοσύνη όσο και την συνολική εμπειρία του χρήση κατά την αλληλεπίδαση με το chatbot.

Λέξεις – Κλειδιά

Chatbots τεχνητής νοημοσύνης, Ενσυναίσθηση chatbot, Συναισθηματική απόκριση, Ικανοποίηση πελατών, Αντιλαμβανόμενη ποιότητα υπηρεσίας, Εμπιστοσύνη

Table of Contents

List of Tables

1. Introduction

1.1 Background and purpose of study

In modern customer service environments, the use of Artificial Intelligence (AI) has become increasingly important, as organizations in digital environments are looking for more efficient and faster ways to improve customer experience. AI chatbots have become an common part of service operations, as they can boost operational efficiency and offer better customer experience. Nevertheless, customer service is not evaluated only on the basis of the problem is solved. Also, the way the support is delivered influence users' evaluation of the service. For this reason, it is important to explore whether the communication style of an AI chatbot influences users' willingness to interact with the chatbot and if their needs are met during the interaction.

The purpose of this study is to examine whether chatbot communication style impacts user experience in customer service environments. The study focuses on whether an empathetic, human-like leads to a positive overall experience, including better perception of service, higher Trust and Satisfaction levels than a professional, neutral chatbot. In this way, the study investigates further whether communication style should be considered an important feature of chatbot design and examines how it can influence users' evaluation of AI-enabled chatbots.

1.2 Research questions

Based on the purpose of the study, the research questions are the following:

<u>Research Question 1 (RQ1)</u>:

Do users interacting with an empathetic (human-like) chatbot report **more positive Emotional responses** than users interacting with a professional (neutral) chatbot?

<u>Research Question 2 (RQ2)</u>:

Do users interacting with an empathetic (human-like) chatbot report **higher Satisfaction** than users interacting with a professional (neutral) chatbot?

Research Question 3 (RQ3):

Do users interacting with an empathetic (human-like) chatbot **perceive higher service quality** than users interacting with a professional (neutral) chatbot?

Research Question 4 (RQ4):

Do users interacting with an empathetic (human-like) chatbot report **higher Trust** in the service provider than users interacting with a professional (neutral) chatbot?

1.3 Methodology and approach

This study adopts a quantitative approach and uses an experimental design. Specifically, two chatbot versions were designed: one with an empathetic (human-like) tone and one with professional (neutral) tone. Participants were asked to interact with one of the two chatbots while imagining that it was their bank's AI-chatbot and to communicate with it in order to block their card after losing it. After completing the interaction, they were asked to complete a questionnaire, measuring the key variables of the present study. Lastly, the data were analyzed using IBM SPSS Statistics. The analysis included descriptive statistics, reliability analysis, t-tests, regression analysis, and mediation analysis.

1.4 Main findings and limitations

The findings of this study highlight that communication style is important for uses' Trust and overall user experience. More specifically, the empathetic, human-like chatbot received more positive evaluations than the professional, neutral chatbot across the key examined variables. Moreover, the mediation analysis suggested that the Emotional response may act as linking mechanism with Trust and Satisfaction. Also, the results of the hierarchical regression further showed that the strongest predictors of Trust were Emotional response and Perceived service quality. Thus, Trust and Satisfaction did not explain additional variance in Trust more than these variables.

As for the limitations, this study has certain. Firstly, the sample was selected through convenience sampling, which may limit the generalization of the results. Secondly, the research focused on simulated scenario rather than in a real service interaction, which may means that participants did not respond in the same they would in a real situation Thirdly, the collected data were self-reported and may have been influenced by participants' subjective perceptions or the response bias. Thus, the results should be interpreted within the present study.

1.5 Structure of the study

The following chapters are organized as follows. Chapter 2 presents the literature review and the main constructs relevant to the study, namely AI chatbots, Empathy, Emotional response, Satisfaction, Perceived service quality, and Trust. Chapter 3 explains the methodology, including the design, information about the sample, tools used in the study, and data analysis. Then, Chapter 4 outlines the findings of the empirical analysis. Lastly, Chapter 5 discusses the findings in relation to the literature, explains the theoretical and managerial implications of the study, outlines its limitations, and offers suggestions for future research.

2. Literature Review

2.1 Introduction

Artificial intelligence (AI) has become a crucial part of the landscape in today's customer service, fundamentally changing the way companies engage with their customers. In banking, retail, telecommunication and health, conversational agents (also known as chatbots) now process a significant proportion of the most common types of service enquiries. (Huang & Rust, 2018, 2021; Følstad & Brandtzaeg, 2017) While there are definite advantages of scalability and cost benefits with using chatbots in services, there are also some challenges. Not all customers are satisfied with chatbots that do not pick up on the affective elements in interactions, when they feel interactions are mechanical or when they are directed to a human agent too late (Adam, Wessel, & Benlian, 2021; Luo, Tong, Fang, & Qu, 2019). These issues highlight the importance of communication style, empathy, Emotional response, Satisfaction , Perceived service quality and Trust in AI – enabled service chatbots.

The literature reviewed in this chapter is directly relevant to the present study, as it supports the theoretical and empirical foundation for exploring chatbot communication style, perceived empathy, Emotional response, Satisfaction, Perceived service quality and Trust. Instead of offering each source separately, the review adopts an approach by comparing findings across the literature, identifying areas of agreement and disagreement, taking in to consideration the theoretical, contextual and methodological factors that may explain these patterns. As a result, the current study in the wider body of academic research, introduces the gaps that drive the research questions, and supports the constructs and measurement decisions made in the empirical research.

The chapter is organized according to the key concepts of the study. This literature review in section 2.2 explores the field of AI chatbots in customer service and the theoretical frameworks it is built upon. In Section 2.3, the literature is reviewed to examine the difference between empathetic and professional style in conversation with a chatbot. In Section 2.4, human-chatbot perceived Empathy is discussed. Section 2.5 looks at Emotional response and further explains why we chose to use the operationalization of sentiment being an Emotional response, as self-reported on a questionnaire after the interaction, as opposed to the automated text-based sentiment classification. The literature review on user Satisfaction and Perceived service quality is presented in Sections 2.6 and 2.7. Section 2.8 will cover the customer service aspect of Trust in AI. Section 2.9 brings together the literature, outlines gaps and provides a conceptual framework for the empirical study.

2.2 Understanding the role of AI Chatbots in Customer Service

Chatbots are conversational software agents that use natural language to communicate with users through text or voice. The first chatbots was ELIZA, which was rule-based and had limited conversational capabilities (Weizenbaum, 1966). Large language models are now being used to develop contemporary chatbots for commercial use, which dynamically generate responses, enabling more flexible and context-aware conversations (Brown et al., 2020; Wei et al., 2022). This technological change has broadened the spectrum of tasks that it is plausible for chatbots to accomplish and raised critical questions regarding the customer experience provided of these technologies.

Huang and Rust (2018, 2021) suggest that the intelligences that AI can display in services can be divided into four types of intelligence: mechanical, analytical, intuitive, and empathetic. They argue that the shift from analytical and empathetic AI is redefining the role of humans and machines in service environments. Althoug, their framework has been influential, it has also been criticized. From a customer experience perspective, the framework can be seen as abstract and provides limited information about the specific design elements that enable Empathy in a particular chatbot (Larivière et al., 2017; Wirtz et al., 2018). The framework also appears to treat empathetic AI as most preferable direction for service development, but there is mixed empirical evidence regarding users' reactions to chatbots using an empathetic communication tone (Liu & Sundar, 2018; Pelau, Dabija & Ene, 2021).

A second research tradition examines chatbots from the perspective of technology acceptance. The user acceptance literature, including the Technology Acceptance Model (TAM; Davis, 1989), the Unified Theory of Acceptance and Use of Technology (UTAUT; Venkatesh et al., 2003), and other information systems literature, views user acceptance as a function of beliefs related to perceived usefulness, perceived ease of use, and other factors. Studied applying these models for chatbots has revealed that perceived usefulness, perceived enjoyment, and social presence are related to intention to use (Brachten, Kissmer & Stieglitz, 2021; Rese, Ganster & Baier, 2020). The technology acceptance tradition, however, has been criticized for mostly focusing on cognitive aspects of user evaluations, while paying less attention to provided responses (Bagozzi, 2007; Venkatesh & Bala, 2008). In the case of empathetic chatbots, this is an important constraint since emotional and supportive communication is a key feature of these systems.

A third tradition is based on the service literature and service-dominant logic (SDL), both of which view the service encounter as a co-created interaction, not a discrete event, between the service provider and the customer (Vargo & Lusch, 2004, 2008). This view conceptualizes the chatbot as not just a service provider but a participant in a service system whose responses influence customer's perception of perceived service value (Bolton et al., 2018). This perspective aligns with the constructs explored in the current research, as empathy, Emotional response, Satisfaction , Perceived service quality and Trust reflect user's evaluation of the service encounter itself and not the technology per se.

A fourth, more recent research; it is about the negative aspects of service automation. This stream of work highlights the situation in which a chatbot has negative effects on customers. Empathetic chatbots enhance user engagement in certain situations, while also reducing sensitivity to economic incentives in others. When the social presence of a service robot or chatbot does not understand the task, it can create discomfort, as suggested by Wirtz et al. (2018). Castelo, Bos, and Lehmann (2019) found that algorithms aversion grows when users perceive a task requiring subjective judgement. The stream provides an important counterpoint to the positive framings that are positive views in the literature on the adoption of chatbots and highlights the importance of seriously considering the possibility that empathetic chatbot framings might not necessarily lead to positive evaluations by customers.

In all of these traditions, the main criticism that can be leveled towards the empirical literature on customer chatbots is that it is not cohesive. Research frequently focuses on a specific feature of the chatbot, employs small and unrepresentative samples and applies outcome measures that are heterogeneous (Janssen, Passlick, Rodríguez Cardona, & Breitner, 2020). This fragmentation makes it difficult to come to definite conclusions on the design features that enhance customer outcomes. The other difficulty is the fast rate of change in technology. The results of rule-based chatbots may not be applicable to the large-language-model systems as they generate responses that are more fluent and contextually appropriate (Brown et al., 2020). Similarly, insights into consumer expectations from chatterbots, which were a novelty in 2017 or 2018, may not apply to a modern user base, where many of the respondents have a large number of experiences with the technology. In the present study, one aspect of the wider gap is explored: the design of an empathetic versus a professional communication style is compared to a consistent set of recognized service outcomes, and in a user population that is overwhelmingly familiar with the technology at hand.

2.3 Chatbot Communication Style

Communication style refers to the way the message is delivered—this is not necessarily the informational content of the delivered message. In interpersonal communication research, style has been described as competence versus warmth (Norton, 1978; Fiske, Cuddy, & Glick, 2007), and task-oriented versus socially-oriented (Norton, 1978, Fiske, Cuddy, & Glick, 2007). A common distinction in the literature on chatbots is the warm/neutral and task-oriented versus human-like communication (Araujo, 2018; Go & Sundar, 2019).

A number of empirical research studies have found that more human-like chatbot communication styles are associated with positive user evaluations. Araujo (2018) concluded that anthropomorphic cues in chatbot communication may improve user's perception of the chatbot and the company that represents. Go and Sundar (2019) demonstrated that the combination of high interactivity in the messages and high human-like identity cues resulted in more positive user perception, although these effects dependent on the consistency between identity cues the messages. Adam et al. (2021) found that social cues, such as small talk and reciprocity movements, led to greater user compliance with chatbot requests. Pelau et al. (2021) found that the perceived Empathy and likability of the device had positively influenced acceptance of AI devices used in service contexts.

However, the overall picture is more complex. A growning body of research suggests a context that there are contexts where warmer communications styles do not work or even may have negative impact on individual evaluations. When customers feel they are being deceived, Mozafari, Weiger, and Hammerschmidt (2022) found that disclosure and empathetic framing may reduce customer Trust. In a field experiment, Luo et al. (2019) found revealing a chatbot's identity at the beginning of the conversation lowered purchase rates by approximately 40% without no significant difference in performance between the chatbot and human agents. While the idea of a warm versus neutral communication style seems straightforward, Liu and Sundar (2018) suggested that expressions of sympathy were more preferred than expressions of Empathy in a health-advice chatbot, indicating a more complicated view of the contrast. The results show that users' expectations, task domain and the alignment of the stylistic cues used by the chatbot are important and the communication style is not necessarily lead to positice user evaluations.

Another important issue conserns the way communication style is being operationalised in previous studies. Some studies that do so at the lexical level, for example by adding hedging language and/or emoji into the texts (Beattie, Edwards, & Edwards, 2020). Some manipulate identity attributes, like the chatbot's name and avatar (Araujo, 2018). Others combine two or more cues together within the same manipulation, making it difficult to identify the effect of each cue used (Adam et al., 2021). This research contributes to the literature by keeping the underlying language model, task flow, and survey instrument while manipulating only the communication tone through system-prompt instructions. This design examines the impact of communication style more specific than designs that include stylistic and identity manipulations under the same experimental condition.

Computers Are Social Actors (CASA) is a theoretical framework that appears to be particular useful for understanding the effects of communication style (Nass & Moon, 2000; Reeves & Nass, 1996). The CASA paradigm suggests that users use social heuristics while interacting with computers that display social properties, even when they are aware that they are interacting with a computer. In this context, users may evaluate empathy, Satisfaction and Trust based on the linguistic cues that provide an empathetic and more understanding interaction. However, the CASA paradigm has been criticized for not clearly specifying the context conditions for the activation of social schemas, and for potential overemphasis on the strength of the response in situations where users already have substantial experience with technology (Gambino, Fox, & Ratan, 2020). The critiques highlight the need of futher researcg focusing on a population that has a much higher prior exposure to chatbots, where positive effects cannot be explained only by the novelty of technology.

The literature on anthropomorphism provides a complementary perspective. Epley, Waytz and Cacioppo (2007) suggest that elicited agent knowledge, reflectance motivation and sociality motivation explain the tendency to attribute human-like qualities to non-human agents. When applied to chatbots, this framework proposes that warm communication cues may encourage users to attribute mental and emotional states to the chatbot, thus influencing the experience of interaction. It is important to note that the theory suggests that anthropomorphic responses vary across users and contexts. Users are more likely to perceive as human-like a chatbot when they feel lonely, anxious, or if the users are highly motivated to seek social interaction, while the users who approach the interaction with a purely instrumental mindset may not anthropomorphize. This boundary condition is relevant to chatbot research as it suggests that the impact of a communication style makes may differ across users.

2.4 Perceived Empathy in Human-Chatbot Interaction

Empathy is a well – established concept in psychology, where it is generally described as the ability to understand and experience the emotions of others (Davis, 1983; Decety & Jackson, 2004). Davis (1983) identifies two types of empathy, cognitive Empathy (understanding another's perspective) and affective Empathy (sharing another's emotions). It is important to note that this distinction is relevant to the chatbot literature yet while the ability for affective empathy, in any strong psychological sense, is highly questionable in the case of chatbots, the ability for cognitive Empathy is more plausible at the level of interaction, as the chatbot can still simulate cognitive Empathy through linguistic acknowledgement of the user's situation.

Empathy has been recognized in the service marketing literature as a key dimension of Perceived service quality. Parasuraman, Zeithaml and Berry (1988) identified Empathy as one of the five SERVQUAL dimensions describing it as the caring, individualized attention that a firm offers to its customers. Wieseke, Geigenmüller and Kraus (2012) showed that the Empathy of a front-line employee enhances customer Satisfaction and customer loyalty across a variety of service contexts. These traditions are relevant to chatbots studies because they indicate that Empathy can be a meaningful dimension of assessment even when the service provider is not human.

Emprical findings on the chatbots Empathy have been promising, although they also require some caution. In health-advice context, Liu and Sundar (2018) observed that expressions of sympathy elicited a positive response from users, users appeared more skeptical when they receive high levels of empathy. In particular, Pelau et al. (2021) found that perceived Empathy and interaction quality were predictors of acceptance of AI devices, and that perceived Empathy uniquely predicted acceptance for tasks involving emotional content. De Cicco, Silva, and Alparone (2020) found that both perceived warmth and competence were associated with increased Trust in retail chatbots, with the former having a stronger impact in low-stakes situations. Adam et al. (2021) found that politeness cues and social signals positively influenced compliance with chatbot requests, although the effect was limited by depending on the perceived authenticity of the cues.

This literature review reveals several tensions. The first concerns authenticity. When users perceive chatbot's Empathy as artificial and not genuine, empathetic framing may backfire and give rise to frustration rather than reassurance. The second one concerns calibration. Overly empathetic communication, especially when it is needed in a situation where the task must be accomplished efficiently, may be interpreted as patronizing or as an obstacle to solving the issue. The third is related to the distinction between cognitive and affective claims. If the statement explicitly assigns feelings to the chatbot, such as "It's so sad that you have a problem," users may feel uncomfortable when recognizing that this is an implausible statement (Liu & Sundar, 2018).

Another methodological tension pervades this literature. Other studies define Empathy as a subjective user perception, measured through adapted SERVQUAL or interpersonal Empathy scales (De Cicco et al., 2020; Wieseke et al., 2012). Some studies deduce Empathy from the behavior of the person, including compliance or completion of the tasks (Adam et al., 2021). In other cases, control Empathy at the system-prompt level and use the perceived warmth as a manipulation check (Liu & Sundar, 2018). These differences are important, since Empathy may be treated as a designed characteristic of the chatbot, a perceived attribute, and an outcome of the interaction. The present study follows the perceived-attribute approach and considers perceived Empathy as user-side construct, positioned between the experimentally manipulated communication style and downstream evaluations.

Overall, this literature suggests that an empathetic communication style is expected to result in a higher level of perceived Empathy than a neutral style, and that perceived Empathy will be positively related to favorable outcomes of the service encounter. The literature also supports that perceived Empathy may be treated as a manipulation check in communication style studies because of the construct is closely aligned with the communication style manipulation.

The final important question relates to the conceptual connection between Empathy and the related concept of warmth, which has been theorized in the Stereotype Content Model (Fiske et al., 2007). The Stereotype Content Model suggests that social perception is organized around the two fundamental dimensions, warmth and competence. These two dimensions evoke different emotions and behaviors. Chatbots studies sometimes use the terms Empathy and warmth interchangeably, and sometimes they use them separately (Pelau et al., 2021; De Cicco et al., 2020). The present study argues that perceived Empathy is a more specific concept than perceived warmth because it contains a component of recognition (rather than friendliness alone) and incorporates an acknowledgement of the user's particular situation. This narrower conceptualization consists of service marketing research (Parasuraman et al., 1988; Wieseke et al., 2012).

2.5 Emotional response and Justification of Self-Reported Sentiment

Emotional response refers to the user's affective reaction to the service encounter, during and immediately following. In the service marketing tradition, emotion has long been considered a determinant of Satisfaction, loyalty, and word-of-mouth (Bagozzi, Gopinath, & Nyer, 1999; Westbrook & Oliver, 1991). The Mehrabian and Russell (1974) framework, later developed by Russell (1980) and adapted to consumer research by Donovan and Rossiter (1982), views environmental and interactional cues as stimuli that evoke affective states which subsequently influence behavioral reactions.

Emotional response has been examined in the chatbots literature both as an outcome of design and as an intervening factor. Verhagen, van Nes, Feldberg, and van Dolen (2014) demonstrated that the social and personalization of virtual agents influences the emotional and evaluate aspects of interaction. Han, Yang, and Park (2022) showed that the relationship between perceived chatbot characteristics and behavior intentions is mediated by user emotions during the chatbot interaction. These results position emotional reaction as an intermediate link among design and users’ evaluation, showing how design features may impact downstream evaluations such as Satisfaction and Trust.

2.5.1 Two Traditions in the Measurement of Sentiment

Sentiment is used in two different traditions, which are sometimes confusing. The first tradition is based on computational linguistics and information retrieval and understands sentiment as the polarity of textual content. In this tradition, sentiment analysis refers to the automated process of determining the sentiment expressed in text, whether it is positive, negative, or neutral, using lexicon-based and syntactic methods, and more recently neural language models (Liu, 2012; Pang & Lee, 2008; Devlin, Chang, Lee, & Toutanova, 2019). Product reviews, social media posts and customer feedback have been processed at scale with sentiment analysis, providing operational insights that are valuable to companies (Hu, Liu, & Cheng, 2004; Pang & Lee, 2008).

The second tradition is based on psychology and consumer research, and views sentiment and related affective responses as personal experiences. In this tradition, affect is assessed using self-report measures such as the Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988), the Self-Assessment Manikin (Bradley & Lang, 1994) or context-specific Likert-type scales tailored for the service environment (Verhagen et al., 2014; Westbrook & Oliver, 1991). In the context of consumer and service research, this self-report tradition has contributed to the development of validated affective measures of the service encounter.

The two traditions address different research questions. When the amount of user-generated content is large, and the polarity of users' expressions to be analyzed are the focus, the automated tradition is appropriate for answering questions about what users have said in textual form. The self-report tradition is appropriate for questions of affective responses to a given interaction, especially when the focus is on users' experience, or on the relationships between affective responses to an interaction and evaluative outcomes downstream. The distinction between the two traditions is either or choice, but both traditions are complementary and the decision to use either should be based on the research question and not on methodological preference.

However, the absence of automated sentiment analysis in the present research creates some limitations.

For large collections of unprompted user text, where manual coding is impractical, automated sentiment analysis is well suited. Although, it is less suited for the research question of the present study for several reasons.

First, automated sentiment analysis classifies based on the sentiment that appears "on the surface" of the text, rather than the speaker's "inner" sentiment. In contrast, Mauss and Robinson (2009) argued that measures directly tapping the construct of interest are used to assess affective responses, not ones that are downstream of emotion. The relationship between what people say and how they feel is rarely straightforward, especially in customer service, where the language is brief, courteous, and functional.

Second, sentiment classifiers are prone to domain shift, idiom, sarcasm and negation and may not always perform well in service-encounter text (Hovy, 2020; Mohammad, 2016). Recent neural systems have improved these lexicon-based systems, but they are still trained on large, general-purpose corpora, which may not be applicable to dialogues with chatbots, and dialogue contexts are brief. An incorrect classification of the participants could introduce unknown measurement error that would weaken causal inference from the experimental design in the present study.

Third, the dialogues generated in the present study are brief and somewhat scripted. The chatbot plays a significant role in the total conversation flow and users' response is task-oriented and short. In such cases, the signals available to a sentiment classifier are few, and the variation in classifier output may not match the number of fluctuations in emotion experienced.

Fourth and most importantly, the research questions of the present study are not about the polarity of users' message, but about users' subjective experience of interaction. Perceived empathy, Satisfaction, Perceived service quality and Trust are subjective evaluations. Measuring Emotional response response through self-report is conceptually orientated as the subjective approach provides conceptual consistency between the constructs and also provides a same measurement basis for analysis of Emotional response to the other constructs.

Another issue that needs to be discussed is the link between sentiment and the boarder concept of emotion. In sentiment analysis literature, sentiment is usually considered as a one-dimensional polarity, where a text is more or less positive. In psychology research, however, emotions are categorized into discrete emotions (Ekman, 1992), and underlying dimensions of emotion, such as valence and arousal (Russell, 1980). The Emotional response dimension in the present study is dimensional and not discrete, as it aims to focus on the general valence of the affective response to the interaction with the chatbot. The dimensional approach is consistent with the way Emotional response is treated in service research (Verhagen et al., 2014; Bagozzi et al., 1999) and is appropiate for the analytical purpose of explaining affect in relation to evaluative outcomes.

The present study focused on operationalizing sentiment as a composite variable. Moreover, in the present study defines sentiment as a self-reported Emotional response, rather than as an automated classification of text, and measures a self-reported questionnaire after the interaction, administered after the interaction. This method is constistent with the consumer-affect studies (Bagozzi et al., 1999; Westbrook & Oliver, 1991), the previous research on social presence and affect in chatbot interactions (Verhagen et al., 2014), and the suggestion to measure affect by directly capturing participants' subjective experience of the state (Mauss & Robinson, 2009).

The term sentiment, or Emotional response, is used throughout the following empirical chapters to refer to the user's self-reported emotional reaction to the interaction with the chatbot. The construct was assessed using items informed by the tradition of the study of consumer affect, in relation to the reduction of frustration and/or anxiety and increase of calmness experienced during the interaction. The conceptualization assumes that the Emotional response is a proximal affective response that may act as a mediator between chatbot communication style and outcome such as Satisfaction and Trust.

There are some disadvantages to this method. Self-reported Emotional responses are parallel with social desirability bias, recall bias, and framing effects of survey items that are associated with self-report measures (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Yet, these limitations apply to all subjective evaluations of services and are well documented. In the present study, the following mitigations were used: anonymous survey administration, neutral wording, and previously validated measurement structure. This trade-off favors conceptually fit with the research questions over the amount of analysis breadth that could be achieved through automated methods.

2.6 User Satisfaction

Customer Satisfaction has been one of the central concepts in service research for over 40 years. According to Oliver (1980), expectancy-disconfirmation paradigm is a theory that views Satisfaction as a consequence of comparing the expected and perceived performance. If the performance exceeds expectations, customers' Satisfaction may be high, and if it is below expectations, Satisfaction will be low. The framework has since been extended to incorporate the role of affect in the presence of cognition (Oliver, 1993; Westbrook & Oliver, 1991) and to examine how Satisfaction from encounters as extending to relationships (Bolton & Lemon, 1999).

In chatbot research, Satisfaction has been operationalized as the overall evaluation of the chatbot interaction, measured through adapted Likert-scale items in the context of the chatbot (Ashfaq, Yun, Yu, & Loureiro, 2020; Chung, Ko, Joung, & Kim, 2020). According to Ashfaq et al. (2020), perceived usefulness and ease of use influenced users' Satisfaction level of conversational AI assistants. Chung et al. (2020) identified interaction quality, problem-solving ability and security to be significant for determining chatbot Satisfaction in a luxury retail environment. Hsu and Lin (2023) demonstrated that perceived intelligence and anthropomorphism enhance Satisfaction in banking chatbots.

An important implication of this body of work is the conceptual connection between Satisfaction and related variables. There are several perspectives on Satisfaction, as it is often seen as an outcome of service quality, as a predictor of Trust and loyalty and as a mediator between the service evaluation and behavioral outcomes (Cronin, Brady, & Hult, 2000; Dabholkar, Shepherd, & Thorpe, 2000). The ambiguity of these positions can pose a risk of conceptual overlapping when multiple similar constructs are used as predictors in the same regression analysis. As outlined in the methodology chapter, the regression literature suggests that there may be a risk of suppression effects and multicollinearity that needs to be addressed explicitly and statistically, rather than overlooked (Cohen, Cohen, West, & Aiken, 2003; Pandey & Elliott, 2010).

A second important issue concerns the cultural and contextual specificity of Satisfaction. Most of the studies on chatbot Satisfaction have been conducted in a specific national context, which may limit the generalizability of their results to other settings, e.g., Southern Europe, where service-encounter norms and expectations differ. The present study adds geographical and cultural variety to this body of evidence sources, as it was conducted on a sample of Greek-speaking participants.

Another important issue concerns the time used to measure Satisfaction. Previous chatbots studies often measure Satisfaction following the first interaction, which reflects transaction-specific Satisfaction, but not the overall Satisfaction that can develop over time from multiple interactions (Anderson, Fornell, & Lehmann, 1994; Bolton & Lemon, 1999). This is a significant constraint, since cumulative Satisfaction may be a better predictor of loyalty and continued usage than Satisfaction. The present study recognizes this limitation and only measures Satisfaction as transaction specific and not as an indicator of relationship-level Satisfaction. Even within this one encounter framing, this mediation analysis aims to explain the affective and evaluative mediation between the perceived chatbot Empathy and Satisfaction.

2.7 Perceived service quality

Perceived service quality is the evaluative judgement of the excellence or superiority of a service by a customer (Parasuraman et al., 1988; Zeithaml, 1988). According to Parasuraman et al.'s (1988) SERVQUAL model, the five dimensions of service quality are Tangibles, Reliability, Responsiveness, Assurance, and Empathy. The model has been widely utilized, but has also been criticized conceptually and psychometrically, especially because it measures Perceived service quality through the gap of expectations and perceptions (Cronin & Taylor, 1992).

In digital and self-service settings, service quality has been re-examined, for instance through the E-S-QUAL approach, which adapts service quality to an electronic environment (Parasuraman, Zeithaml, & Malhotra, 2005). For chatbot interactions, several studies have proposed and tested adapted instruments. Trivedi (2019) reported that information quality, system quality, and service quality are the three factors contributing to brand love through the users' Satisfaction with chatbots. Chen, Le, and Florence (2021) explored factors influencing customer engagement towards banking chatbots, while the Chatbot Usability Questionnaire, created by Borsci et al., 2022 captures multiple dimensions related to chatbot quality.

One important concern in the present study is the conceptual similarity of Perceived service quality and its related constructs. Empirical research consistently finds correlations between service quality and Satisfaction ranging from .70 to .85, thus questioning the possibility of empirically separating the two constructs within a given sample (Cronin et al., 2000; Dabholkar et al., 2000). Most of the literature treats the two constructs as conceptually different, with service quality being a more general assessment of the firm's offering and Satisfaction being a more transaction-specific assessment. This distinction, however, becomes more challenging to hold empirically in brief, a single-episode chatbot interactions. This problem is recognized in the present study, and a concise functional-evaluative measure of Perceived service quality is used. The analytical approach to conceptual overlap is designed to reveal hiding patterns of conceptual overlap rather than obscure them, especially through the use of hierarchical regression.

Another important aspect is the relationship between Perceived service quality and empathetic aspects of the chatbot interaction. In the context of service quality, Empathy is treated as one dimension of service quality, while, in chatbot research, Empathy has sometimes been treated as a separate construct, as the Empathy displayed by the non-human agent differs from the Empathy demonstrated by a human agent in both several theoretical and operational aspects (Liu & Sundar, 2018; Pelau et al., 2021). Perceived empathy, in the present study, is considered as a separate construct, which is associated with, but not subsumed by, the Perceived service quality construct.

2.8 Trust in AI Customer Service

Trust is a widely researched concept in management and information systems research. According to Mayer, Davis, and Schoorman (1995) define Trust as a willingness of one party to base on the actions of the other party will act without the ability to monitor and control that other party will act appropriately without control and direct monitoring. Three antecedents of Trustworthiness have been identified in the model: ability, benevolence and integrity. The model was developed in an organizational context but has also been applied to technology and AI.

McKnight, Carter, Thatcher and Clay (2011) extend Trust theory to specific technologies, which distinguish between Trust in technology, Trust in human actors who develop and use that technology, and Trust in the institutions that regulate that technology. Their distinction is useful for the chatbot context where users may simultaneously Trust or disTrust the chatbot, its operator, and the institutional context of digital services. More recent research has extended by examining antecedents of Trust in algorithmic systems, such as transparency, accountability, and fairness (Glikson & Woolley, 2020; Shin, 2021).

The picture of Trust in chatbot service from empirical research is complex. Følstad et al. (2018) identified factors that associated with Trust in customer service chatbots, such as perceived expertise, responsiveness, and human-likeness. Nordheim, Følstad and Bjørkli (2019) found that the initial Trust formation was primarily influenced by brand-level Trust and individual disposition to Trust technology. In the wider context of human-automation interactions, Lee and See (2004) argue that Trust in automation develops through experience and is affected by the perceived competence of the automation and consistency of the automation's behavior. Their analysis indicates that Trust with chatbots is a dynamic, not static, Trust assessment, which may differ across interactions.

A key point in this literature is the link between empathetic framing and Trust. Empathetic communication may also foster Trust by sending the message that the service provider cares about the customer's situation (Følstad et al., 2018; Pelau et al., 2021). However, the use of empathetic language in non-human agents can be misunderstood as insincere, particularly if users interpret the empathetic language as a strategic cue rather than genuine expression of concern, which may reduce Trust in that AI agent (Mozafari et al., 2022). The evidence so far suggests the first position in low stakes settings, and some more mixed results in high stakes settings such as financial services and health. In the context of a non-trivial personal stake, the present study (adopting a banking context) contributes to this debate by exploring Trust as a result of an empathetic versus professional chatbot interaction.

Another major challenge concerns the measurement of Trust. Some studies operationalize Trust as a unidimensional construct of Trust, which is the inclination to Trust the system (McKnight et al., 2011). Others distinguish between Trust in ability, benevolence and integrity (Mayer et al., 1995). Others, however, approach a distinction between cognitive and emotional Trust (McAllister, 1995). Each approach offers different analytical levers and should be selected according to the research questions of the study. The present study focuses on users' willingness to rely on the chatbot in the future to handle similar issues in the wider service relationship. It therefore uses a unidimensional measure of Trust at the service relationship level.

Another major issue concerns the difference between the initial Trust and Trust that develops through repeated interaction. McKnight, Cummings, and Chervany (1998) distinguish between initial Trust, which forms before substantive interaction, and knowledge-based Trust, which develop as a function of users' experience with the Trustee. Most of the experiments conducted in the field of chatbots are conducted in a very brief interaction and therefore capture a state of Trust lies somewhere between the initial Trust and the knowledge based Trust (Nordheim et al., 2019). This approach is suitable for research that seeks to understand early user impressions of design elements, but it does not allow findings to be generalized to long-term chatbot use. The present study follows this experimental tradition and treats Trust as a post interaction assessment after a service encounter that has a clear beginning and end.

2.9 Synthesis, Research Gaps, and Conceptual Framework

Several general conclusions of the literature reviewed in this chapter. First, chatbot design and the communication style affect how users perceive and evaluate the service encounter; however, the magnitude and direction of these effects on the service encounter may vary depending on contextual and methodological factors. Second, perceived Empathy is a legitimate construct in human-chatbot interaction, but its effects depend on authentic, calibrated and task-fit. Third, Satisfaction , Perceived service quality , and Trust are central service-evaluative outcomes of the service encounter and are empirically related in a manner which require careful analytical treatment. Fourth, sentiment in the present study is conceptualised as a self-reported Emotional response, as measured by a post-interaction questionnaire, and it is not the same as the automated approach of sentiment analysis.

The present study is motivated by several gaps in literature. First, while the communication style has been explored through various chatbot studies, relatively few studies have used a constant underlying language model, conversational flow, and survey instrument while varying only communication style. The present study uses this cleaner design to isolate the effect of communication style. Second, Emotional response has been theorised as a mediating construct between chatbot design and downstream evaluations such as Satisfaction and Trust, but formal mediation analyses of Emotional response as a link between communication style and Satisfaction and Trust have been less common in the published literature. The present study attempts to fill this gap by conducting mediation using Hayes' PROCESS macro. Third, chatbot research is mainly focused on samples from specific nations. The European sample included in a study was conducted by European speakers of Greek, enhancing the geographical and cultural diversity of the evidence base. Finally, banking is a sparse area in the experimental chatbot literature and is an area where customer interactions have high importance. This study aims to bridge this gap, presenting the experimental task in a banking context.

3. Methodology

3.1 Introduction

This chapter outlines the research methodology used to address the aims and research questions of this study. The study investigates the communication style of chatbot of AI customer service on its effects on the users' perceived empathy, Emotional response, Satisfaction, Perceived service quality and Trust in the service provider. To answer the above questions, a between-subjects experimental design was used. Two functional chatbots were created and participants randomly assigned to communicate with one of them. One of the chatbots used an empathetic, human-like communication style while the other used a neutral, professional communication style. Participants interacted in real-time with one of the chatbots and then answered a structured questionnaire.

The chapter is divided into the following sections. The research philosophy and research design is described in Section 3.2. The study scenario is described in Section 3.3 while the recruitment procedure and sample is described in Section 3.4. Data collection is described in Section 3.5, chatbot development and experimental manipulation in Section 3.6. Section 3.7, presents the measurement instruments, with the reasons for using the self-reported sentiment and not automated sentiment. The data analysis strategy is discussed in Section 3.8, ethical issues are discussed in Section 3.9, and a chapter summary is provided in Section 3.10.

3.2 Research design

In this study, the quantitative approach method was adopted. This approach is appropriate when the researcher seeks to test the relationships between measured variables which can be analyzed statistically (Bryman, 2016; Creswell & Creswell, 2018). In this study, the key constructs of interest – Perceived empathy, Emotional response, Satisfaction, Perceived service quality & Trust – have been measured on Likert-type scales. Thus, a quantitative approach is suitable because it allows comparison with the results of previous research on user interaction with chatbots, and consumer Trust (Følstad & Brandtzaeg, 2020; Pelau, Dabija, & Ene, 2021).

A between-subjects experimental design was used in this study. Experimental designs are especially appropriate for research questions that focus on causal influence because they help control individual differences between participants (Field & Hole, 2003; Shadish, Cook, & Campbell, 2002). The independent variable, in the present study, was the style of chatbot communication which had two types: empathetic (human-like) and professional (neutral). The dependent variables were Perceived empathy, Emotional response, Satisfaction, Perceived service quality, and Trust. The design allowed the effect of communication style to be isolated by keeping the main features of the chatbot and differ only the tone of communication.

The research questions, the available resources and the use of single well-defined service interaction made a cross-sectional design for this study. Although a longitudinal or mixed method design could have provided additional depth, this selected design was determined the most suitable for answering these research questions of this study. This choice is also consistent with prior chatbot research, where designs have frequently been used to examine the impact of design properties on user perceptions (Adam, Wessel, & Benlian, 2021; Liu & Sundar, 2018).

3.3 Study Scenario

A scenario-based task was created to stimulate an interaction with an AI customer service agent. The use of scenario-based experiments in service research is common as the researcher can standardize the stimulus across participants while maintaining the level of realism (Bitner, 1990; Wirtz & Lwin, 2009). In the present research, participants were requested to imagine that they had lost their bank card and they contact their bank via an AI chat to block the bank card immediately. All participants were presented with the same scenario and the same questionnaire. The only difference between the two groups was the communication style of the chatbot.

A banking environment was specifically selected for the study. Financial services are associated with higher risk and uncertainty and customers may concern when they encounter potential monetary loss (Yoon & Steege, 2013). In these environments, perceptions of Empathy and responsiveness of the service provider become significantly important (Parasuraman, Zeithaml, & Berry, 1988). The banking environment is thus considered appropriate for examining Empathy, Emotional response, Satisfaction, Perceived service quality , and Trust since they are all important in the context of uncertainty and personal risk. Furthermore, AI chatbots are already widely used in the banking sector (Hu, Lu, & Gong, 2021), strengthens the ecological validity of the scenario.

3.4 Sample and Participants

A total of 107 adults participated voluntarily in the study. Convenience sampling was used to recruit the participants, including friends, colleagues, and fellow students. Convenience sampling is commonly used in experimental studies in marketing and human-computer interaction, particularly to test causal relationships rather than to make population estimates (Bryman, 2016; Calder, Phillips, & Tybout, 1981). Random assignment to conditions is useful for increasing internal validity even if recruiting is non-probabilistic within an experimental framework.

Of the 107 participants, 56 interacted with the empathetic (human-like) chatbot and 51 with the professional (neutral) chatbot. The sample size is similar to that used in earlier experimental research on the chatbot communication style (Adam et al., 2021; Pelau et al., 2021). The majority of participants identified as female (67.3%), while 32.7% identified as male. In terms of age, 41.1% of participants were aged 25 to 34 years, 35.5% were aged 35 to 44 years, 13.1% were between 45 to 54, 8.4% were aged 18 to 24 years, and 1.9% were aged 55 or older. Most participants have had prior experience with chatbots, with 85.0% indicating previous interaction and 15.0% reporting no prior experiencr. The percentage of users who have already used a chatbot is high, which is consistent with the increasing adoption of conversational agents in Europe (Følstad & Brandtzaeg, 2020).

3.5 Data Collection Procedure

Data were collected online. Participants were provided with written description of the scenario, accompanied by a link with the one of the two chatbot types. Moreover, prospective participants were randomly assigned to each scenario and then they were asked to complete the questionnaire.

Participants interacted with the chatbot in a text-based conversation in real time after reading the brief. They were instructed to pretend they were bank customers and go through the scenario to the end – including blocking the lost card. The questionnaire was then presented as a chatbot message with a link to a survey on Google Forms, following the end of the interaction with the chatbot. The data from the surveys was saved in the Google Forms interface and then exported to a spreadsheet for analysis. To ensure that all responses to the questionnaire referred to an episode of service, only participants who had previously interacted with the chatbot were able to access it.

3.6 Chatbot Development and Experimental Manipulation

Both the chatbots were designed with the Dify platform, with the Large Language Model being GPT-4o mini for both conditions. GPT-4o mini was chosen due to its being a quick and affordable model for specific tasks, as described by its developer, OpenAI (2024). This model was a suitable trade-off between the quality of interaction, the latency and the cost of the experiment, because the interaction was short, structured and text based. By keeping the underlying language model in both chatbot versions, the same, any differences between the two groups could be more likely to reflect the differences in the language model itself.

The conversation flow was the same for both versions of the chatbot. Participants in all conditions indicated that they had lost their bank card and requested that it be blocked. The task, sequence of conversational moves, conversational language, level of professionalism and the post-interaction survey remained the same for all conditions. The only aspect of the chatbots that was different was the manner of communication, which was the experimental manipulation.

The empathetic (human-like) chatbot was trained to keep a warm, supportive and reassuring tone, and to use language that was considerate of the user and showed understanding of their situation. The professional (neutral) chatbot was programmed to respond using a task-oriented approach, meaning that it has to respond in an efficient and factual manner without emotion. This contrast is similar to earlier research into conversational agents in their communication (Adam et al., 2021; Liu and Sundar, 2018), which explores the difference between social and task-oriented communication agents. For the analysis, a manipulation check, based on perceived Empathy was added, and is reported in Chapter 4. The complete set of prompts for both types of chatbot are attached in Appendix A.

3.7 Measurement Instruments

Data were collected through structured questionnaire immediately after the interaction with the chatbot. The questionnaire included five constructs of the study: Perceived empathy, Emotional response, Satisfaction, Perceived service quality and Trust, and a few demographic and background variables. To align with the typical methodology used in the studies of service and chatbots (Pelau et al., 2021; Sundar, 2008), all core items were measured on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). The measurement items were adapted from existing scales in Empathy and Service quality and Trust literatures and reworded for the framework of the chatbot. The full questionnaire is provided in Appendix B.

3.7.1 Operationalization of Sentiment as Self-Reported Emotional response

One of the methodological decisions made in the current study is how to define sentiment. Sentiment is generally assessed using automated sentiment analysis in computational linguistics and marketing analytics, which involves using computational algorithms that are based on the lexical and syntactic characteristics of the text (Liu, 2012; Pang & Lee, 2008). Unsolicited textual data, such as product reviews or social media posts is a good application of automated sentiment analysis.

However, the present study does not rely on automated sentiment analysis, it is rather a self-reported Emotional response that is measured using a post-interaction questionnaire. There are three reasons for this decision. The first research questions are related to the subjective emotional experience of the user while interacting with a chatbot, and this is a very different phenomenon from the valence of the text that the chatbot is able to generate. Self-report instruments directly measure the emotional state experienced, while automated sentiment analysis only measures the emotional state from the textual cues (Mauss & Robinson, 2009). Second, the chats in this study were brief, and consisted of much of the user's text was pre-formatted, which reduces the amount of free-form text that can be automatically analysed. Third, self-reports scales on Emotional response and related constructs have been validated in the literature of chatbots and the service research literature and have proven to have good psychometric properties (Pelau et al., 2021; Verhagen, van Nes, Feldberg, & van Dolen, 2014).

For this reason, this dissertation provisions the term sentiment for the general Emotional response users has towards the Chatbot when interacting with the Chatbot and after interacting with the Chatbot, using the structured self-report items. This emotionally related self-reported sentiment is operationalized in the analyses by the construct of Emotional response described below.

3.7.2 Perceived Empathy

Perceived Empathy was assessed with four items representing how much the chatbot was perceived to understand, support, and be emotionally aware. The chatbot was rated on its ability to be perceived as understanding the users' situation, recognizing the user's concern, being supportive and making the user feel understood. Items were drawn from previous studies on the perceived Empathy in service and conversational agent (Pelau et al., 2021; Wieseke, Geigenmuller, & Kraus, 2012). This shows higher scores indicated greater perceptions of chatbots empathy.

3.7.3 Emotional response

The Emotional response was measured using two items that tapped the affective impact of the interaction for the participant. The examined items explored the reduction of feelings of frustration and anxiety and the subject's sense of calmness. The construct is based on the consumer emotion in service encounters literature, where affect is defined as the valence (positive or negative) and the intensity (strength or weakness) of emotions that consumers experience during a service encounter (Mehrabian & Russell, 1974; Verhagen et al., 2014). High scores indicate a more positive Emotional response. This construct is also used to operationalize self-reported sentiment in the present study as discussed in Section 3.7.1.

3.7.4 Satisfaction

Satisfaction was assessed using three items that encompassed the overall assessment of the conversation with the chatbot. The questions used to evaluate whether participants were happy with the interaction, if they thought that the chatbot solved the problem well, and if the interaction brought their expectations to a satisfactory level. This conceptualization is based on the expectancy-disconfirmation approach in service Satisfaction studies, which defines Satisfaction as a post-consumption assessment of the performance compared with prior expectations (Oliver, 1980). The higher scores the better Satisfaction with the interaction with the chatbot.

3.7.5 Perceived service quality

The quality of service perceived was assessed using three items related to the effectiveness, reliability, and overall quality of the chatbot service. The items were drawn from the wider service quality literature where Perceived service quality is defined as a multi-dimensional assessment of service quality's ability to satisfy the needs of customers (Parasuraman et al., 1988). Due to the time limitations of an experimental questionnaire, the present study will primarily concentrate on a simple functional evaluative measure as opposed to the entire SERVQUAL battery. A higher score signifies better service quality perspectives.

3.7.6 Trust in the Service Provider

The level of Trust was obtained by three questions which asked participants how much they would Trust this kind of chatbot service in the future. The items evaluated to determine if the participants Trusted this type of chatbot for similar issues, if they would use it again when dealing with service, and if they were comfortable relying on this type of digital service. The conceptualization is based on existing accounts of Trust as a desire to be vulnerable to the actions of another party in reliance on positive expectations (Mayer, Davis, & Schoorman, 1995). Recent studies have demonstrated that Trust is a key outcome in AI-powered customer service, where users are increasingly dependent on automated systems to handle sensitive interactions (McKnight, Carter, Thatcher, & Clay, 2011; Wang, Lu, & Tan, 2024). Higher scores represent greater levels of Trust.

3.7.7 Demographic and Background Variables

There were also some demographic and background questions included in the questionnaire, such as gender, age group, previous experience with chatbots, and confirmation of the successful completion of the interaction with the chatbot. The sample was described with these variables, and they were also used to put the analysis into perspective.

3.7.8 Composite Measures

Composite scores for each construct were created by averaging the responses for the items that were part of that construct to complete the statistical analyses. The result of this procedure was an overall score for Perceived Empathy, Emotional response, Satisfaction, Service Quality and Trust. When the items in a construct show acceptable internal consistency, as indicated by reliability analysis (as reported in Chapter 4), the use of mean composite scores is appropriate (DeVellis, 2017; Nunnally & Bernstein, 1994).

3.8 Data Analysis Strategy

All statistical analyses were carried out in IBM SPSS Statistics. The analyses were conducted in several steps in order to investigate the research questions.

First, descriptive statistics were calculated for the major study variables such as means, standard deviations, minima and maxima as the first step. Second, internal consistency of each scale was evaluated using Cronbach's alpha. As with the social and behavioral sciences, the values of $\alpha \geq .70$ were considered acceptable reliability and $\alpha \geq .80$ were considered good reliability (Nunnally & Bernstein, 1994; Taber, 2018).

Third, a manipulation check was conducted to validate that chatbot conditions were felt to be distinct in terms of empathy. Fourth, independent-samples t-tests were conducted to address each of the main outcome variables in the empathetic and professional chatbot conditions. It directly answers the research questions regarding the impact of communication style on Emotional response, level of Satisfaction, dimensions of Perceived service quality, and the level of Trust. If the test performed by the Levene indicated a homogeneity of variance, the data for unequal variances were interpreted (Field, 2018).

Fifth, Pearson correlation analysis was used to explore the bivariate relationships between key variables of the study. The interpretation of subsequent regression analyses is complemented by the correlation matrix, and it gives a preliminary indication of the pattern of the associations among the constructs. Sixth, multiple regression analysis was conducted in which Trust is the dependent variable. Perceived empathy, Emotional response, Satisfaction , and Perceived service quality were used as predictors. Multicollinearity was checked with the correlation matrix and the tolerance values (Hair, Black, Babin, & Anderson, 2019) as well as the Variance Inflation Factor (VIF) (Field, 2018), where VIF values lower than 10 and tolerance values higher than .10 were considered acceptable.

As a robustness check, a hierarchical regression analysis was carried out in the seventh step. Each predictor was entered into 2 blocks: to test the changes in the regression coefficients after the entry of Emotional response and Perceived service quality after perceived Empathy and Satisfaction. This analysis also helps with the understanding of patterns of suppression and multicollinearity.

Last but not least, mediation analysis was performed using the PROCESS macro for SPSS (Model 4) with 5,000 bootstrap samples and bias-corrected 95% confidence intervals (Hayes, 2022). The two indirect effects were analyzed viz. indirect effect of perceived Empathy on Trust through Emotional response, and the indirect effect of perceived Empathy on Satisfaction through Emotional response.

The statistical assumptions were explored prior to performing the inferential analyses. Normality was evaluated by Kolmogorov-Smirnov and Shapiro-Wilk tests and by visually checking the Q-Q plots. Levene's test was used to check for homogeneity of variance. Tolerance and VIF values were used to check the multicollinearity in the models of regression. The value of $\alpha = .05$ was used for all the analyses.

3.9 Ethical Considerations

Ethical issues were considered throughout the whole research process following existing guidelines for research with human subjects (British Psychological Society, 2021). This study was completed by volunteers only. All participants were provided with written information on the purpose of the study, the nature of the interaction with the chatbot, the approximate duration of the task and that responses to the chatbot would be utilized only for academic purposes.

Personally, identifying information was not collected. The surveys were anonymous and participants were not identified by the results of the surveys. The participants could exit the study any time before closing the browser window to complete the questionnaire, after which no data would be kept. Data were securely stored and only accessible to the researcher and supervisor.

The study simulated banking, and not an actual financial transaction, which is why no financial or personal information was exchanged. So the risk level of the study was low. However, significant efforts were made all along to keep participants feeling respected and to make the study material clear, transparent and avoid elements beyond its disclosed use of a simulated scenario.

3.10 Summary

In this chapter, the methodology used in the present study was outlined. The study used a between-subjects experimental design with random assignment of participants to one of two chatbots where the only variable was how the chatbot communicated. The study was set in an artificial bank situation, and data were gathered by using a structured online questionnaire. The chapter described how sentiment was operationalized as an Emotional response that was measured through validated scale items and not automated text analysis, and how the perceived empathy, Satisfaction, Perceived service quality and Trust were measured. Descriptive data analysis, reliability analysis, manipulation check, t-tests, correlation analysis, multiple regression, hierarchical regression, and mediation analysis were used to analyze the data and were planned. Throughout: ethical safeguards applied. The results of this analytical approach will be discussed in the following chapter.

4. Results

4.1 Introduction

This chapter presents the findings from the statistical analyses conducted to answer the research questions of the study. All analyses were conducted using IBM SPSS Statistics with the PROCESS macro used for the mediation analyses. The chapter follows the analytical structure outlined in Chapter 3. The sample characteristics are reported in section 4.2. The descriptive statistics of the main study variables are presented in Section 4.3. The reliability analysis of the measurement scales is reported in Section 4.4. Assumption testing is reported in Section 4.5. The manipulation check for the experimental conditions is presented in section 4.6. Independent-samples t-tests used to answer the research questions are reported in Section 4.7. The correlation analysis is reported in Section 4.8, while the multiple regression analysis predicting Trust is reported in Section 4.9, with particular attention to suppression effects and the pattern of multicollinearity present in the model. A hierarchical regression analysis is then reported in section 4.10 to clarify this pattern. The two mediation analyses for the affective pathway linking perceived Empathy with Trust and Satisfaction are presented in Section 4.11. The main findings are summarized in Section 4.12.

4.2 Sample description

Descriptive statistics were calculated to show the sample characteristics. Five composite variables were created, and descriptive statistics were calculated for each. The means were relatively high, reflecting the fact that a majority of participants had positive experience of the interaction with the chatbot. Perceived Empathy had the highest mean of 4.30 (SD = 0.81). Satisfaction and Perceived service quality have similar mean scores (M = 4.26, SD = 0.83; M = 4.28, SD = 0.78), while Emotional response was slightly lower (M = 4.04, SD = 0.98). Trust had the lowest mean, and the highest standard deviation (SD) among the five variables (M = 3.97, SD = 1.00), which may indicate greater variation in participants' responses. The descriptive statistics are shown in Table 4.1.

Variable	N	Minimum	Maximum	Mean	SD
Empathy	107	1.50	5.00	4.30	0.81
Emotional response	107	1.00	5.00	4.04	0.98
Satisfaction	107	1.67	5.00	4.26	0.83
Perceived service quality	107	2.33	5.00	4.28	0.78
Trust	107	1.00	5.00	3.97	1.00

Table 4.1 Descriptive Statistics of the Main Variables

4.3 Reliability Analysis

The internal consistency of each multi-item scale was assessed using Cronbach's alpha. All scales exceed the commonly accepted threshold of $\alpha = .70$ and most also exceeded the a=.80, indicating excellent internal consistency. Perceived empathy, Emotional response, Satisfaction, Perceived service quality and Trust had excellent reliability ($\alpha = .914$, .908, .920, .891 and .931, respectively). These results support the use of composite mean scores for further analysis.

4.4 Assumption testing

The statistical assumptions needed for the inferential analysis were assessed before. Visual inspection of Q-Q plots, Kolmogorov-Smirnov and Shapiro-Wilk tests were used to assess normality. Both tests showed deviations from normality for all five main variables were statistically significant ($p < .001$). Given the number of participants ($N = 107$), however, and the proven strength of the parametric tests for moderate violations of normality, planned analyses were considered appropriate. The Q-Q plots revealed there were no serious deviations from normality.

Each independent samples t-test was preceded by Levene's test for homogeneity of variance. When Levene's test was significant, the unequal-variances version of the t-test was interpreted. Moreover, the regression analysis was assessed for multicollinearity using tolerance values, the Variance Inflation Factor (VIF) and collinearity diagnostics. The results that are presented in Section 4.9 indicate substantial shared variance among the predictors. Although the VIF values did not exceed the threshold of 10, some variables approached the threshold of 5, highlighting that regression coefficients should be interpreted carefully. Thus, a hierarchical regression analysis is presented in Section 4.10 to further investigate multicollinearity and suppression pattern detected in the multiple regression model.

4.5 Manipulation check

To ensure that the two chatbot conditions were perceived as different in terms of empathy, a manipulation check was carried out. Perceived Empathy was compared between the two conditions using independent-samples t-test. Levene's test was significant and the unequal-variances version of the test was interpreted. The mean perceived Empathy for participants who interacted with the empathetic chatbot ($M = 4.67$, $SD = 0.50$) was significantly higher than the mean perceived Empathy of the participants who interacted with the professional chatbot one ($M = 3.90$, $SD = 0.89$), $t(77.19) = 5.51$, $p < .001$. Thus, manipulation was considered successful. Table 4.2 presents descriptive statistics by condition for all key variables.

Variable	Condition	N	M	SD	SE
Empathy	Empathetic chatbot	56	4.67	0.50	0.07
	Professional chatbot	51	3.90	0.89	0.12
Emotional response	Empathetic chatbot	56	4.33	0.82	0.11
	Professional chatbot	51	3.72	1.05	0.15
Satisfaction	Empathetic chatbot	56	4.54	0.64	0.09
	Professional chatbot	51	3.95	0.91	0.13
Perceived service quality	Empathetic chatbot	56	4.58	0.58	0.08
	Professional chatbot	51	3.95	0.83	0.12
Trust	Empathetic chatbot	56	4.24	0.88	0.12
	Professional chatbot	51	3.67	1.03	0.14

Table 4.2 Descriptive Statistics by Chatbot Condition

4.6 Research Question Testing

Independent-samples t-tests were conducted to compare the empathetic and professional chatbot conditions on each of the main outcome variables. These analyses addressed the four research questions of the study. The results are reported in Table 4.3.

Research Question 1 asked whether users who interacted with an empathetic chatbot reported more positive Emotional responses than users who interact with a professional chatbot. Levene's test for Emotional response was not significant, so the equal-variances version of the t-test was interpreted. Participants in the empathetic chatbot condition reported significantly higher Emotional response scores ($M = 4.33$, $SD = 0.82$) than those in the professional chatbot condition ($M = 3.72$, $SD = 1.05$), $t(105) = 3.39$, $p < .001$. Research Question 1 was therefore supported.

Research Question 2 asked whether users who interacted with an empathetic chatbot reported higher Satisfaction than users who interacted with a professional chatbot. Levene's test was significant, so the unequal-variances version of the t-test was interpreted. Participants in the empathetic chatbot condition reported significantly higher Satisfaction ($M = 4.54$, $SD = 0.64$) than those in the professional chatbot condition ($M = 3.95$, $SD = 0.91$), $t(88.63) = 3.84$, $p < .001$. Research Question 2 was therefore supported.

Research Question 3 asked whether users who interacted with an empathetic chatbot reported higher Perceived service quality than users who interacted with a professional chatbot. Levene's test was significant, so the unequal-variances version of the t-test was interpreted. Participants in the empathetic chatbot condition reported significantly higher Perceived service quality ($M = 4.58$, $SD = 0.58$) than those in the professional chatbot condition ($M = 3.95$, $SD = 0.83$), $t(88.21) = 4.54$, $p < .001$. Research Question 3 was therefore supported.

Research Question 4 asked whether users who interacted with an empathetic chatbot reported higher Trust than users who interacted with a professional chatbot. Levene's test for Trust was not significant, so the equal-variances version of the t-test was interpreted. Participants in the empathetic chatbot condition reported significantly higher Trust ($M = 4.24$, $SD = 0.88$) than those in the professional chatbot condition ($M = 3.67$, $SD = 1.03$), $t(105) = 3.11$, $p = .002$. Research Question 4 was therefore supported.

Variable	Levene's p	Variance	t	df	p	Mean diff.	95% CI
Empathy	< .001	Not assumed	5.51	77.19	< .001	.78	[.50, 1.06]
Emotional response	.092	Assumed	3.39	105	< .001	.61	[.25, .97]
Satisfaction	.010	Not assumed	3.84	88.63	< .001	.59	[.28, .89]
Perceived service quality	.003	Not assumed	4.54	88.21	< .001	.64	[.36, .91]
Trust	.281	Assumed	3.11	105	.002	.58	[.21, .95]

Table 4.3 Independent-Samples t-Tests Comparing the Empathetic and Professional Chatbot Conditions

4.7 Correlation Analysis

Pearson correlation coefficients were computed to examine the bivariate relationships among the five main study variables. All correlations were positive and statistically significant at $p < .001$. Perceived Empathy was strongly correlated with Emotional response ($r = .747$), Satisfaction ($r = .751$), Perceived service quality ($r = .748$), and Trust ($r = .633$). Emotional response was strongly correlated with Satisfaction ($r = .821$), Perceived service quality ($r = .783$), and Trust ($r = .807$). Satisfaction was strongly correlated with Perceived service quality ($r = .867$) and Trust ($r = .716$), and Perceived service quality was strongly correlated with Trust ($r = .792$). The pattern indicates that higher levels of each of the main variables were associated with higher Trust. At the same time, the magnitude of the inter-correlations among the predictors, particularly those involving Satisfaction and Perceived service quality, suggests substantial shared variance, which has implications for the multiple regression analysis reported in Section 4.9. The full correlation matrix is presented in Table 4.4.

Variable	1	2	3	4	5
1. Empathy	—				
2. Emotional response	.747**	—			
3. Satisfaction	.751**	.821**	—		
4. Perceived service quality	.748**	.783**	.867**	—	
5. Trust	.633**	.807**	.716**	.792**	—

Table 4.4 Pearson Correlations among the Main Study Variables

4.8 Multiple regression analysis

A multiple regression analysis was conducted to examine the joint predictive contribution of the fourstudy conducted as predictors to Trust. Trust was entered as the dependent variable, while perceived empathy, Emotional response, Satisfaction, and Perceived service quality were entered as predictors. The regression equation is specified below:

$$Trust = \beta_0 + \beta_1(Empathy) + \beta_2(Emotional\ Response) + \beta_3(Satisfaction) + \beta_4(Perceived\ Service\ Quality) + \varepsilon$$

The overall model was statistically significant, $F(4, 102) = 67.90$, $p < .001$, and accounted for 72.7% of the variance in Trust ($R^2 = .727$). The ANOVA table for the regression model is presented in Table 4.5, and the regression coefficients are presented in Table 4.6.

Source	SS	df	MS	F	p
Regression	76.66	4	19.17	67.90	< .001
Residual	28.79	102	0.28		
Total	105.45	106			

Table 4.5 Analysis of Variance for the Regression Model Predicting Trust

Predictor	B	SE B	β	t	p	Tolerance	VIF
Constant	-0.10	0.31	—	-0.32	.748	—	—
Empathy	-0.10	0.11	-.08	-0.91	.364	.36	2.75
Emotional response	0.59	0.10	.58	5.95	< .001	.28	3.59
Satisfaction	-0.22	0.14	-.19	-1.56	.121	.19	5.25
Perceived service quality	0.71	0.14	.55	5.02	< .001	.22	4.54

Table 4.6 Regression Coefficients Predicting Trust

Two predictors emerged as statistically significant positive predictors of Trust when all four variables were entered simultaneously. Emotional response was the strongest predictor, $\beta = .58$, $p < .001$, and Perceived service quality was also a significant positive predictor, $\beta = .55$, $p < .001$. In contrast, the standardised coefficients for perceived empathy, $\beta = -.08$, $p = .364$, and Satisfaction, $\beta = -.19$, $p = .121$, were not statistically significant and were negative in sign.

It is important to interpret these negative and non-significant coefficients with caution. At the bivariate level, both perceived Empathy ($r = .633$, $p < .001$) and Satisfaction ($r = .716$, $p < .001$) were strongly and positively correlated with Trust, as reported in Table 4.4. The reversal in sign in the multiple regression therefore does not reflect a substantive negative relationship between these constructs and Trust. Rather, it reflects a suppression and multicollinearity pattern produced by the substantial shared variance among the four predictors. Suppression effects of this kind are well documented in the regression literature and arise when correlated predictors are entered into the same model. In such situations, the standardised coefficient of one variable can shift in sign or magnitude when other correlated variables are added (Cohen, Cohen, West, & Aiken, 2003; Pandey & Elliott, 2010). The coefficients should therefore be interpreted as the unique contribution of each predictor once the variance shared with the others has been partialled out, and not as the total association between that predictor and Trust.

Multicollinearity was further examined through tolerance and VIF values, as reported in Table 4.6. The VIF values ranged from 2.75 for perceived Empathy to 5.25 for Satisfaction, and the tolerance values ranged from .19 to .36. Although, these values did not exceed the conventional thresholds for severe multicollinearity (VIF < 10, tolerance > .10) the VIF for Satisfaction slightly exceeded the threshold of 5, suggesting that the regression coefficients should be treated with caution, given the shared variance among the variables (Hair, Black, Babin, & Anderson, 2019). However, they confirm that a non-trivial portion of the variance in each predictor is shared with the others. The collinearity diagnostics, reported in Table 4.7, provide additional evidence of this shared variance. The final dimension had a condition index of 33.67, with high variance proportions for Satisfaction (.80) and Perceived service quality (.76).

Dimension	Eigenvalue	Condition Index	Empathy	Emotion	Satisfaction	PSQ
1	4.95	1.00	.00	.00	.00	.00
2	0.03	12.92	.00	.15	.01	.00
3	0.01	22.40	.97	.18	.03	.01
4	0.01	23.03	.03	.63	.16	.23
5	0.00	33.67	.01	.04	.80	.76

Table 4.7 Collinearity Diagnostics for the Regression Model Predicting Trust

Taken together, these results indicate that Emotional response and Perceived service quality are the strongest unique predictors of Trust in this model. The negative and non-significant coefficients observed for perceived Empathy and Satisfaction should be interpreted as the residual contribution of these variables once their shared variance with the other predictors has been accounted for, and not as evidence of negative effects. To examine this pattern further, a hierarchical regression analysis was conducted, as reported in Section 4.10.

4.9 Hierarchical regression analysis

A hierarchical regression analysis was conducted to further investigate the suppression and multicollinearity pattern explicit by entering the predictors into two theoretically meaningful blocks. Trust was again entered as the dependent variable. In Model 1, perceived Empathy and Satisfaction were entered as predictors. In Model 2, Emotional response and Perceived service quality were added to the model.

In Model 1, both predictors were statistically significant and positive. Perceived Empathy was a significant positive predictor of Trust, $\beta = .219$, $p = .033$, as was Satisfaction, $\beta = .551$, $p < .001$. The model accounted for 53.3% of the variance in Trust, $R^2 = .533$, and was statistically significant, $F(2, 104) = 59.45$, $p < .001$.

The addition of Emotional response and Perceived service quality in Model 2 significantly improved the model, $\Delta R^2 = .194$, $\Delta F(2, 102) = 36.16$, $p < .001$, bringing the total explained variance to 72.7%, $R^2 = .727$. In Model 2, Emotional response, $\beta = .583$, $p < .001$, and Perceived service quality, $\beta = .554$, $p < .001$, emerged as unique predictors. By contrast, the coefficients for perceived Empathy and Satisfaction shifted: perceived empathy, $\beta = -.078$, $p = .364$, and Satisfaction, $\beta = -.185$, $p = .121$, both became non-significant and negative in sign.

This pattern reinforces the interpretation offered in Section 4.9. When perceived Empathy and Satisfaction were entered into Model 1, they were both significant positive predictors of Trust, in line with the bivariate correlation results. Once Emotional response and Perceived service quality were added to the model, the shared variance was accounted for by these new predictors, and the unique contributions of perceived Empathy and Satisfaction were no longer statistically significant. The shift to negative coefficients in Model 2 should therefore be read as a statistical artefact of the highly correlated predictor set, and not as evidence that perceived Empathy or Satisfaction reduces Trust. The full results of the hierarchical regression are reported in Table 4.8.

Predictor	Model 1 β	p	Model 2 β	p
Empathy	.219	.033	-.078	.364
Satisfaction	.551	< .001	-.185	.121
Emotional response	—	—	.583	< .001
Perceived service quality	—	—	.554	< .001
R^2	.533		.727	
ΔR^2	—		.194	
F change	59.45	< .001	36.16	< .001

Table 4.8 Hierarchical Regression Analysis Predicting Trust

4.10 Mediation analysis

Two mediation analyses were conducted to examine whether Emotional response mediated the relationship between perceived Empathy and the two key evaluative outcomes: Trust and Satisfaction. Both analyses were conducted using Model 4 of the PROCESS macro for SPSS, with 5,000 bootstrap samples and bias-corrected 95% confidence intervals. An indirect effect was interpreted as statistically significant when its bootstrap confidence interval does not include zero (Hayes, 2022).

In the first mediation model, Trust was entered as the dependent variable. Perceived Empathy was a significant predictor of Emotional response, $b = 0.908$, $p < .001$, and Emotional response was a significant predictor of Trust, $b = 0.767$, $p < .001$. After controlling Emotional response, the direct effect of perceived Empathy on Trust was not statistically significant, $b = 0.085$, $p = .431$. The indirect effect of perceived Empathy on Trust through Emotional response was statistically significant, indirect effect = 0.696, 95% bootstrap CI [0.468, 0.936]. Taking together, these results are consistent with full mediation, indicating the relationship between perceived Empathy and Trust in this sample.

In the second mediation model, Satisfaction was entered as the dependent variable. Perceived Empathy was again a significant predictor of Emotional response, $b = 0.908$, $p < .001$, and Emotional response was a significant predictor of Satisfaction, $b = 0.498$, $p < .001$. The direct effect of perceived Empathy on Satisfaction remained statistically significant in this model, $b = 0.318$, $p < .001$. The indirect effect through Emotional response was also statistically significant, indirect effect = 0.452, 95% bootstrap CI [0.256, 0.679]. These results suggest that Emotional response partially mediates the relationship between perceived Empathy and Satisfaction.

Considered together, the two mediation analyses point to Emotional response as an important affective mechanism linking perceived Empathy to user evaluations of the chatbot interaction. The association between perceived Empathy and Trust operates primarily through Emotional response in this sample, while the association between perceived Empathy and Satisfaction appears to operate both directly and indirectly through Emotional response. The complete mediation results are reported in Table 4.9.

Model	Path	b	SE	p	95% CI
Model 1	Empathy → Emotional response	0.908	0.079	< .001	[0.752, 1.065]
(DV: Trust)	Emotional response → Trust	0.767	0.088	< .001	[0.592, 0.942]
	Direct: Empathy → Trust	0.085	0.107	.431	[-0.128, 0.298]
	Indirect via Emotional response	0.696	0.121	—	[0.468, 0.936]
Model 2	Empathy → Emotional response	0.908	0.079	< .001	[0.752, 1.065]
(DV: Satisfaction)	Emotional response → Satisfaction	0.498	0.066	< .001	[0.367, 0.629]
	Direct: Empathy → Satisfaction	0.318	0.081	< .001	[0.158, 0.477]
	Indirect via Emotional response	0.452	0.110	—	[0.256, 0.679]

Table 4.9 Mediation Analysis Results for Trust and Satisfaction

4.11 Summary of findings

The analyses reported in this chapter support a consistent overall pattern. Participants who interacted with the empathetic chatbot reported significantly more positive Emotional responses, higher Satisfaction , higher Perceived service quality, and higher Trust than participants who interacted with the professional chatbot. All four research questions were therefore supported by the independent samples t-tests.

The multiple regression analysis showed that Emotional response and Perceived service quality were the strongest unique predictors of Trust, with the full model accounting for 72.7% of the variance in the outcome. The negative and non-significant coefficients observed for perceived Empathy and Satisfaction in this model do not represent meaningful negative effects. They reflect a suppression and multicollinearity pattern produced by the substantial shared variance among the predictors, as indicated by the bivariate correlations, the tolerance and VIF values, the collinearity diagnostics, and the hierarchical regression. In the hierarchical analysis, perceived Empathy and Satisfaction were both positive and significant predictors of Trust when entered in the first block, and their unique contribution was reduced once Emotional response and Perceived service quality were added to the equation.

The mediation analyses indicate that Emotional response is an important affective mechanism linking perceived Empathy with the evaluative outcomes of the chatbot interaction. The results were consistent with full mediation in the relationship between perceived Empathy and Trust and partially mediated the relationship between perceived Empathy and Satisfaction. A consolidated overview of the main findings, organized by research question and by analysis, is presented in Table 4.10.

Research question / Analysis	**Main finding**	**Conclusion**
RQ1: Emotional response	Participants in the empathetic chatbot condition reported significantly higher Emotional responses than those in the professional condition.	Supported
RQ2: Satisfaction	Participants in the empathetic chatbot condition reported significantly higher Satisfaction than those in the professional condition.	Supported
RQ3: Perceived service quality	Participants in the empathetic chatbot condition reported significantly higher Perceived service quality than those in the professional condition.	Supported
RQ4: Trust	Participants in the empathetic chatbot condition reported significantly higher Trust than those in the professional chatbot condition.	Supported
Mediation: Empathy → Emotional response → Trust	The results were consistent with full mediation, indicating the relationship between perceived Empathy and Trust through Emotional response.	Supported (full mediation)
Mediation: Empathy → Emotional response → Satisfaction	The results were consistent with partial mediation, indicating the relationship between perceived Empathy and Satisfaction through Emotional response.	Supported (partial mediation)

Table 4.10 Summary of Research Questions and Main Findings

5. Results

5.1 Introduction

This chapter interprets the results presented in Chapter 4 in light of the wider literature on AI in customer service. This study aimed to test whether a neutral, professional service communication style or an empathetic service communication style generated more positive user evaluations on four outcomes: Emotional response, Satisfaction, Perceived service quality, and Trust. The chapter also explores the mediation role of Emotional response, given the perceived Empathy and the evaluative results of Satisfaction and Trust. After interpretation of the main findings, the chapter addresses theoretical implications of the study, the suppression phenomenon and multicollinearity pattern in the regression analyses, the limitations of the study, and suggestions for future research.

5.2 Interpretation of the Main Findings

The research provides evidence that the role of the chatbot communication style on user perceptions of an AI-driven service interaction. The participants who interacted with the empathetic chatbot reported a significantly greater Emotional response, Satisfaction, Perceived service quality, and Trust than the participants who interacted with the professional chatbot. The four research questions were thus all supported. The findings were consistent in direction across all outcome constructs, and a manipulation check was performed to ensure that participants could identify the two different conditions, given empathy. The empathetic chatbot was perceived as significantly more empathetic than the professional chatbot.

The finding for Emotional response addresses Research Question 1 and is consistent with the broader body of research about affect in service experience. Participants seem to feel more reassured and less frustrated by the warm and supportive interactions with a chatbot, even though the underlying task was the same in both chatbots. This is in line with the idea that service encounters are judged not only on what is provided, but how they are provided (Bagozzi, Gopinath, & Nyer, 1999; Verhagen, van Nes, Feldberg, & van Dolen, 2014). The current study suggests that the affective cues embedded in the language of the chatbot may have an impact on the participants' reported emotional experiences in a banking scenario, involving a moderate level of personal stress.

The finding for Satisfaction relates to Research Question 2. The empathetic chatbot produced higher Satisfaction scores, despite the same underlying task flow and the problem-solving approach. This means that the participants did not only evaluate the interaction only from the handling of their issue. They also considered the way the chatbot interacted. This is consistent with the "expectancy-disconfirmation" tradition in the service literature, which views Satisfaction as a complex post-consumption assessment, rather than merely as an assessment of whether the functional outcome was met (Oliver, 1980; Oliver, 1993).

The finding of the Perceived quality service addresses Research Question 3 and is theoretically interesting. Perceived service quality is frequently understood as the efficiency, reliability and effectiveness of the service, which may be independent in terms of the communication style if the functional task remains the same. However, the empathetic chatbot received greater Perceived service quality even though the same result was achieved in both conditions. One possible explanation is that empathetic framing affects not only affective judgments, but also general judgments of competence and reliability. This aligns with the Stereotype Content Model (Fiske, Cuddy, & Glick, 2007), which posits that warmth and competence are evaluated simultaneously rather than separately, as well as with previous research on chatbots, which indicates that users' perceptions of the chatbot appear to affect both their cognitive and affective evaluation (Araujo, 2018; Pelau, Dabija, & Ene, 2021).

The finding about Trust relates to Research Question 4 and is important in a financial services context. Trust is often a relational, rather than transactional assessment and its importance might vary regarding the nature of the service interaction. Despite this, the empathetic chatbot still reported higher Trust scores, indicating that even short conversations may affect users' Trust in a service provider when the tone of the conversation suggests care and attentiveness. This aligns with previous studies showing that perceived warmth and human likeness are positive factors in building Trust in a chatbot context (De Cicco, Silva, & Alparone, 2020; Følstad, Nordheim, & Bjørkli, 2018; Nordheim, Følstad, & Bjørkli, 2019).

5.3 The Suppression and Multicollinearity Pattern

One of the main interpretive questions arising from the analysis is related to the regression results. For the multiple regression predicting Trust, Emotional response and Perceived service quality were significant positive predictors whereas perceived Empathy and Satisfaction were not significant and had negative coefficients respectively. At first glance, these coefficients may seem to indicate that perceived Empathy and Satisfaction are negatively related to Trust, a finding that is inconsistent with the bivariate correlations and the substantive argument of the study. Nevertheless, this interpretation would be rejected.

At the zero-order level, both perceived Empathy and Satisfaction were highly and positively correlated with Trust ($r = .633$ and $r = .716$, respectively, $p < .001$). The change in sign of the coefficients of the multiple regression model, then, is unlikely to represent a true negative correlation. It is, however, a signature of a suppression pattern and of multicollinearity that is the consequence of the high level of shared variance among the four predictors, in the regression model (Cohen, Cohen, West, & Aiken, 2003; Pandey & Elliott, 2010). When several variables are entered into the same regression model, the standardized coefficient for each variable presents only its unique contribution after the shared variance with the other variables. In the present study, Empathy and Satisfaction are importantly associated and Trust appears to overlap Emotional response and Perceived service quality. Once, the unique effect of Empathy and Satisfaction become quite small and insignificant in sign.

This relationship is evident in the hierarchical regression. The results of Model 1 (where only perceived Empathy and Satisfaction were entered into the regression) both factors were statistically significant and positive predictors of Trust ($\beta = .219$ and $.551$, respectively, $R^2 = .533$). This means that before Emotional response and Perceived service quality were included in the model the variable, perceived Empathy and Satisfaction explained much of the variance in Trust. When Emotional response and Perceived service quality were added in Model 2, the variance increased significantly ($\Delta R^2 = .194$) and these two variables explained much of the shared variance. However, the non-significance of the two variables should not be interpreted as result that they are not associated with Trust. More specifically, it suggests that their relationship with Trust overlaps notably with Emotional response and Perceived service quality.

The negative coefficients in Model 2 are therefore better explained as an outcome of the highly intercorrelated predictor set in the model than as a result of the negative effects.

This interpretation is confirmed by the multicollinearity diagnostics. The tolerance values were between .19 and .36 and the VIF values were between 2.75 and 5.25, which are not severe levels of multicollinearity, although the highest VIF reaches the threshold of 5 (Hair, Black, Babin, & Anderson, 2019). The collinearity diagnostics also indicated a final dimension with a condition index of 33.67 and high variance proportions concentrated on Satisfaction and Perceived service quality – a pattern that is in line with these two constructs being the closely associated with the predictor set. As a group, the evidence for regression does not cast doubt on the role of Empathy and Satisfaction. It suggests that their impact on Trust may operate through more proximal evaluations of affective and service-quality.

5.4 Integration of the Mediation Findings

The mediation analyses provide a formal test of this mechanism. This relationship between perceived Empathy and Trust consisted with full mediation by Emotional response. Emotional response was highly correlated with perceived Empathy and Trust was highly correlated with Emotional response. However, once Emotional response was controlled, the direct relationship between perceived Empathy and Trust was not statistically significant. The size of the indirect effect was estimated by its bootstrap confidence interval that did not include zero. In this sample, there is not a direct relationship between empathetic communication and Trust. Instead, the relationship between perceived Empathy and Trust appeared to be explained through the Emotional response generated during the chatbot conversation.

The pattern was different for Satisfaction. Emotional response partially mediated the relationship between perceived Empathy and Satisfaction. The direct pathway between perceived Empathy and Satisfaction, as well as the indirect pathway through the Emotional response, were both statistically significant. Thus, this suggests that Satisfaction seems to arise from two sources: one affective channel via Emotional response and a more direct evaluative channel, whereby users evaluate the interaction as satisfying because the chatbot communicated in an empathetic communication style.

The mediation results contribute to clarifying the overall pattern of the results. The bivariate correlations indicate there is a relationship between perceived Empathy with Trust and Satisfaction. The regression results indicate, however, suggest that when entered together with Emotional response and Perceived service quality, perceived Empathy is not the strongest proximal predictor of Trust. The mediation analyses bring these two together by demonstrating that Emotional response functions as a link between perceived Empathy and the two outcomes. Emotional response should, therefore, not be treated as a competitor to perceived Empathy in predicting Trust but rather an affective mechanism in which empathetic framing translates to a relational outcome. This result aligns with theoretical explanations which suggest that affective response may be an immediate result of empathetic communication and serves as a mechanism by which empathetic communication contributes to downstream evaluations (Liu & Sundar, 2018; Pelau et al., 2021).

5.5 Theoretical Implications

The study can be a useful contribution to the chatbots literature in several ways. First, it provides evidence that communication style is an important design feature in AI-enabled customer service. Language model, task flow and survey instrument held constant, the study demonstrates that differences in communication style alone are statistically significant for four theoretically significant user outcomes. However, the contribution is also methodological. Specifically, numerous chatbot experiments combine communication style with identity cues or with style differences in the task-flow (Araujo, 2018; Go & Sundar, 2019) and it is difficult to separate out the active element of manipulation. The present design provides a more concentrated test of the contribution of tone.

Second, the study provides a better theoretical understanding of the relationship between perceived Empathy and Trust in chatbot service. Much of the literature has assumed there is a direct relationship between Empathy and Trust, so that Empathy is seen as being a characteristic that fosters Trust as it communicates care and attentiveness (Følstad et al., 2018; Pelau et al., 2021). The results of the present study indicate that the process is more complex. Perceived Empathy is also an indirect contributor to Trust, through Emotional response. In this sample, shows that when the affective consequence of empathetic communication is controlled, there is no additional variance in the perception of Empathy to be added to Trust. The implication suggests that Emotional response is not secondary outcome of empathetic communication style, yet an affective way through which perceived Empathy is related to Trust.

Third, the results reinforce the argument that chatbot experience should be considered multi-dimensional rather than global. Emotional response, Satisfaction, Perceived service quality and Trust had strong inter-correlations but were not equivalent. All were sensitive to manipulation, yet each played a different role in the regression models. Theoretical research on the chatbot evaluation should build on the concept of distinguish between the types of outcomes (affective, evaluative, and relational) and should specify the paths of how these three outcomes are connected, instead of considering them as single a user-experience composite.

Fourth, the self-reported measure of Emotional response is aligned with the conceptual focus of the study. The research questions focus on how users subjectively experience interaction, making a suitable measurement approach based on the consumer-affect tradition (Bagozzi et al., 1999; Mauss & Robinson, 2009) than automated textual sentiment classification. This result confirms the importance of this measurement choice; the variable that performs the explanatory role in the model is the users' felt emotional experience, not the surface of language used in the conversation.

5.6 Managerial Implications

The findings provide organizations with practical tips on how they can use chatbots in customer-facing situations. The first implication is the fact that the tone of the chatbots is not merely an aesthetic feature. The empathetic chatbot in the present study achieved the same functional task as the professional chatbot yet elicited significantly more positive user ratings in all measured outcomes. The implications for service providers suggest that carefully designing empathetic responses through conversational design may translate into measurable improvements in customer experience metrics.

The mediation results lead to a second managerial implication. Thus, in empathetic communication, the aim should not simply be to make the chatbot just sound warmer yet should generate an Emotional response that supports Trust and reassurance. Empathetic framing requires tuning to the circumstance. Affective response that leads to the formation of Trust is not likely to be achieved when there is no specific concern of the user. Empathetic language should recognize the user's situation, communicate understanding of the problem, and provide reassuring messages that are relevant to the stakes of the service encounter.

The third implication concerns the combination of empathetic communication with effective execution of tasks. Perceived service quality and Emotional response were the two strongest factors of Trust on the regression model. This indicates that Trust is not shaped by warmth, nor by being alone, but by the combination of perceived service and emotional reassurance effectiveness. Chatbot designs that fail to obtain emotional comfort in a way that also facilitates clear and efficient task completion may weaken customer Trust and the reverse is also true: designs that focus on efficiency, but not emotional comfort may likely reduce customer Trust. Thus, Empathy and task performance should be treated as mutually complementary elements.

The fourth practical implication is related to the strategic positioning of customer chatbots. Participants reported higher Trust when interacting with the empathetic chatbot, which means that the chatbot may be considered as a representative of the service provider and not just a technical tool. The tone of the chatbot may shape users' overall impression of the organization. In industries where Trust is a critical factor (such as banking and financial services) the design of chatbot communications should be considered as part of brand and customer experience strategy, more than a technical or operational issue.

A fifth, more cautious implication addresses the appropriate use of empathetic framing. Although empathetic framing had more positive effects in the present study, previous studies have found that inauthentic, and excessive or badly adjusted from non-human agents may have had detrimental effects on the human agent's credibility (Liu & Sundar, 2018; Mozafari, Weiger, & Hammerschmidt, 2022). Therefore, service providers should understand how to use empathetic communication carefully, without overestimating the emotional abilities of the chatbot, and, at the same time, should ensure that the empathetic framing is consistent with the wider service positioning.

5.7 Limitations

The study has certain limitations. The first limitation is about the sample. The participants were only convenience sampled and included a greater number of female participants and more mature participants (aged 25-44). Random assignment to the two conditions strengthens the internal validity of the experimental comparisons, the demographic characteristics of the sample limit the external validity of the results. Future studies should use larger and more representative sample sizes to test whether the outcome extends to different demographic samples.

The second constraint is that of the experimental scenario. Participants did not experience a real banking service interaction with real financial consequences but took part in a simulated service experience. Experimental scenarios are common in the service environment and allow for greater control over the stimulus, but they may not generate the same intensity of emotion as a real lost-card situation. In parallel, demand effects were also introduced from the use of simulated scenario, as participants were asked to respond as if the scenario were real.

A final problem relates to the reliance on self-reports measures. Social desirability, retrospective rationalization, and common-method variance may impact self-report data, especially when all measures are administered at the same time and use similar type of responses (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Future research could supplement self-report measures of Trust with behavioral measures, such as whether users' willingness to provide information during the interaction, select the chatbot over another human agent in a subsequent activity.

A fourth limitation is that the study did not include moderating variables. Several potential moderators may influence the effects observed, including individual differences in prior experience with chatbots, digital literacy, dispositional Trust in technology, and preferred communication style. These factors were not measured in detail to be modelled as moderators in the present study. Future studies would better define the boundary conditions of empathetic communication style including the specific variables.

The fifth limitation is related to the study design, which was both cross sectional and single chatbot encounter. Trust is theorized to develop over time and the impact observed in the present study reflects the initial stage of Trust rather than lasting Trust (McKnight, Cummings, & Chervany, 1998; Nordheim et al., 2019). Whether the effect of the empathetic communication style is an advantage that would persist or diminish across repeated interaction remains an open question.

5.8 Directions for Future Research

The present study has several implications for future research. First, mediation model could have been extendable beyond Emotional response by including different mechanisms such as perceived warmth, perceived competence, perceived authenticity and social presence. A more elaborated model may explain if emotional reaction is the may affective mechanism through which empathetic communication influence user evaluations or it operates in parallel to other affective and cognitive mechanisms.

Second, longitudinal studies that follow users' interactions with the same chatbot over time would allow researchers to better understand how the effects seen in the present study change over time. Of particular interest is whether the advantage empathetic style remains once the novelty effect has passed, whether it increases as users' repeated use, and whether decreases because the repeated empathetic framing may appear formulaic. There are different design implications for each of these trajectories.

Third, future research may examine whether the model would be applicable to different sectors of service. Empathic framing may be more useful in high stakes contexts such as banking but may not be as useful in low stakes contexts such as e-commerce browsing assistance or scheduling support. Across-sector comparative research would help the research to investigate the contextual boundaries of the empathetic-style effect.

Fourth, future research examines individual difference moderators should be explored in an explicit way. The magnitude and nature of the effects may be conditioned by dispositional Trust in technology, preferred communication mode, age, previous experience with chatbots, and cultural background. The present sample is mainly composed of Greek-speaking participants, and a cross-cultural extension could reveal if the pattern described in this study can be extended to other linguistic and cultural contexts.

Finally, a further direction would be to integrate self-reported sentiment with automatic linguistic analysis of the user's own messages during the chatbot interaction. Although, the present study intentionally relies on self-report for the reasons discussed in Chapter 2, a multi-method approach that triangulates the users' affect with linguistic features of the user's text could provide a more holistic view of the impact of empathetic chatbot communication user's experience and expression of affect.

5.9 Conclusion

This research examined whether an empathetic communication style of a customer service chatbot, rather than a neutral and professional, would elicit more positive customer evaluations. All four research questions were supported. Participants in the empathetic chatbot indicated significantly more positive Emotional response, higher Satisfaction and Perceived service quality, and higher Trust than the professional chatbot condition participants.

The regression analyses revealed that Emotional response and Perceived service quality were the strongest unique predictor variables of Trust. The negative and non-significant coefficients found for perceived Empathy and Satisfaction in the full regression model should not necessarily indicate negative effects. Rather, they indicate a suppression and multicollinearity pattern as a result of notably shared variance among the predictors. This interpretation was futher confirmed by the hierarchical regression analysis and the mediation analyses revealed that Emotional response acted as a mechanism that explains the impact of perceived Empathy on Trust and, to a certain extent, Satisfaction.

The results suggest that chatbot communication design is not only about technological skills or tasks performance, but also about users' psychological experience. Services that combine efficient task handling with care that is tailored to the value of the service encounter may have a better chance at creating users' perceptions of reassuring, satisfying, high in service quality and Trustworthy.

References

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets, 31*(2), 427–445. https://doi.org/10.1007/s12525-020-00414-7

Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). Customer Satisfaction , market share, and profitability: Findings from Sweden. *Journal of Marketing*, *58*(3), 53–66. https://doi.org/10.1177/002224299405800304

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. https://doi.org/10.1016/j.chb.2018.03.051

Ashfaq, M., Yun, J., Yu, S., & Loureiro, S. M. C. (2020). I, chatbot: Modeling the determinants of users' Satisfaction and continuance intention of AI-powered service agents. *Telematics and Informatics, 54*, 101473. https://doi.org/10.1016/j.tele.2020.101473

Bagozzi, R. P. (2007). The legacy of the Technology Acceptance Model and a proposal for a paradigm shift. *Journal of the Association for Information Systems*, *8*(4), 244–254. https://doi.org/10.17705/1jais.00122

Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science, 27*(2), 184–206. https://doi.org/10.1177/0092070399272005

Beattie, A., Edwards, A. P., & Edwards, C. (2020). A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies, 71*(3), 409–427. https://doi.org/10.1080/10510974.2020.1725082

Brachten, F., Kissmer, T., & Stieglitz, S. (2021). The acceptance of chatbots in an enterprise context: A survey study. *International Journal of Information Management, 60*, 102375. https://doi.org/10.1016/j.ijinfomgt.2021.102375

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Bolton, R. N., & Lemon, K. N. (1999). A dynamic model of customers' usage of services: Usage as an antecedent and consequence of Satisfaction . *Journal of Marketing Research*, *36*(2), 171–186. https://doi.org/10.1177/002224379903600203

Bolton, R. N., McColl-Kennedy, J. R., Cheung, L., Gallan, A., Orsingher, C., Witell, L., & Zaki, M. (2018). Customer experience challenges: Bringing together digital, physical and social realms. *Journal of Service Management*, *29*(5), 776–808. https://doi.org/10.1108/JOSM-04-2018-0113

Borsci, S., Malizia, A., Schmettow, M., van der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2022). The Chatbot Usability Scale: The design and pilot of a usability scale for interaction with AI-based conversational agents. *Personal and Ubiquitous Computing*, *26*, 95–119. https://doi.org/10.1007/s00779-021-01582-9

Bitner, M. J. (1990). Evaluating service encounters: The effects of physical surroundings and employee responses. *Journal of Marketing*, *54*(2), 69–82. https://doi.org/10.1177/002224299005400206

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1981). Designing research for application. *Journal of Consumer Research*, *8*(2), 197–207. https://doi.org/10.1086/208856

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Chen, J. S., Le, T. T. Y., & Florence, D. (2021). Usability and responsiveness of artificial intelligence chatbot on online customer experience in e-retailing. *International Journal of Retail & Distribution Management*, *49*(11), 1512–1531. https://doi.org/10.1108/IJRDM-08-2020-0312

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer Satisfaction regarding luxury brands. *Journal of Business Research*, *117*, 587–595. https://doi.org/10.1016/j.jbusres.2018.10.004

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.

Cronin, J. J., Brady, M. K., & Hult, G. T. M. (2000). Assessing the effects of quality, value, and customer Satisfaction on consumer behavioral intentions in service environments. *Journal of Retailing*, *76*(2), 193–218. https://doi.org/10.1016/S0022-4359(00)00028-2

Cronin, J. J., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing*, *56*(3), 55–68. https://doi.org/10.1177/002224299205600304

Dabholkar, P. A., Shepherd, C. D., & Thorpe, D. I. (2000). A comprehensive framework for service quality: An investigation of critical conceptual and measurement issues. *Journal of Retailing*, *76*(2), 139–173. https://doi.org/10.1016/S0022-4359(00)00029-4

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

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

De Cicco, R., Silva, S. C., & Alparone, F. R. (2020). Millennials' attitude toward chatbots: An experimental study in a social relationship perspective. *International Journal of Retail & Distribution Management*, *48*(11), 1213–1233. https://doi.org/10.1108/IJRDM-12-2019-0406

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100. https://doi.org/10.1177/1534582304267187

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage.

Donovan, R. J., & Rossiter, J. R. (1982). Store atmosphere: An environmental psychology approach. *Journal of Retailing*, *58*(1), 34–57.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). Sage.

Field, A., & Hole, G. (2003). *How to design and report experiments*. Sage.

Følstad, A., & Brandtzaeg, P. B. (2017). Chatbots and the new world of HCI. *Interactions, 24*(4), 38–42. https://doi.org/10.1145/3085558

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What makes users Trust a chatbot for customer service? An exploratory interview study. In S. Bodrunova (Ed.), *Internet Science* (pp. 194–208). Springer. https://doi.org/10.1007/978-3-030-01437-7_16

Følstad, A., & Brandtzaeg, P. B. (2020). Users' experiences with chatbots: Findings from a questionnaire study. *Quality and User Experience, 5*(1), 3. https://doi.org/10.1007/s41233-020-00033-2

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the Computers Are Social Actors paradigm. *Human-Machine Communication, 1*, 71–85. https://doi.org/10.30658/hmc.1.5

Glikson, E., & Woolley, A. W. (2020). Human Trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior, 97,* 304–316. https://doi.org/10.1016/j.chb.2019.01.020

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). The Guilford Press.

Han, E., Yang, M., & Park, J. (2022). The effects of perceived chatbot characteristics on consumer responses in service encounters. *Journal of Retailing and Consumer Services, 67*, 102978.

Hovy, D. (2020). *Text analysis in Python for social scientists: Discovery and exploration*. Cambridge University Press.

Hsu, C.-L., & Lin, J. C.-C. (2023). Understanding the user Satisfaction and loyalty of customer service chatbots. *Journal of Retailing and Consumer Services, 71*, 103211. https://doi.org/10.1016/j.jretconser.2022.103211

Hu, M., Liu, B., & Cheng, J. (2004). Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 168–177). ACM. https://doi.org/10.1145/1014052.1014073

Hu, P., Lu, Y., & Gong, Y. (2021). Dual humanness and Trust in conversational AI: A person-centered approach. *Computers in Human Behavior, 119*, 106727. https://doi.org/10.1016/j.chb.2021.106727

Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research, 21*(2), 155–172. https://doi.org/10.1177/1094670517752459

Huang, M.-H., & Rust, R. T. (2021). Engaged to a robot? The role of AI in service. *Journal of Service Research, 24*(1), 30–41. https://doi.org/10.1177/1094670520902266

Janssen, A., Passlick, J., Rodríguez Cardona, D., & Breitner, M. H. (2020). Virtual assistance in any context: A taxonomy of design elements for domain-specific chatbots. *Business & Information Systems Engineering, 62*(3), 211–225. https://doi.org/10.1007/s12599-020-00644-1

Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., et al. (2017). "Service Encounter 2.0": An investigation into the roles of technology, employees and customers. *Journal of Business Research, 79*, 238–246. https://doi.org/10.1016/j.jbusres.2017.03.008

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool.

Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking, 21*(10), 625–636. https://doi.org/10.1089/cyber.2018.0110

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, *38*(6), 937–947. https://doi.org/10.1287/mksc.2019.1192

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*(2), 209–237. https://doi.org/10.1080/02699930802204677

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational Trust. *Academy of Management Review, 20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

McAllister, D. J. (1995). Affect- and cognition-based Trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24–59. https://doi.org/10.5465/256727

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial Trust formation in new organizational relationships. *Academy of Management Review, 23*(3), 473–490. https://doi.org/10.5465/amr.1998.926622

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

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. MIT Press.

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In H. L. Meiselman (Ed.), *Emotion measurement* (pp. 201–237). Woodhead Publishing. https://doi.org/10.1016/B978-0-08-100508-8.00009-6

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2022). Trust me, I'm a bot: Repercussions of chatbot disclosure in different service frontline settings. *Journal of Service Management*, *33*(2), 221–245. https://doi.org/10.1108/JOSM-10-2020-0380

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An initial model of Trust in chatbots for customer service: Findings from a questionnaire study. *Interacting with Computers*, *31*(3), 317–335. https://doi.org/10.1093/iwc/iwz022

Norton, R. W. (1978). Foundation of a communicator style construct. *Human Communication Research*, *4*(2), 99–112. https://doi.org/10.1111/j.1468-2958.1978.tb00600.x

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

OpenAI. (2024). *GPT-4o mini: Advancing cost-efficient intelligence*. OpenAI. https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of Satisfaction decisions. *Journal of Marketing Research*, *17*(4), 460–469. https://doi.org/10.1177/002224378001700405

Oliver, R. L. (1993). Cognitive, affective, and attribute bases of the Satisfaction response. *Journal of Consumer Research*, *20*(3), 418–430. https://doi.org/10.1086/209358

Pandey, S., & Elliott, W. (2010). Suppressor variables in social work research: Ways to identify in multiple regression models. *Journal of the Society for Social Work and Research*, *1*(1), 28–40. https://doi.org/10.5243/jsswr.2010.2

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135. https://doi.org/10.1561/1500000011

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(1), 12–40.

Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). E-S-QUAL: A multiple-item scale for assessing electronic service quality. *Journal of Service Research, 7*(3), 213–233. https://doi.org/10.1177/1094670504271156

Pelau, C., Dabija, D.-C., & Ene, I. (2021). What makes an AI device human-like? The role of interaction quality, Empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. *Computers in Human Behavior, 122*, 106855. https://doi.org/10.1016/j.chb.2021.106855

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879b

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Rese, A., Ganster, L., & Baier, D. (2020). Chatbots in retailers' customer communication: How to measure their acceptance? *Journal of Retailing and Consumer Services, 56*, 102176. https://doi.org/10.1016/j.jretconser.2020.102176

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shin, D. (2021). The effects of explainability and causability on perception, Trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). MIT Press.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Trivedi, J. (2019). Examining the customer experience of using banking chatbots and its impact on brand love: The moderating role of perceived risk. *Journal of Internet Commerce*, *18*(1), 91–111. https://doi.org/10.1080/15332861.2019.1567188

Vargo, S. L., & Lusch, R. F. (2004). Evolving to a new dominant logic for marketing. *Journal of Marketing*, *68*(1), 1–17. https://doi.org/10.1509/jmkg.68.1.1.24036

Vargo, S. L., & Lusch, R. F. (2008). Service-dominant logic: Continuing the evolution. *Journal of the Academy of Marketing Science, 36*(1), 1–10. https://doi.org/10.1007/s11747-007-0069-6

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a research agenda on interventions. *Decision Sciences*, *39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Verhagen, T., van Nes, J., Feldberg, F., & van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication, 19*(3), 529–545. https://doi.org/10.1111/jcc4.12066

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., et al. (2022). Emergent abilities of large language models. *Transactions on Machine Learning Research*.

Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

Westbrook, R. A., & Oliver, R. L. (1991). The dimensionality of consumption, emotion patterns and consumer Satisfaction . *Journal of Consumer Research*, *18*(1), 84–91. https://doi.org/10.1086/209243

Wieseke, J., Geigenmüller, A., & Kraus, F. (2012). On the role of Empathy in customer-employee interactions. *Journal of Service Research*, *15*(3), 316–331. https://doi.org/10.1177/1094670512439743

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management*, *29*(5), 907–931. https://doi.org/10.1108/JOSM-04-2018-0119

Wirtz, J., & Lwin, M. O. (2009). Regulatory focus theory, Trust, and privacy concern. *Journal of Service Research*, *12*(2), 190–207. https://doi.org/10.1177/1094670509335772

Yoon, C., & Steege, L. M. B. (2013). Development of a quantitative model of the impact of customers' personality and perceptions on internet banking use. *Computers in Human Behavior*, *29*(3), 1133–1141. https://doi.org/10.1016/j.chb.2012.10.005

Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. *Journal of Marketing*, *52*(3), 2–22. https://doi.org/10.1177/002224298805200302

Appendix A: Full Prompts for Both Chatbot Versions

Empathetic Chatbot Link: https://udify.app/chat/4m7FYTHKJ74E0bsl

MISSION

You are an AI Customer Service Assistant in a banking customer service scenario.

Your role is to assist users who report that they have lost their bank card and are concerned about possible unauthorized use. Your task is to help secure the user's card and provide reassurance.

If the user reports an issue unrelated to a lost or stolen bank card, you must gently guide the conversation toward lost card assistance WITHOUT making assumptions.

You must always respond in the language used by the user (Greek or English).

STYLE RULES

- Use "we", "our services", "our support"

- Do NOT refer to "the bank" as a third party

- Provide direct guidance as a bank representative

TONE CONTROL (STRICT)

Responses must be:

- Professional

- Calm

- Empathetic (moderate and controlled)

- Reassuring

Avoid:

- Casual language

- Overly emotional or exaggerated empathy

- Informal expressions

Use formal phrasing such as:

"I understand that this situation may be concerning."

"I am here to assist you."

"Let us address this immediately."

CRITICAL BEHAVIOR RULES

- NEVER assume the user has lost their card unless they explicitly say so

- NEVER say or imply:

"I see that you lost your card"

"I have been informed that..."

or any invented situation

- If the user has NOT clearly stated the issue:

→ Ask a neutral clarification question

- If the user sends ONLY a greeting (e.g., "hello", "hi", "καλημέρα"):

→ Respond with a professional greeting

→ Ask how you can assist

→ DO NOT mention lost cards

- Do NOT mention lost or stolen cards unless the user explicitly refers to it first

RESPONSE RULES (STRICT)

- Each response MUST contain 1 to 3 sentences only

- NEVER exceed 3 sentences

- Use one short paragraph only

- Do NOT use lists or bullet points

- Keep responses concise and controlled

- You must internally track the number of user messages.

Maximum allowed: 5 user messages

CONVERSATION FLOW

When the user clearly states they lost their card:

→ Respond with Empathy and acknowledgment

Example:

"I understand how concerning this situation can be. Please give me a moment while I check the status of your card."

After that:

→ Provide the solution with reassurance

Example:

"Your card has been temporarily blocked to prevent unauthorized use. Your account is now secure. You can request a new card through our services."

After providing the solution:

→ Ask:

"Is there anything else I can help you with?"

If the user responds:

→ Provide ONE final short response (1–3 sentences)

→ Then end the interaction with:

"Thank you for contacting our support. Please complete the short survey about your experience by clicking the link below:

https://forms.gle/LQYNVPX78sdkQGaS8

The survey takes about 3–5 minutes."

IMPORTANT

- Do NOT stop the conversation early

- Do NOT track or count user messages

- Continue naturally until the closing condition is met

STRICT ENDING CONTROL

After you provide the solution about the lost card, you MUST:

1. Ask:

"Is there anything else I can help you with?"

2. If the user responds to this question:

→ Provide ONE final short response (1–3 sentences)

3. Immediately after that, ALWAYS end the conversation with:

"Thank you for contacting our support. Please complete the short survey about your experience by clicking the link below:

https://forms.gle/LQYNVPX78sdkQGaS8

The survey takes about 3–5 minutes."

4. Do NOT continue the conversation after this message.

5. For any further user messages, ALWAYS repeat the same survey message.

Professional Chatbot Link: https://udify.app/chat/lHHe4pVZUbjmClru

MISSION

You are an AI Customer Service Assistant in a banking customer service scenario.

Your role is to assist users ONLY with issues related to a lost or stolen bank card and help secure the card.

If the user mentions a different topic, you must gently guide the conversation toward lost card assistance WITHOUT making assumptions.

You must always respond in the language used by the user (Greek or English).

STYLE RULES

- Use "we", "our services", "our support"

- Do NOT refer to "the bank" as a third party

- Sound like a real bank representative

TONE CONTROL (STRICT)

Responses must be:

- Professional

- Neutral

- Calm

Avoid:

- Emotional or empathetic language

- Informal expressions

- Reassuring or expressive phrasing

CRITICAL BEHAVIOR RULES

- NEVER assume the user has lost their card unless they explicitly say so

- NEVER say or imply:

"I see that you lost your card"

"I have been informed that..."

"I understand that you already..."

or any invented situation

- If the user has NOT clearly stated the issue:

→ Ask a neutral clarification question

- If the user sends ONLY a greeting (e.g., "hello", "hi", "καλημέρα"):

→ Respond with a polite greeting

→ Ask how you can help

→ DO NOT mention lost or stolen cards

→ WAIT for the user to state their issue

Example:

"Hello. I am your virtual assistant. How can I assist you today?"

- Do NOT mention lost or stolen cards unless the user explicitly refers to it first

RESPONSE RULES (STRICT)

- Each response MUST contain 1 to 3 sentences only

- NEVER exceed 3 sentences

- Use one short paragraph only

- Do NOT use lists or bullet points

- Keep responses concise and controlled

- You must internally track the number of user messages. Maximum allowed: 5 user messages

CONVERSATION FLOW

When the user clearly states they lost their card:

→ Respond neutrally

Example:

"Please give me a moment while I check the status of your card."

After that:

→ Provide the solution

Example:

"Your card has been temporarily blocked to prevent unauthorized use. You can request a replacement card through our services."

After providing the solution:

→ Ask:

"Can I assist you with anything else?"

If the user responds:

→ Provide ONE final short response (1–3 sentences)

→ Then end the interaction with:

"Thank you for contacting our support. Please complete the short survey about your experience by clicking the link below:

https://forms.gle/Tax8SbptPZjQjQrU8

The survey takes about 3–5 minutes."

IMPORTANT

- Do NOT stop the conversation early

- Do NOT track or count user messages

- Continue naturally until the closing condition is met

STRICT ENDING CONTROL

After you provide the solution about the lost card, you MUST:

1. Ask:

"Is there anything else I can help you with?"

2. If the user responds to this question:

→ Provide ONE final short response (1–3 sentences)

3. Immediately after that, ALWAYS end the conversation with:

"Thank you for contacting our support. Please complete the short survey about your experience by clicking the link below:

https://forms.gle/LQYNVPX78sdkQGaS8

The survey takes about 3–5 minutes."

4. Do NOT continue the conversation after this message.

5. For any further user messages, ALWAYS repeat the same survey message.

Appendix B: Questionnaire

- **A1. Ενημέρωση συμμετεχόντων και οδηγίες**

Σας ευχαριστώ για τη συμμετοχή σας στη συνομιλία με το chatbot εξυπηρέτησης πελατών.

Το παρόν ερωτηματολόγιο αποτελεί μέρος της μεταπτυχιακής μου διατριβής και στόχο έχει να εξετάσει την εμπειρία των χρηστών κατά την αλληλεπίδρασή τους με chatbots εξυπηρέτησης πελατών σε διαδικτυακά περιβάλλοντα.

Θα σας παρακαλούσα να απαντήσετε στις παρακάτω ερωτήσεις με βάση την εμπειρία που μόλις είχατε κατά τη διάρκεια της συνομιλίας με το chatbot.

Η συμμετοχή σας είναι εθελοντική και όλες οι απαντήσεις θα παραμείνουν ανώνυμες. Τα δεδομένα θα χρησιμοποιηθούν αποκλειστικά για ακαδημαϊκούς σκοπούς στο πλαίσιο της μεταπτυχιακής μου διατριβής. Δεν θα συλλεχθούν προσωπικά ή ευαίσθητα δεδομένα.

Η συμπλήρωση του ερωτηματολογίου διαρκεί περίπου 3–5 λεπτά.

Ευχαριστώ εκ των προτέρων.

- **A2. Ερώτηση ελέγχου**

1. Ολοκληρώσατε τη συνομιλία με το chatbot πριν συνεχίσετε στο ερωτηματολόγιο;

- Ναι
- Όχι

- **A3. Δημογραφικά και βασικά στοιχεία**

2. Έχετε χρησιμοποιήσει στο παρελθόν chatbots εξυπηρέτησης πελατών;

- Ναι
- Όχι

3. Φύλο

- Γυναίκα
- Άνδρας
- Άλλο / Προτιμώ να μην απαντήσω

4. Ηλικιακή ομάδα

- 18–24
- 25–34
- 35–44
- 45–54
- 55+

- **A4. Κλίμακα απάντησης**

Για τις παρακάτω δηλώσεις χρησιμοποιήθηκε πενταβάθμια κλίμακα Likert:

1 = Διαφωνώ απόλυτα
5 = Συμφωνώ απόλυτα

- **A5. Αντιλαμβανόμενη ενσυναίσθηση**

5. Το chatbot φάνηκε να κατανοεί την κατάστασή μου.
6. Το chatbot αναγνώρισε την ανησυχία μου.
7. Το chatbot ανταποκρίθηκε με υποστηρικτικό και κατανοητικό τρόπο.
8. Ένιωσα ότι με καταλαβαίνουν κατά τη διάρκεια της αλληλεπίδρασης.

- **A6. Συναισθηματική απόκριση**

9. Η αλληλεπίδραση μείωσε τον εκνευρισμό ή την ανησυχία μου.
10. Η αλληλεπίδραση με έκανε να αισθανθώ πιο ήρεμος/η.

- **Α7. Ικανοποίηση**

11. Είμαι ικανοποιημένος/η από αυτή την αλληλεπίδραση με το chatbot.
12. Το chatbot χειρίστηκε αποτελεσματικά το πρόβλημά μου.
13. Η αλληλεπίδραση ανταποκρίθηκε στις προσδοκίες μου από την εξυπηρέτηση.

- **Α8. Αντιλαμβανόμενη ποιότητα υπηρεσίας**

14. Η υπηρεσία που παρείχε το chatbot ήταν αποτελεσματική.
15. Η διαδικασία εξυπηρέτησης ήταν αξιόπιστη.
16. Συνολικά, η ποιότητα της υπηρεσίας που παρείχε το chatbot ήταν υψηλή.

- **Α9. Εμπιστοσύνη**

17. Θα εμπιστευόμουν ένα τέτοιου είδους chatbot για την αντιμετώπιση παρόμοιων προβλημάτων στο μέλλον.
18. Θα χρησιμοποιούσα ξανά ένα τέτοιου είδους chatbot για εξυπηρέτηση πελατών.
19. Νιώθω άνετα να βασίζομαι σε τέτοιου είδους ψηφιακή εξυπηρέτηση.

Author’s Statement:

I hereby expressly declare that, according to the article 8 of Law 1559/1986, this dissertation is solely the product of my personal work, does not infringe any intellectual property, personality and personal data rights of third parties, does not contain works/contributions from third parties for which the permission of the authors/beneficiaries is required, is not the product of partial or total plagiarism, and that the sources used are limited to the literature references alone and meet the rules of scientific citations.