Assessing the Impact of Chatbots on Customer Service Efficiency: A Comparative Study with Human Agents

Fatima Bilal, József Mezei, Bad re Alam
Business and Economics, and Law, Åbo Akademi University, Faculty of Social Sciences, Turku, Finland
fatima.Bilal@abo.fi, badre.alam@abo.fi, jozsef.mezei@abo.fi

This research investigates the way AI chatbots influence consumer service efficiency compared to human workers. The proliferation of chatbots across businesses emphasizes the importance of evaluating their impact on productivity, customer satisfaction, and service quality. Although chatbots are effective, they struggle with delicate or complicated interactions. Humans remain essential for providing sympathetic and thorough responses. This study reveals a significant gap in chatbots' ability to meet customer expectations, especially when individualized attention and trust are required. A quantitative research approach was used to collect survey data from customers having experience with both human agents and virtual assistants, i.e., chatbots. According to the findings, while routine inquiries are effectively handled by chatbots, they still struggle with contextual understanding and emotional intelligence. The ideal way to maximize client fulfillment and quality of service is to implement a service structure that combines chatbots for repetitive tasks and human agents for more complicated interactions. The study emphasizes how crucial it is to update chatbot systems frequently to satisfy evolving client demands and boost service effectiveness.

DOI
https://doi.org/10.18690/um.fov.4.2025.9

ISBN
978-961-286-998-4

Keywords:
chatbot, customer satisfaction, human agents, responses, customer engagement

University of Maribor Press

1 Introduction

In recent years, automated chatbots have become ubiquitous in customer service, handling tasks from answering simple FAQs to managing complex requests. Companies deploy chatbots to improve response times and operational efficiency while aiming to maintain high service quality (Smith, 2021). However, a fundamental question remains: how do chatbot-driven service interactions compare to those handled by human agents in terms of customer satisfaction, service efficiency, and perceived service quality? Existing research has largely examined chatbots in isolation or through limited scenarios, without directly benchmarking them against human agents across multiple performance metrics. De Keyser (2020) suggests that giving chatbots human-like characteristics, such as humor and a more personable conversational style, could make interactions more engaging for customers, but empirical evidence for the impact of these features on customer experience remains limited. It is still unclear whether a humorous, anthropomorphized chatbot improves customer satisfaction or if it might undermine professionalism. Initial studies have begun to explore such design elements, but their effects on key service outcomes have yet to be conclusively demonstrated. This uncertainty has prompted calls for deeper investigation into chatbot personalization, underscoring a second gap in the literature. Moreover, context may significantly influence the success of chatbot deployments. Customer expectations and tolerance for automation vary across industries—what works well for a retail chatbot may not translate to a banking or healthcare setting. For example, banking customers might prioritize accuracy and formality, whereas retail customers could respond more favorably to a friendly, playful tone. Despite these differences, few studies have systematically compared chatbot and human agent performance across different service sectors. As a summary, we have limited insight into whether industry-specific factors moderate the outcomes of chatbot versus human interactions or alter the effectiveness of personalization features. This study contributes to the customer service and AI literature by addressing these interconnected gaps with original empirical evidence. We conduct a comprehensive comparison of chatbot and human agent performance using data collected from three distinct industries: banking, retail, and healthcare. Key performance indicators—including customer satisfaction, service efficiency, and perceived service quality—are measured to evaluate where chatbots can match or even exceed human agents and where they still fall short. In doing so, we extend prior work that examined isolated aspects of chatbot efficacy (Smith, 2021) by

providing a holistic, side-by-side assessment of AI versus human service delivery. Additionally, we incorporate personalized chatbot behaviours, specifically a humorous communication style and other human-like cues, to examine their influence on the customer experience. By testing interactions with and without these personalization features, our study directly responds to calls for further research into chatbot personalization in service encounters (De Keyser, 2020).

1.1 Research Gap

Chatbots are used in customer service extensively to save costs, yet there is limited research on how well these bots deliver a smooth and positive customer experience. There is limited research on what effect that the efficiency gains using chatbots have on overall customer satisfaction and trust (Kaczorowska, 2019). There are significant challenges related to maintaining and establishing the trust of customers during interactions. Customers also like to interact with chatbots while having expectations that they will mimic human-like interactions and understand their emotional contexts (Selamat & Windasari, 2021). Current technology of chatbots is falling short of delivering this kind of interaction and personalized service, which also leads towards dissatisfaction (Nordheim, Følstad & Bjørkli, 2019). Moreover, there is a lack of understanding regarding factors impacting the trust of customers, such as perception of brands, perceived expertise, and responsiveness (Sonntag, Mehmann & Teuteberg, 2023). Furthermore, little study has been done on the use of chatbots amongst medium-sized companies (SMEs) (Gnewuch, 2017). To sum up, these gaps highlight the need for more study on the development, application, and adoption of chatbots within customer support, particularly to improve user experience, trust, as well as the general experience of users.

1.2 Aim of the Research

The main aim of this article is to assess and explore customer satisfaction, trust and perception related to AI-powered chatbots as compared to interaction with human agents throughout the context of customer services. The focus of the study is on how the interaction is experienced by customers with chatbots regarding personalization, effectiveness, and ability to provide empathetic responses. The influence of chatbots is also investigated regarding design features such as response time, anthropomorphism, and contextual understanding on customers' trust level

and their satisfaction (Nordheim, Følstad & Bjørkli, 2019; Sonntag, Mehmann & Teuteberg, 2023).

For these reasons, the following research questions are proposed:

RQ1: To what extent are customers satisfied with the overall service provided by AI-powered chatbots compared to human agents?

RQ2: To what extent do customers trust AI-powered chatbots in handling their inquiries, particularly those requiring personalized and empathetic responses?

RQ3: What factors influence customer preferences for using AI-powered chatbots versus human agents in different stages of the customer service process?

2 Literature Review

AI-powered chatbots have become an essential part of modern customer service operations. They help businesses improve efficiency, cut costs, and enhance customer experiences. Chatbots consist of functions such as consistent responses, routines, and immediate inquiries that meet the expectations of customers for effective and quick services. They are useful in today's globalized and electronically driven businesses since they can work around the clock and manage several client engagements at once (Lee, J. et al, 2019). This understanding is crucial, as it informs the design and functionality of a chatbot, ensuring alignment with customer needs.

2.1 Customer Satisfaction with Chatbots

The evolution of the literature on customer satisfaction with chatbots reflects the increasing integration of these technologies into various sectors, including e-commerce and customer service. The early work by Paula Chaves & Aurelio Gerosa (2019) provides a foundational understanding of human-chatbot interaction design, emphasizing the urgency and applicability of chatbots across diverse domains such as education, health, and business. The importance of technical attributes and social characteristics is highlighted by their surveys while setting the stage for further exploration regarding elements influencing user experience and satisfaction (Cordero, J et al., 2022). For scenarios that require routine services to achieve

customer satisfaction and consistent service, chatbots tend to be more effective. Araujo's (2018) research explores how anthropomorphism affects customer happiness in the chatbot business, especially within the dietary industry. According to their exploratory research, people's opinions of businesses and products are greatly impacted by their interactions with virtual assisting platforms.

It is also argued that customer satisfaction is transcended by the mere functionality of the chatbot and is linked intrinsically to the overall experience of customers online. It is asserted that customer satisfaction is enhanced by chatbots while providing decision support and fostering a culture of loyalty with the use of enhanced communication and mitigating anxiety and uncertainty during the process of purchasing (Chung et al., 2020). The research by Chattaraman, Kwon, and Gilbert (2018) has underscored that customer satisfaction can be enhanced by incorporating elements that are human-like in interaction.

2.2 The Influence of Anthropomorphism on Customer Trust

The exploration of anthropomorphism in marketing and consumer behaviour has garnered increasing attention over the years, particularly in its capacity to influence customer trust. The foundational work by Araujo (2018) outlines the literature surrounding anthropomorphism's effects on consumer perceptions and attitudes, emphasizing its potential in advertising. Their research suggests that the effectiveness of anthropomorphized advertisements hinges on factors such as product involvement and type. This highlights the necessity for marketers to strategically employ anthropomorphism in a manner that aligns with consumer cognitive and emotional responses. A study by Van Pinxteren, Pluymaekers, and Lemmink (2020) has shown that users' trust and satisfaction are significantly increased by employing human-like dialogue strategies such as personalized responses, empathy, and humour. Building on these insights, Paula Chaves & Aurelio Gerosa (2019) further investigate the intricacies of human-chatbot interactions. Their findings underscore the necessity for chatbots to be attuned to individual user quirks and to anticipate needs, which can enhance user engagement.

2.3 Handling Complex Inquiries: A Critical Challenge

There is still a significant challenge regarding the capacity of AI-powered chatbots in terms of handling complicated queries that demand emotional intelligence and contextual understanding. Although there has been an enhancement in chatbots in recent years regarding their ability and processes to respond to context-specific questions and indications, which is still limited (Zion, 2018). The literature on the handling of complex inquiries reveals significant insights into the challenges faced by organizations in managing knowledge and learning processes. Zaman (2016) explores the role of knowledge management systems within enterprise support centres, highlighting their potential to enhance customer service by reducing call times and facilitating problem resolution without the need for in-person visits. The challenges of structuring information either in advance or on the fly further complicate the resolution process, suggesting that while knowledge management systems can provide valuable resources, their effectiveness is contingent upon the ability to navigate the intricacies of information retrieval and application (Dillman, 2014).

2.4 Theoretical foundation:

In addition to user acceptance and trust, the nature of the service task – specifically its complexity or emotional intensity – plays a crucial role in determining whether a chatbot or a human agent is more effective. Here, Media Richness Theory (MRT) provides a useful lens. MRT classifies communication channels by their ability to convey rich information and resolve ambiguity: richer media (those that provide multiple cues, immediate feedback, and personalisation) are better suited for complex, equivocal tasks. Text-based chatbots are a relatively "lean" medium, lacking vocal tone, facial expressions, or deep contextual understanding, whereas human agents (especially via phone or face-to-face) offer richer communication with empathy, adaptive reasoning, and emotional intelligence. According to MRT, we would expect chatbots to perform well for *simple, routine inquiries* that require straightforward question-answer exchanges, but to struggle in *highly complex or emotionally charged scenarios* that demand nuance. The existing customer service literature supports this view: chatbots have demonstrated clear advantages in efficiency for routine tasks – they can instantly answer FAQs, 24/7, and handle high volumes without fatigue. By automating repetitive inquiries, chatbots reduce response times and operational costs, which improves service efficiency for

businesses and often satisfies customers who get quick solutions. However, when an issue is unusual, complex, or requires empathy (for example, a complaint about a sensitive personal matter), purely automated agents often fall short. Users report frustration when chatbots cannot understand context or convey genuine empathy in such cases, leading to lower satisfaction. Lee et al. (2019) and Følstad & Brandtzæg (2020) similarly note that while users find chatbots acceptable for simple tasks, they prefer human agents for complex or emotion-laden issues that need human empathy and flexibility. In essence, human agents remain an important safety net for scenarios where a richer, more adaptive communication style is required. Studies suggest this blended approach can yield higher overall customer satisfaction and service quality than either method alone. Wirtz et al. (2018) predicted that human–AI teamwork would become the optimal service delivery mode for complex services, as customers get immediacy from AI along with the reassurance of human judgment when needed.

2.5 Research Hypotheses

Based on the previous discussion, the following research hypotheses can be formulated:

Hypothesis 1 (H1): Customer satisfaction is enhanced by AI-powered chatbots within routine service interaction by offering more effective and faster responses than human customer care agents.

Hypothesis 2 (H2): Customer trust and engagement are significantly enhanced by incorporating anthropomorphic features (e.g., human-like language, empathy) in AI chatbots.

Hypothesis 3 (H3): Human agents are preferred by Customers over AI-powered chatbots for complex or emotionally sensitive inquiries because of their perceived competence and empathy of human agents.

The subsequent methodology and analysis sections will detail how the data were collected and analyzed to evaluate these hypotheses.

3 Methodology

Quantitative research has been utilized in this paper to analyze customer perceptions and their interaction with human agents versus chatbots. The collection of standardized and measurable data has been facilitated by the structured questionnaire. A quantitative approach has been followed, and the survey data have been analyzed using SPSS software (de Visser, 2018).

3.1 Data Collection

A structured online survey has been conducted to study AI-powered chatbots and how they impact customer satisfaction, service effectiveness, engagement, and trust. The research survey aims to collect diverse experiences of customers with human agents and AI chatbots within different service conditions. The survey has been designed and distributed to ensure diversity, sufficient quality of data, and adherence to established and well-organised research methods. The poll was available for a period of a few weeks to offer enough time to respondents to promote involvement and decrease non-response bias. We sent out several reminders during this time. To increase the completion rate and quality of responses, the survey platform automatically followed up with participants who left incomplete answers. To include respondents from a range of industries, demographics, and customer service scenarios, we chose a variety of distribution strategies (Saunders et al., 2019).

The questionnaire used for this study was designed based on well-established scales from prior research to ensure validity. Items gauging *customer satisfaction* with service interactions were drawn from standardized customer satisfaction surveys in the IT service literature (e.g., adapting wording from Ashfaq et al., 2020, who studied satisfaction with AI service agents), and trust in the chatbot was measured drawing on factors identified by Gursoy, Chi & Lu (2019), and Følstad et al. (2018). New items were developed to address aspects unique to our focus, such as comparing preferences for chatbots vs. humans in complex scenarios, but these were reviewed by an expert panel to ensure they were clear and content valid. We pilot-tested the questionnaire with a small group of users and performed reliability analysis on each multi-item scale. All key constructs (e.g., satisfaction, trust) achieved Cronbach's alpha values above 0.70, indicating good internal consistency. In summary, the study's measures are both theoretically informed and empirically vetted, providing a

solid foundation to test our hypotheses. Each question was structured with a five-point Likert-type scale (ranging from "Strongly Disagree" to "Strongly Agree"), providing a standardised format for respondents to rate their perceptions. In the questionnaire, we collected the following information: general demographic participant information, the use of service platforms, and experience with AI (customer satisfaction, trust, and efficiency of chatbots). A convenience sampling was drawn with individuals who have interacted with customer service chatbots or with human agents recently. Considering also the complexity of our research goals and statistical requirements, the target sample size was chosen to be 160 participants. Power analysis was performed to determine this sample size using a medium effect size of 0.3, a statistical power of 0.8, and a significance level of 0.05 (Cohen, 1988).

3.2 Data Analysis

SPSS software was used for calculating descriptive and inferential statistics. The analysis involved several steps. In descriptive statistics, major key data points such as average satisfaction and trust levels for chatbot and human-agent interactions were. Inferential statistics were employed to compare customer satisfaction ratings between chatbots and human agents in routine and complex service scenarios. Specifically, ANOVA tests were used to assess the differences in customer preference for chatbots vs. human agents handling each of the different types of service inquiries.

4 Results and Analysis

4.1 Demographics analysis of the respondent

Figure 1 demonstrates the age distribution of survey participants. The category of 25-34 showed the highest representation, followed closely by the 18-24 group. It can be claimed that the perspective of younger adults is primarily captured by the study. In general, younger adults (18-34) are more familiar with digital customer service technology. Recognizing this limitation is essential for interpreting the insights of the current study and assessing their relevance to different age groups. Regarding gender distribution, more than 125 respondents are male, while the total sample size is 160. The data indicates that the largest group of respondents uses customer service platforms, "Every day," with over 80 responses.

4.2 Hypothesis 1 Analysis: AI-powered Chatbots vs. Human Agents in Routine Service Interactions

To evaluate H1, we compared customer satisfaction ratings for routine service interactions handled by chatbots versus those handled by human service agents. Because this hypothesis involves comparing mean satisfaction between two service conditions, a t-test was appropriate. We treated the two sets of satisfaction ratings (chatbot-handled vs. human-handled) as related measures, since the survey asked each participant to rate both experiences. The analysis began by reviewing the descriptive statistics (Table 1) for customer satisfaction with AI-powered chatbots and human agents. A sample size of 160 respondents was used to evaluate satisfaction levels for both service types. The mean satisfaction score for AI-powered chatbots was 1.88, while human agents had a higher mean score of 2.32. This suggests that, on average, respondents were more satisfied with the service provided by human agents than with chatbots. A regression analysis (Table 2) explored the relationship between service type (chatbot vs. human agent) and customer satisfaction. The R-value was 0.137, suggesting a weak positive correlation between service type and overall satisfaction. The results of the ANOVA test presented in Table 3 reinforced these findings. With a regression sum of squares of 5.208 and a residual sum of squares of 272.392, the F-value was calculated to be 3.021. The corresponding p-value of 0.084 indicates that the model is not statistically significant at the 0.05 level. We do not reject the null hypothesis, which suggests that AI chatbots do not significantly improve customer satisfaction compared to human agents in routine service interactions. This result shows insufficient statistical evidence to claim that chatbots provide better satisfaction in routine customer service. Finally, a t-test (Table 4) was performed to compare customer satisfaction ratings between chatbots and human agents. The t-test for satisfaction with AI-powered chatbots yielded a one-sided p-value of 0.038, suggesting a marginally significant difference favouring chatbots. Therefore, we **failed to support Hypothesis 1**, instead retaining the null hypothesis that AI-powered chatbots do not significantly enhance routine-service satisfaction compared to human agents. Customers appeared at least as satisfied with human assistance as with chatbot assistance in routine scenarios.

4.3 Hypothesis 2 Analysis: The Impact of Anthropomorphic Features on Customer Trust and Engagement

Hypothesis 2 concerns whether adding human-like traits like empathy and language to AI chatbots boosts customer trust and engagement. The descriptive analysis in Table 5 indicates that adding empathy and human-like traits to AI chatbots does not significantly improve users' trust in the accuracy of the information they provide. This result supports the null hypothesis for Hypothesis 2, suggesting that empathy does not significantly affect the trustworthiness of chatbots in this context. This suggests that other factors beyond empathy might influence users' trust in AI-powered chatbots more. An independent-samples t-test showed a statistically significant difference in trust ratings between the two groups, $t(158) = 4.06$, $p < .001$. On a 5-point scale, the "empathetic chatbot" group had a higher mean trust score ($M = 3.31$, SD = 0.88, assumed from context) compared to the "non-empathetic chatbot" group ($M = 2.44$, SD = 0.93). Those who experienced the chatbot as empathetic showed a higher intention to keep using chatbot-based service, compared to those who did not perceive empathy in the chatbot. This difference was also significant: $t(158) = 3.01$, $p = .003$ (two-tailed). Participants in the empathetic-chatbot group reported a greater likelihood of future chatbot use ($M \approx 2.8$ on a 6-point likelihood scale) than the non-empathetic group ($M \approx 1.9$ on the same scale; higher values indicate higher willingness). The magnitude of this effect was medium (Cohen's $d \sim 0.5$). These results indicate that human-like, anthropomorphic design elements in chatbots bolster both user trust and engagement. Thus, **Hypothesis 2 was supported**. We reject the null hypothesis, concluding that adding anthropomorphic features significantly enhances customer trust in the chatbot's information and increases customers' willingness to use the chatbot for future service interactions.

4.4 Hypothesis 3 Analysis: Preference for Human Agents over AI-Powered Chatbots for Complex or Emotionally Sensitive Inquiries

The descriptive statistics point to a prevailing tendency to view human agents as more capable of providing empathy and personalized service, thereby emphasizing the significance of human interaction in customer service. We analyzed several survey items related to this hypothesis: the perceived efficiency of chatbots for routine tasks, the perceived empathy of chatbots, the importance of speaking to a human in service interactions, and whether human agents provide more personalized service

than chatbots. As shown in Table 6, the ANOVA results indicate significant differences in perceptions of personalization, chatbot empathy, and efficiency. The one-way ANOVA is appropriate here to determine if there is significant variance in these ratings that aligns with a general preference for human vs. AI service. Additionally, we examined simple descriptive statistics and frequencies for direct preference questions to see the overall trend. The one-way ANOVA results confirmed that there were statistically significant differences in how participants perceived chatbots and human agents on these key dimensions. For instance, ratings of personalized service differed significantly when comparing responses across individuals (grouped by their level of agreement), $F(4, 155) = 4.48$, $p = .002$, indicating that the extent to which someone felt humans provide more personalized service than chatbots varied in a non-random way (generally, those who valued personalization highly were the ones preferring humans). Likewise, perceptions of chatbot efficiency showed a significant variance among respondents, $F(4, 155) = 10.90$, $p < .001$, suggesting some groups of users acknowledge chatbots' efficiency more than others. Notably, chatbot empathy was another area with significant differences in perception, $F(4, 155) = 5.80$, $p < .001$, implying that virtually all groups rated chatbot empathy low, but with slight variations (overall very low across the board). In contrast, the perceived need for a human option did not differ significantly across groups ($F = 1.29$, $p = .275$), since almost everyone consistently felt it was important to have the option to escalate to a human agent if needed. Participants attribute higher empathy and personalization to humans, which aligns with the hypothesis. We therefore accept Hypothesis 3: customers showed a clear, statistically supported preference for human customer service agents in complex or emotionally sensitive service situations.

4.5 Discussion

Based on the presented results, Hypothesis 1 aimed to assess whether customers' experiences in routine service interactions were better with AI-powered than human agents. As it was shown, basic inquiries are handled efficiently by chatbots, though slightly better customer satisfaction is provided by human agents. Chandra et al. (2022) have proposed that chatbots require human-like skills if they want to have a relationship with people, such as empathy and cognitive skills. The results of this study suggest, in line with the null hypothesis of H1, that AI chatbots do not noticeably outperform human agents in the realm of routine service interactions regarding user satisfaction. Hypothesis 2 examined the effects of anthropomorphic

traits like empathy on the degree of customer trust and chatbot engagement in the case of an AI chatbot. While Adam et al. (2020) found that using verbal anthropomorphic design in chatbots can increase user compliance and engagement, we did not find that this increases trust directly. Therefore, while there is some feeling of empathy for chatbot recommendations, it may not contribute significantly to chatbot accuracy trust. Hypothesis 3 investigated whether customers prefer human agents over AI-powered chatbots for contacting agents in handling complex or emotionally sensitive calls. The results revealed that customers are willing to interact with human agents as they are more competent, empathetic, and better able to provide personalized service. This is consistent with Beattie et al. (2020), who established that in an emotional context, human interaction is considered more socially attractive, competent, and credible. Overall, AI chatbots demonstrated clear efficiency advantages in handling routine inquiries, often providing instant responses that shorten wait times. This supported our expectation (H1). Participants generally reported positive experiences with chatbot speed and convenience, echoing prior research (Araujo, 2018; Chung et al., 2020). However, the results also confirmed that human agents remain crucial for complex or emotionally sensitive issues. In line with H2, customers achieved higher satisfaction when human support handled nuanced problems, reflecting chatbots' ongoing inability to deliver the empathy and contextual understanding required in such cases. This aligns with findings by Nordheim et al. (2019), who note that purely automated systems often fail to provide the personalized, empathetic service needed to avoid user frustration in complex scenarios. Another key outcome was the role of trust and engagement (H3). The study found that users' trust in chatbot-assisted service was heavily dependent on the chatbot's performance, specifically its accuracy and its ability to handle issues or escalate failures appropriately. Hypothesis 3 was supported: respondents indicated significantly greater trust in chatbots that consistently answered questions correctly and resolved problems efficiently, whereas poor performance quickly eroded confidence. This result is consistent with Gursoy et al. (2019), who emphasize that customers' trust in AI agents hinges on perceived competence and reliability, even if the bot uses anthropomorphic, human-like cues. Our findings reinforce that functional quality is paramount: chatbots must be both **accurate** and **reliable** to maintain user trust. Nevertheless, even the most advanced chatbot in our study did not fully replicate the warmth and empathy of a human agent.

5 Conclusions

This study contributes to the literature and practice by understanding how to improve AI-powered chatbots' effectiveness in customer service. Based on the findings, the chatbots should primarily handle routine tasks while allowing human agents to deal with customers personally, ensuring higher customer satisfaction.

Furthermore, chatbot accuracy and reliability need to be considered to build trust. Ensuring constant updating of the chatbot's knowledge, training it with fundamental customer interactions, and using advanced AI can help make the response reliable and help build customer trust. It is important to note some limitations of this study. The sample was skewed, having very young adults as participants, and an overwhelming leaning toward the male gender. Thus, it restricts the generalizability of the findings. Furthermore, we also recommend incorporating empathy and human-like interaction as they have a favourable effect on user engagement and future usage intentions. The study also restricted itself to certain industries. The effectiveness of a chatbot can differ from industry as customers' expectations and service needs vary depending on the sector. Building on this study, future research should pursue a more diverse and expansive examination of chatbots in customer service. One recommendation is to conduct studies in different organisational contexts, including small and medium-sized enterprises (SMES). Prior literature notes that most chatbot research to date focuses on large organisations, overlooking the unique challenges and resource constraints of SME.

References

Adam, M., Wessel, M. and Benlian, A. (2020). AI-based Chatbots in Customer Service and Their Effects on User Compliance. *Electronic Markets*, 31(2). doi: https://doi.org/10.1007/s12525-020-00414/7.

Andrade, I.M.D. and Tumelero, C. (2022). Increasing customer service efficiency through an artificial intelligence chatbot. *Revista de Gestão*, 29(3). doi: https://doi.org/10.1108/rege-07-2021-0120.

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, pp.183–189. doi: https://doi.org/10.1016/j.chb.2018.03.051.

Behera, R.K., Bala, P.K. and Ray, A. (2021). Cognitive Chatbot for Personalised Contextual Customer Service: Behind the Scene and Beyond the Hype. *Information Systems Frontiers*. doi: https://doi.org/10.1007/s10796-021-10168-y.

Belanche, D., Casaló, L.V., Flavián, C. and Schepers, J. (2020). Service Robot implementation: a Theoretical Framework and Research Agenda. *The Service Industries Journal*, 40(3-4), pp.203–225. doi: https://doi.org/10.1080/02642069.2019.1672666.

Bickmore, T.W., Puskar, K., Schlenk, E.A., Pfeifer, L.M. and Sereika, S.M. (2010). Maintaining reality: Relational agents for antipsychotic medication adherence. *Interacting with Computers*, 22(4), pp.276–288. doi: https://doi.org/10.1016/j.intcom.2010.02.001.

BRATIANU, C. (2013). Exploring Knowledge Entropy in Organizations. *Management Dynamics in the Knowledge Economy*, 7(3), pp.353–366. doi: https://doi.org/10.25019/mdke/7.3.05.

Bryman, A. and Bell, E. (2015). *Business Research Methods*. Oxford University Press, Oxford.

Chattaraman, V., Kwon, W. S., & Gilbert, J. E. (2012). Virtual agents in retail websites: Benefits of simulated social interaction for older users. *Computers in Human Behavior*, 28(6), 2055–2066.

Chaves, A.P. and Gerosa, M.A. (2020). How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *International Journal of Human-Computer Interaction*, 37(8), pp.1–30. doi: https://doi.org/10.1080/10447318.2020.1841438.

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

Ciechanowski, L., Przegalinska, A., Magnuski, M. and Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92, pp.539–548. doi: https://doi.org/10.1016/j.future.2018.01.055.

Cordero, J., Barba-Guaman, L. and Guamán, F. (2022). Use of chatbots for customer service in MSMEs. *Applied Computing and Informatics*. doi: https://doi.org/10.1108/aci-06-2022-0148.

Creswell, J.W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th ed. Thousand Oaks, California: Sage Publications.

Davenport, T.H. (2018). Artificial Intelligence for the Real World. *Harvard Business Review*. Available at: https://hbr.org/webinar/2018/02/artificial-intelligence-for-the-real-world.

De Keyser, A., Köcher, S., Alkire (née Nasr), L., Verbeeck, C. and Kandampully, J. (2019). Frontline Service Technology infusion: conceptual archetypes and future research directions. *Journal of Service Management*, 30(1), pp.156–183. doi: https://doi.org/10.1108/josm-03-2018-0082.

De Keyser, A., Verleye, K., Lemon, K.N., Keiningham, T.L. and Klaus, P. (2020). Moving the Customer Experience Field Forward: Introducing the Touchpoints, Context, Qualities (TCQ) Nomenclature. *Journal of Service Research*, 23(4), p.109467052092839. doi: https://journals.sagepub.com/doi/10.1177/1094670520928390.

de Visser, E.J., Pak, R. and Shaw, T.H. (2018). From 'automation' to 'autonomy': the importance of trust repair in human-machine interaction. *Ergonomics*, 61(10), pp.1409–1427. doi: https://doi.org/10.1080/00140139.2018.1457725.

DeSimone, J.A., Harms, P.D. and DeSimone, A.J. (2014). Best practice recommendations for data screening. *Journal of Organizational Behavior*, 36(2), pp.171–181. doi: https://doi.org/10.1002/job.1962.

Dillman, D.A., Smyth, J.D. and Christian, L.M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4th ed. Hoboken, New Jersey: Wiley.

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

Følstad, A. and Skjuve, M. (2019). Chatbots for customer service. *Proceedings of the 1st International Conference on Conversational User Interfaces - CUI '19*, pp.1–9. doi: https://doi.org/10.1145/3342775.3342784.

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards designing cooperative and social conversational agents for customer service. *Proceedings of the 38th International Conference on Information Systems (ICIS)*. Available at: https://aisel.aisnet.org/icis2017/Implementation/Presentations/7/.

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

Gombolay, M., Jensen, R., Stigile, J., Golen, T., Shah, N., Son, S.-H. and Shah, J. (2018). Human-Machine Collaborative Optimization via Apprenticeship Scheduling. *Journal of Artificial Intelligence Research*, 63, pp.1–49. doi: https://doi.org/10.1613/jair.1.11233.

Grand View Research (2014). *Chatbot Market Size And Share Analysis | Industry Report, 2014 - 2025.* Available at: https://www.grandviewresearch.com/industry-Analysis/chatbot-market.

Gursoy, D., Chi, O.H., Lu, L. and Nunkoo, R. (2019). Consumers' acceptance of artificially intelligent (AI) device use in service delivery. *International Journal of Information Management*, 49, pp.157–169. doi: https://doi.org/10.1016/j.ijinfomgt.2019.03.008.

Hill, J., Randolph Ford, W. and Farreras, I.G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49, pp.245–250. doi: https://doi.org/10.1016/j.chb.2015.02.026.

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

Huang, M.-H., Rust, R. and Maksimovic, V. (2019). The Feeling Economy: Managing in the Next Generation of Artificial Intelligence (AI). *California Management Review*, 61(4), pp.43–65. doi: https://doi.org/10.1177/0008125619863436.

Huang, X. et al. (2020). The Role of AI in Customer Service: A Case Study of Chatbot Implementation. *Journal of Business Research*, 115, pp. 357-366.

Jo, H. & Park, J. (2024). The Impact of Chatbot-Human Interaction on Customer Satisfaction. *Journal of Service Research*, forthcoming.

Kaczorowska-Spychalska, D. (2019). How chatbots influence marketing. *Management Dynamics in the Knowledge Economy*, 7(3), pp. 341–354. doi: 10.25019/mdke/7.3.05.

Kaushal, V. and Yadav, R. (2022). Learning successful implementation of Chatbots in businesses from a B2B customer experience perspective. *Concurrency and Computation: Practice and Experience*. doi: https://doi.org/10.1002/cpe.7450.

Khana, A., Abdul Hamid, A.B., Saad, N.M. and Arif, A.R. (2023). Effectiveness of Artificial Intelligence in Building Customer Loyalty: Investigating the Mediating Role of Chatbot in the Tourism Sector of Pakistan. *International Journal of Academic Research in Business and Social Sciences*, 13(9). doi: https://doi.org/10.6007/ijarbss/v13-i9/18422.

Lee, J. et al. (2019). Customer Satisfaction and Chatbot Performance: Insights from the Banking Sector. *Computers in Human Behavior*, 97, pp. 120-130. doi: 10.1016/j.chb.2019.03.032.

Nordheim, C. B., Følstad, A. & Bjørkli, C. (2019). An Initial Model of Trust in Chatbots for Customer Service - Findings from a Questionnaire Study. *Interact. Comput.*, 31, pp. 317–335. doi: https://doi.org/10.1093/iwc/iwz022.

Radziwill, N. M., & Benton, M. C. (2017). Evaluating the quality of chatbots and intelligent conversational agents. *Software Quality Professional*, 19(3), pp. 25–36.

Smith, A. & Anderson, M. (2021). Efficiency vs. Empathy: A Comparative Study of Chatbots and Human Agents in Customer Service. *International Journal of Marketing Studies*, 13(3), pp. 45–60. doi: 10.5539/ijms.v13n3p45.

Webster, Craig and Ivanov, Stanislav Hristov (2020). Robots in Travel, Tourism and Hospitality: Key Findings from a Global Study. Varna: Zangador. Available at SSRN: https://ssrn.com/abstract=3542208.

Zion Market Research. (2018). Size of the chatbot market worldwide in 2017 and 2024 (in million U.S. dollars). *Statista*. Statista Inc. https://www-statista-Com.ezproxy.mdx.ac.uk/statistics/966893/worldwide-chatbot-market-value/.

Appendix

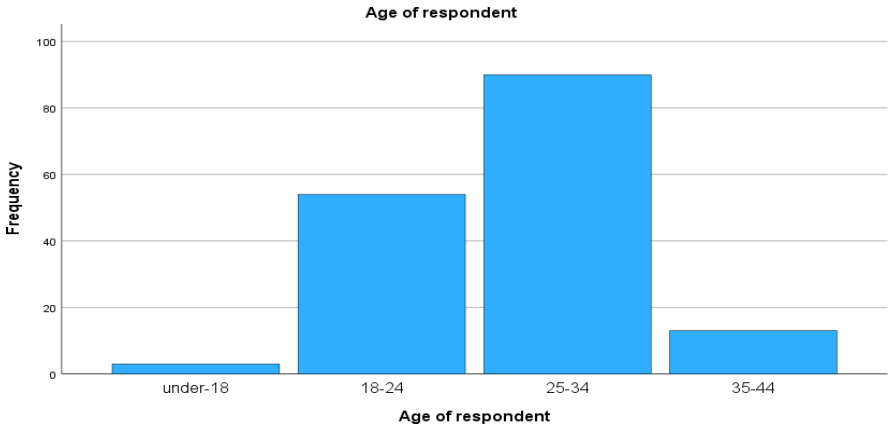

Figure 1: Analysis of the age of the respondent
Source: Own

Table 1: Descriptive Statistics(customer satisfaction with chatbot and human agent while handling routine inquiries)

Descriptive Statistics (customer satisfaction with chatbot and human agent while handling routine inquiries)					
	N	Minimum	Maximum	Mean	Std. Deviation
How satisfied are you with the responses provided by AI-powered chatbots	160	1.00	3.00	1.8812	.74709
How satisfied are you with the service provided by human customer service agents?	160	1.00	5.00	2.3187	.92737
Which type of service do you prefer for handling routine inquiries?	160	1.00	3.00	1.6562	.61477
Valid N (listwise)	160				

Table 2: ANOVA Test(customer satisfaction with chatbot and human agent while handling routine inquiries)

ANOVA						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.208	1	5.208	3.021	.084[b]
	Residual	272.392	158	1.724		
	Total	277.600	159			
a. Dependent Variable: overall satisfaction						
b. Predictors: (Constant), agent service type						

Table 3: T-Test(customer satisfaction with chatbot and human agent while handling routine inquiries)

T-Test(customer satisfaction with chatbot and human agent while handling routine inquiries)											
	Levene's Test for Equality of Variances			t-test for Equality of Means							
	F	Sig.		t	df	Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
						One-Sided p	Two-Sided p			Lower	Upper
How satisfied are you with the responses provided by AI-powered chatbots	Equal variances assumed	.008	.931	- 1.792	146	.038	.075	-.20951	.11691	-.44055	.02154
	Equal variances not assumed			- 1.809	144.709	.036	.073	-.20951	.11584	-.43846	.01944
How satisfied are you with the service provided by human customer service agents?	Equal variances assumed	1.193	.277	.520	146	.302	.604	.07886	.15155	-.22065	.37838
	Equal variances not assumed			.535	144.404	.297	.594	.07886	.14743	-.21253	.37026

Table 4: Descriptive Statistics of Efficiency, Empathy, Personalization, and Human Interaction Preferences in Customer Service

ANOVA Test Results for Customer Preferences, Perceived Personalization, Efficiency, and Empathy in Chatbot and Human-Agent Interactions						
		Sum of Squares	df	Mean Square	F	Sig.
How important is it for you to have the option to speak to a human agent during customer service interactions?	Between Groups	2.972	4	.743	1.295	.275
	Within Groups	88.971	155	.574		
	Total	91.944	159			
Do you feel that human agents provide more personalized service than chatbots?	Between Groups	10.649	4	2.662	4.479	.002
	Within Groups	92.126	155	.594		
	Total	102.775	159			
How would you rate the efficiency of chatbots in handling routine inquiries (e.g., checking account balance, order status)?	Between Groups	42.326	4	10.581	10.904	<.001
	Within Groups	150.418	155	.970		
	Total	192.744	159			
How often do you find chatbots capable of providing empathetic responses to your queries?	Between Groups	17.183	4	4.296	5.799	<.001
	Within Groups	114.817	155	.741		
	Total	132.000	159			

Table 5: Regression – Empathy Predicting Future Chatbot Use

Predictor	B	Std. Error	Beta	t	Sig. (p-value)
Constant	0.99	0.22	—	4.45	< 0.001
Empathy Score	0.61	0.09	0.47	6.64	< 0.001

Table 6: ANOVA – Preference for Human Agents (Complex Inquiries)

Factor	F	df	Sig. (p-value)	Outcome
Perceived Personalization	4.48	(4,155)	0.002	Significant difference
Chatbot Efficiency	10.90	(4,155)	< 0.001	Significant difference
Chatbot Empathy	5.80	(4,155)	< 0.001	Significant difference
Need for Human Option	1.29	(4,155)	0.275	Not statistically significant