Customer Experience Unveiled: A Neuroscientific Exploration of AI-powered Chatbots in Online Retail

Mike Trynczyk
Hochschule Düsseldorf
Ellen Roemer
Hochschule Ruhr West
Carina Eisel-Ende
Hochschule Ruhr West

Cite as:
Trynczyk Mike, Roemer Ellen, Eisel-Ende Carina (2024), Customer Experience Unveiled: A Neuroscientific Exploration of AI-powered Chatbots in Online Retail. *Proceedings of the European Marketing Academy*, (122577)

Paper from EMAC Regional Conference, Lisbon, Portugal, September 25-27, 2024

Customer Experience Unveiled: A Neuroscientific Exploration of AI-powered Chatbots in Online Retail

This study examines the impact of AI-powered chatbots on customer experience in online retailing environments. Drawing on neuroscientific methods, including eye-tracking and galvanic skin response (GSR), we assess the dimensions of customer experience alongside self-reported survey measures. Our findings reveal important insights into the influence of chatbots on customer experience, contrasting experiences with and without chatbot intervention. Our approach contributes to a comprehensive understanding of customer experience. Our research underscores the importance of AI-powered chatbots in shaping customer experiences in online retailing.

Keywords: *artificial intelligence (AI), chatbot, customer experience*

1 Introduction

The integration of artificial intelligence (AI) in today's businesses offers a high potential to generate added value for customers (Hoyer, Kroschke, Schmitt, Kraume, & Shankar, 2020). One of the key applications of AI in online retail are chatbots, which enable direct, two-way interaction between retailers and customers (Bawack, Wamba, Carillo, & Akter, 2022; Jiang, Cheng, Yang, & Gao, 2022; Shumanov & Johnson, 2021). Especially at a time when many countries are experiencing labour shortages and there are limited staff available to take care of customers in customer service centres, AI-powered chatbots represent a promising way out of the dilemma (Jiang et al., 2022).

However, online retailers must ensure that AI-powered chatbots add value to the customer experience and avoid negative impacts (Hoyer et al., 2020; Lou, Kang, & Tse, 2022; Luo, Tong, Fang, & Qu, 2019). Therefore, research is needed to investigate consumers' experiences with AI-powered chatbots. Much of this research has focused on satisfaction (Jiang et al., 2022), customer engagement (Jiang et al., 2022), trust (Cheng, Bao, Zarifis, Gong, & Mou, 2022), and patronage intentions (Lou et al., 2022). Only few studies, have investigated the genuine customer experience with chatbots in online shopping environments (exceptions relate to educational contexts, e.g., Liao & Yan, 2023, Bubaš, Babić, & Čižmešija, 2023). Therefore, we formulate our research question as follows: How do AI-powered chatbots influence customer experience in online shopping environments?

We use a neuroscientific approach using eye-tracking to evaluate customer's cognition as well as galvanic skin response (GSR) to assess emotions. In this line, we follow Lemon and Verhoef (2016, p. 88) who state that "these [neuroscientific] approaches will soon complement attitudinal survey measurements and provide new insights into the factors that influence the customer experience and how they are linked to customer behaviors." In addition, we used a survey to include self-reported measures of cognitive, affective and social experience and to interpret our data.

Therefore, our contribution is twofold. First of all, we assess the impact of an AI-powered chatbot on customer experience by comparing the experience with and without the use of a chatbot. In this way, we provide new insights into the consequences of the use of such a tool on the customer side. Second, we use a neuroscientific approach in order to provide a comprehensive understanding of customer experience. In particular, we use eye-tracking to assess the cognitive component of customer experience and galvanic skin response (GSR) to measure the affective component of customer experience in addition to a survey instrument with traditional measurement scales. Prior studies have only focused on specific components of customer experience, such as the affective component (Verhulst et al., 2020).

Our paper is structured as follows. In the next section, we use the extant literature to derive our hypotheses and our model. Thereafter, we describe the research design, measurements, data collection, data cleaning and the sample. In Section 4, we present our research results before we delineate limitations as well as areas for future research.

2 Theoretical Foundations – The Impact of AI-powered Chatbots on Customer Experience

2.1 Customer experience in offline and online retail

Verhoef et al. (2009, p. 32) defined that "the customer experience construct is holistic in nature and involves the customer's cognitive, affective, emotional, social and physical responses to the retailer." Much of the literature has focused on the analysis and measurement of customer experience in physical shopping environments (Bagdare & Jain, 2013; Bustamante & Rubio,

2017; Lahmeyer & Roemer, 2024; Schmitt, 1999) or multichannel environments (Lemon & Verhoef, 2016).

In online shopping environments, the cognitive and affective components of the customer experience have received increasing research attention (Barari, Ross, & Surachartkumtonkun, 2020). Given that our research topic focuses on chatbots, we also include the social component of customer experience due to the interactive nature of chatbots (see also Hoyer et al., 2020).

2.2 Cognitive experience

AI-powered chatbots offer the opportunity to provide informative and personalised interactions with customers. They are able to answer individual questions, provide specific information and, consequently, support customers' decision-making. In this way, they offer cognitive added value (Hoyer et al., 2020). Therefore, we can assume that visitors of an online shop, who interact with a chatbot, are more active participants in the exchange of information than visitors without chatbot interaction. This will be reflected in more intensive cognitive reactions and experiences (Lou et al., 2022; Xu, Shieh, van Esch, & Ling, 2020). Consequently, we posit that:

H_1: Visitors of an online shop, who interact with a chatbot, have a higher level of cognitive experience compared to visitors without chatbot interactions.

2.3 Affective experience

Empathy, understanding and the ability to create a positive emotional connection with the customer are key characteristics of chatbots. They are user-friendly tools that can provide a feeling of individual support and attention (Bleier, Harmeling, & Palmatier, 2019; Bogula, 2022). Individual interactions can have a positive impact on the customer's emotional experience. They may also prevent from dissatisfaction by processing enquiries faster and more effectively than the customer can perform or with the help of others (Bogula, 2022; Hoyer et al., 2020; Jenneboer, Herrando, & Constantinides, 2022). This leads to our second research hypothesis:

H_2: Visitors of an online shop, who interact with a chatbot, have a higher level of affective experience compared to visitors without chatbot interactions.

2.4 Social experience

AI-powered chatbots are able to simulate human-like interactions and give customers a social feeling to a service employee. A dialogue between humans and computers can be created by responding to direct questions, asking queries and opening up new topics on their own initiative (Bogula, 2022; Hoyer et al., 2020). This can cause a feeling of human contact, warmth and sensitivity in customers (Bleier et al., 2019; Gefen & Straub, 2003). Therefore, we propose that visitors of an online shop who interact with a chatbot during the customer journey on the website develop a higher level of social awareness than visitors without such an interaction.

H_3: Visitors of an online shop, who interact with a chatbot, have a higher level of social experience compared to visitors without chatbot interactions.

Figure 1 provides a model with the above-mentioned hypotheses that we will submit to empirical testing using neuroscientific measurement as well as self-reported measurements using a survey.

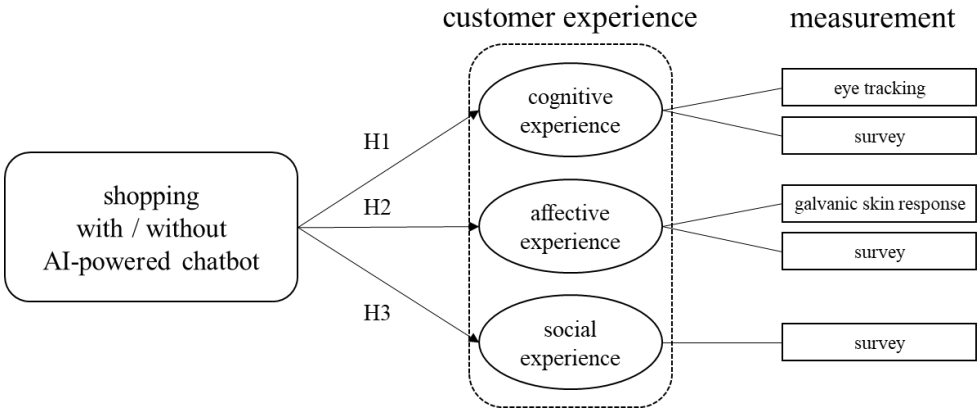

Figure 1: Model with measurement approaches

3 Methodology

3.1 Measurement

In order to assess how an AI-powered chatbot affects the customer's experience, we measured its cognitive, affective and social components with the help of an experimental design. We measured the three components of customer experience while participants were using an AI-powered chatbot on a website and compared them to the measurements without the use of the chatbot on the same website. The groups differed only in the independent variable: In the first group, an AI-powered chatbot was used to complete the task. In contrast, the control group solved the task without using the chatbot.

We conducted an online survey as self-report to assess all three component of customer experience. Inspired by Schmitt's strategic experiential modules (Schmitt, 1999), we adapted the items for the questionnaire measuring cognitive and affective component from Bustamante and Rubio (2017) to an online environment. The items for assessing social experience were adapted from Bleier et al. (2019). All items were measured on 7-point Likert-scales.

To complement our insights from the self-report measurement, we used neuroscientific approaches "that enable more precise, in-the-moment measures of customer experience" (Lemon & Verhoef, 2016, p. 88). To gain deeper insights into the cognitive component, we used an eye tracking instrument (Bojko, 2013). To measure the affective component in addition to the survey instrument, we used galvanic skin response (GSR) measurements to determine the level of emotional arousal (Caruelle, Shams, Gustafsson, & Lervik-Olsen, 2024; Christopoulos, Uy, & Yap, 2019; Verhulst et al., 2020). The aim of our experiment is to provide a comprehensive picture of customer experience (Clore, Wyer, Jr. Robert S., Dienes, Gasper, & Isbell, 2001; Lemon & Verhoef, 2016). The measurement instruments are summarized in Figure 1.

3.2 Research design

The experiment was conducted in a laboratory at a European University. The experiment was designed as a between-subject design to avoid carryover effects (Charness, Gneezy, & Kuhn, 2012). We ran pilot tests with a selection of two participants per group to ensure that the test

procedure and the instruments for measuring the neurophysiological and eye-tracking activities functioned properly (Verhulst et al., 2020).

In the main study, the test subjects were randomly assigned to one of the two groups. The test setup, procedure and tasks were identical in both groups in order to ensure internal validity. The number of test subjects per group was identical for comparability reasons (Charness et al., 2012; Verhulst et al., 2020). The gender distribution in both groups was the same, as women and men differ in certain biological reactions and gender-specific differences can be minimised in this way (Bianchin & Angrilli, 2012). The mean age was comparable in both groups.

Figure 2 illustrates the test procedure. After a participant entered the room, the principal researcher welcomed the participants, the participant signed an informed consent form and sat down on a chair opposite to a screen with eye-tracking sensors (Tobii Pro T60 XL). The principal researcher then gave a brief introduction to the procedure and the special features. The introduction was identical for all test participants. The sensors for the skin resistance measurement (Shimmer3 GSR+) were attached and the eye-tracker was calibrated. After a baseline measurement (participants were informed on the screen about the baseline measurement and asked to remain still for 15 seconds), which is necessary for the skin response measurement, the actual experiment was started.

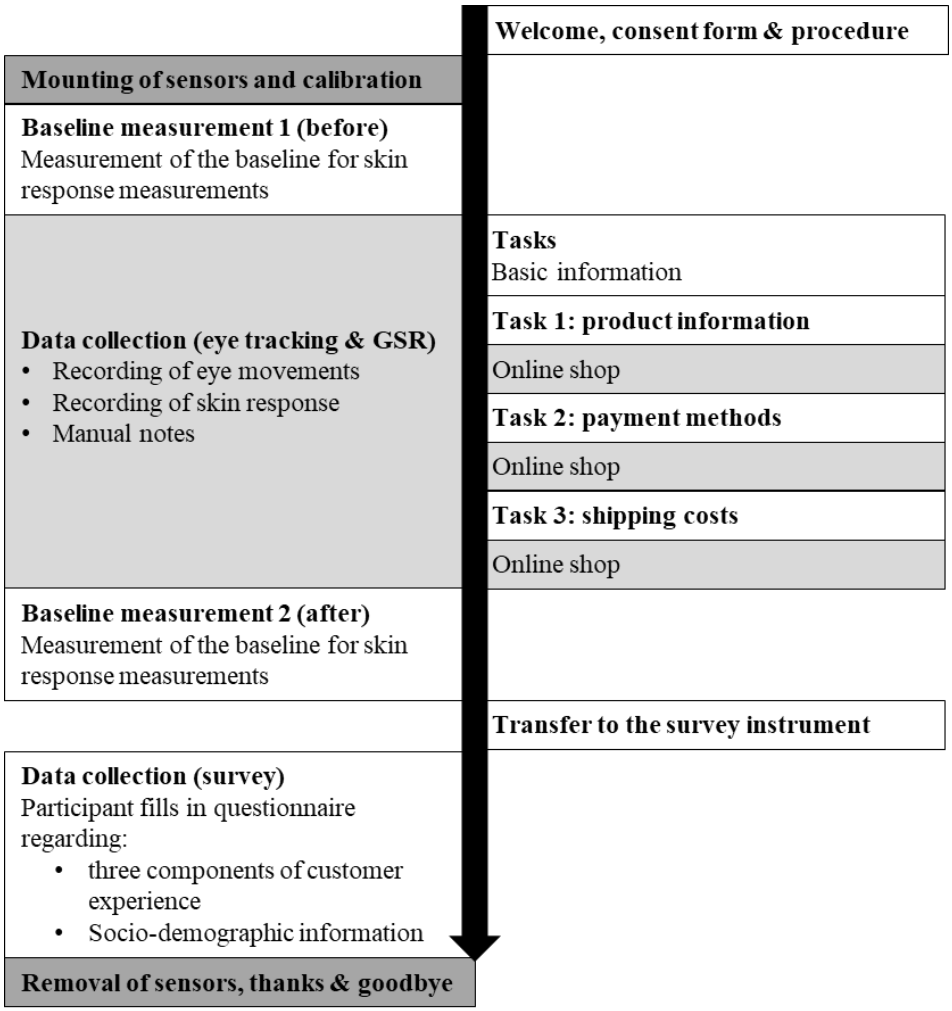

Figure 2: Experimental procedure (adapted from Verhulst et al., 2020)

Participants in both groups were given identical tasks in an online shop of an online pharmacy. Participants had to conduct three tasks at different phases of the customer journey, i.e., (1) finding specific product information, (2) researching payment methods, (3) finding shipping

costs. There was no time limit. The different tasks were chosen to enable the participants to interact with the chatbot at different stages of the customer journey and for different purposes (Bleier et al., 2019; Lemon & Verhoef, 2016). The tasks could be solved both with the chatbot and without the chatbot, i.e., solely relying on the information in the online shop.

After completion of the three tasks, a second baseline measurement was conducted and the participant was transferred to a website for the online questionnaire for the survey. Participants filled out a questionnaire, in which they retrospectively reflected and reported their subjective perception of all three components of customer experience. After completing the questionnaire, the principal researcher removed the sensors, thanked the participant and wished goodbye.

3.3 Data collection, data cleaning and sample

The data was recorded using Tobii Pro Lab and SosciSurvey. We used IBM SPSS Version 29 to further process and analyse the data. A total of 37 test subjects took part in the experiment. They were recruited among family and friends as well as through targeted recruitment at the university. The study was also announced using posters in the buildings and posts in the university's online forums.

The records were checked manually for completeness. Seven results had to be discarded due to missing GSR data. This was due to connection problems between the sensors and the computer during recording. These difficulties resulted in GSR measurement data being partially or completely missing. The questionnaire was answered in full by every participant. In addition, the eye-tracking recordings were successful for every test subject.

The final sample therefore consists of 30 valid cases. Of these, 15 test subjects used the chatbot to complete the tasks (WITH) and 15 other participants belong to the control group and completed the tasks without the chatbot (WITHOUT). Participants were assigned to one or the other group.

In the subsequent section, we will discuss the results. First of all, the results of eye tracking (cognitive experience) and affective experience measured by GSR are presented. Thereafter, we discuss the results from the questionnaire.

4 Results

4.1 Cognitive experience – eye tracking measurement

To measure the cognitive component of customer experience, we used the number of fixations and gaze duration from the eye tracking recordings. Table 1 shows the descriptive statistics.

	WITH			WITHOUT			
Eye tracking metric	**M**	**SD**	**Mean rank**	**M**	**SD**	**Mean rank**	**sig.**
Number of fixations	489.53	166.817	16.23	453.00	177.700	14.77	0.653
Gaze duration (in ms)	252.87	27.720	18.07	230.33	32.629	12.93	0.116

Table 1: Eye tracking metrics with and without chatbot use

Table 1 shows that means and mean ranks of all key eye tracking metrics are higher for the group that used the chatbot. To test for the significance in the differences between the mean ranks of the two groups, the Mann-Whitney-U-test was conducted for each metric. However, the results revealed that there were no significant differences between the groups.

4.2 Affective experience – skin response measurement (GSR)

Throughout the experiment, the skin response values were lower in the chatbot group relative to the control group that did not use the chatbot. Throughout the experiment, the values within both groups exhibited a persistent upward trend. However, the rise compared to the baseline value before is higher in the group with chatbot interaction compared to the control group during and after completion. The results are summarized in Table 2.

	WITH	**WITHOUT**
Baseline value (pre)	2.16	2.59
Task solving	2.77 *(+28.1%)*	3.15 *(+21.4%)*
Baseline value (post)	2.92 *(+34.9%)*	3.43 *(+32.4%)*

Table 2: Galvanic skin response in microsiemens (µS) with and without chatbot use

To test whether there are significant differences between the skin responses of the group with and without chatbot use, i.e., whether an AI-powered chatbot has an impact on the participants' affective component of customer experience, we conducted a one-way analysis of covariance (ANCOVA) with the variable chatbot use (with / without) as the independent variable, the average GSR the baseline after the tasks as the dependent variable, and one covariate, i.e., the average GSR of the baseline before the intervention (pre). The results showed that there were no significant differences between the two groups regarding the skin response measurements, i.e., the arousal (emotional component) did not differ significantly with the use of the chatbot relative to the use of the website without the chatbot.

4.3 All components – self-report measurement

All items for the three components of customer experience scored higher in the group that used the chatbot to accomplish the three tasks compared to the group that refrained from using the chatbot. Table 1 provides an overview of the descriptive statistics and the results of the Mann-Whitney-U-Test to test the differences between the two groups. We used a non-parametric test to test for the differences between the groups due to the small sample sizes.

	WITH			**WITHOUT**			
Items	M	SD	Mean rank	M	SD	Mean rank	sig.
Affective 1	5.07	1.033	19.00	4.20	0.862	12.00	0.029
Affective 2	5.27	1.033	21.03	3.40	1.242	09.97	0.000
Affective 3	5.53	0.834	20.23	4.07	1.335	10.77	0.002
Cognitive 1	4.20	1.740	17.40	3.47	1.506	13.60	0.250
Cognitive 2	4.20	1.821	19.10	2.67	1.397	11.90	0.023
Cognitive 3	4.27	1.792	19.40	2.67	1.345	11.60	0.015
Social 1	4.47	1.959	20.20	2.27	0.961	10.80	0.003
Social 2	3.73	1.751	19.83	1.93	0.884	11.17	0.006
Social 3	3.60	1.844	19.63	1.80	1.014	11.37	0.009

Table 3: Descriptive statistics of the items measuring customer experience

The results comparing the descriptive statistics of the groups with chatbot and without chatbot use (Table 3) show that all means and mean ranks are higher for the group with chatbot use. Table 3 shows the two-sided exact differences since sample sizes are small. This reveals a

more intensive customer experience in all three dimensions of the customer experience as self-reported by the participants. The majority of the mean ranks significantly differ from each other. Conducting principal component analysis and using the elbow criterion suggests three factors: affective, cognitive and social. However, only the affective and social component differed significantly from each other, with the chatbot group yielding higher values. The results are shown in Table 4.

	Mean ranks		
Component	**WITH**	**WITHOUT**	**sig.**
Affective component	19.60	11.40	0.010
Cognitive component	17.20	13.80	0.305
Social component	19.33	11.67	0.016

Table 4: Comparison of mean ranks of the principal components of customer experience

5 Limitations and Future Research

The use of neuroscience methods has the potential to reveal new insights into customer experience. We assessed the affective component using GSR-measurement and the cognitive component using eye tracking. Even though we discovered higher arousal and higher eye tracking measures in the group using the chatbot, the differences were not significant. This may be due to the small sample size. Therefore, further research is needed looking into the dynamics of customer experience (Verhulst et al., 2020) with larger sample sizes and representative sampling procedures. A more detailed analysis of customer's active engagement with the areas of interests on the website and the interaction phases with the chatbot could be worthwhile.

Based on the self-report measures, our study reveals that the affective and social components of customer experience are larger in the group that used the chatbot. Due to the interactive nature of the chatbot, customers are more emotionally and socially attached. Nevertheless, the cognitive component of customer experience is stronger in the group with the chatbot in comparison, however, not significantly stronger. This may be due to the selection of items (see Table 3).

The limitations of our study such as sampling procedure, sample size and also a lack of a dependent marketable variable (such as satisfaction, loyalty, patronage intentions (see, e.g., Lou et al., 2022) open up areas for future research. In addition to the neuroscientific methods used in this study, facial recognition or EEG (electroencephalography) may also be helpful to uncover the emotional component of customer experience.

References

Bagdare, S., & Jain, R. (2013). Measuring retail customer experience. *International Journal of Retail & Distribution Management*, *41*(10), 790–804. https://doi.org/10.1108/IJRDM-08-2012-0084

Barari, M., Ross, M., & Surachartkumtonkun, J. (2020). Negative and positive customer shopping experience in an online context. *Journal of Retailing and Consumer Services*, *53*, 101985. https://doi.org/10.1016/j.jretconser.2019.101985

Bawack, R. E., Wamba, S. F., Carillo, K. D. A., & Akter, S. (2022). Artificial intelligence in E-Commerce: A bibliometric study and literature review. *Electronic Markets*, *32*(1), 297–338. https://doi.org/10.1007/s12525-022-00537-z

Bianchin, M., & Angrilli, A. (2012). Gender differences in emotional responses: A psychophysiological study. *Physiology & Behavior*, *105*(4), 925–932. https://doi.org/10.1016/j.physbeh.2011.10.031

Bleier, A., Harmeling, C. M., & Palmatier, R. W. (2019). Creating Effective Online Customer Experiences. *Journal of Marketing*, *83*(2), 98–119. https://doi.org/10.1177/0022242918809930

Bogula, W. (2022). KI in der Wirtschaft: Sprach-KI und ihre Anwednungen. In M. H. Dahm (Ed.), *Haufe Fachbuch. Wie Künstliche Intelligenz unser Leben prägt: KI verständlich erklärt* (1st ed., pp. 157–164). Freiburg: Haufe.

Bojko, A. (2013). *Eye tracking the user experience: A practical guide to research*. Brooklyn, New York: Rosenfeld Media. Retrieved from https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5198124

Bubaš, G., Babić, S., & Čižmešija, A. (2023). Usability and User Experience Related Perceptions of University Students Regarding the Use of Bing Chat Search Engine and AI Chatbot: Preliminary Evaluation of Assessment Scales. In *2023 IEEE 21st Jubilee International Symposium on Intelligent Systems and Informatics (SISY)* (pp. 607–612). IEEE. https://doi.org/10.1109/SISY60376.2023.10417910

Bustamante, J. C., & Rubio, N. (2017). Measuring customer experience in physical retail environments. *Journal of Service Management*, *28*(5), 884–913. https://doi.org/10.1108/JOSM-06-2016-0142

Caruelle, D., Shams, P., Gustafsson, A., & Lervik-Olsen, L. (2024). Emotional arousal in customer experience: A dynamic view. *Journal of Business Research*, *170*, 114344. https://doi.org/10.1016/j.jbusres.2023.114344

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009

Cheng, X., Bao, Y., Zarifis, A., Gong, W., & Mou, J. (2022). Exploring consumers' response to text-based chatbots in e-commerce: The moderating role of task complexity and chatbot disclosure. *Internet Research*, *32*(2), 496–517. https://doi.org/10.1108/INTR-08-2020-0460

Christopoulos, G. I., Uy, M. A., & Yap, W. J. (2019). The Body and the Brain: Measuring Skin Conductance Responses to Understand the Emotional Experience. *Organizational Research Methods*, *22*(1), 394–420. https://doi.org/10.1177/1094428116681073

Clore, G. L., Wyer, Jr. Robert S., Dienes, B., Gasper, K., & Isbell, L. (2001). Affective Feelings as Feedback: Some Cognitive Consequences. In L. Erlbaum (Ed.), *Theories of Mood and Cognition*. Retrieved from

https://www.researchgate.net/publication/229059906_Affective_Feelings_as_Feedback_Some_Cognitive_Consequences

Gefen, D., & Straub, D. W. (2003). Managing User Trust in B2C e-Services. *E-Service Journal*, *2*(2), 7–24. https://doi.org/10.2979/esj.2003.2.2.7

Hoyer, W. D., Kroschke, M., Schmitt, B., Kraume, K., & Shankar, V. (2020). Transforming the Customer Experience Through New Technologies. *Journal of Interactive Marketing*, *51*, 57–71. https://doi.org/10.1016/j.intmar.2020.04.001

Jenneboer, L., Herrando, C., & Constantinides, E. (2022). The Impact of Chatbots on Customer Loyalty: A Systematic Literature Review. *Journal of Theoretical and Applied Electronic Commerce Research*, *17*(1), 212–229. https://doi.org/10.3390/jtaer17010011

Jiang, H., Cheng, Y., Yang, J., & Gao, S. (2022). AI-powered chatbot communication with customers: Dialogic interactions, satisfaction, engagement, and customer behavior. *Computers in Human Behavior*, *134*, 107329. https://doi.org/10.1016/j.chb.2022.107329

Lahmeyer, S., & Roemer, E. (2024). The link between consumer-facing technologies and customer experience in physical retail environments: a critical literature review. *The International Review of Retail, Distribution and Consumer Research*, 1–32. https://doi.org/10.1080/09593969.2024.2332224

Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, *80*(6), 69–96. https://doi.org/10.1509/jm.15.0420

Liao, T., & Yan, B. (2023). Let's Chat If You Are Unhappy – The Effect of Emotions on Interaction Experience and Trust Toward Empathetic Chatbots. In *Volume 3B: 49th Design Automation Conference (DAC)*. American Society of Mechanical Engineers. https://doi.org/10.1115/DETC2023-115318

Lou, C., Kang, H., & Tse, C. H. (2022). Bots vs. humans: how schema congruity, contingency-based interactivity, and sympathy influence consumer perceptions and patronage intentions. *International Journal of Advertising*, *41*(4), 655–684. https://doi.org/10.1080/02650487.2021.1951510

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*, 937–947. https://doi.org/10.1287/mksc.2019.1192

Schmitt, B. (1999). Experiential Marketing. *Journal of Marketing Management*, *15*(1-3), 53–67. https://doi.org/10.1362/026725799784870496

Shumanov, M., & Johnson, L. (2021). Making conversations with chatbots more personalized. *Computers in Human Behavior*, *117*, 106627. https://doi.org/10.1016/j.chb.2020.106627

Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer Experience Creation: Determinants, Dynamics and Management Strategies. *Journal of Retailing*, *85*(1), 31–41. https://doi.org/10.1016/j.jretai.2008.11.001

Verhulst, N., Vermeir, I., Slabbinck, H., Larivière, B., Mauri, M., & Russo, V. (2020). A neurophysiological exploration of the dynamic nature of emotions during the customer experience. *Journal of Retailing and Consumer Services*, *57*, 102217. https://doi.org/10.1016/j.jretconser.2020.102217

Xu, Y., Shieh, C.-H., van Esch, P., & Ling, I.-L. (2020). AI Customer Service: Task Complexity, Problem-Solving Ability, and Usage Intention. *Australasian Marketing Journal*, *28*(4), 189–199. https://doi.org/10.1016/j.ausmj.2020.03.005