

Resolving the Chatbot Disclosure Dilemma: Leveraging Selective Self-Presentation to Mitigate the Negative Effect of Chatbot Disclosure

Nika Mozafari
University of Goettingen
nika.mozafari@wiwi.uni-goettingen.de

Welf H. Weiger
Alfaisal University
wweiger@alfaisal.edu

Maik Hammerschmidt
University of Goettingen
maik.hammerschmidt@wiwi.uni-goettingen.de

Abstract

Chatbots are increasingly able to pose as humans. However, this does not hold true if their identity is explicitly disclosed to users—a practice that will become a legal obligation for many service providers in the imminent future. Previous studies hint at a chatbot disclosure dilemma in that disclosing the non-human identity of chatbots comes at the cost of negative user responses. As these responses are commonly attributed to reduced trust in algorithms, this research examines how the detrimental impact of chatbot disclosure on trust can be buffered. Based on computer-mediated communication theory, the authors demonstrate that the chatbot disclosure dilemma can be resolved if disclosure is paired with selective presentation of the chatbot’s capabilities. Study results show that while merely disclosing (vs. not disclosing) chatbot identity does reduce trust, pairing chatbot disclosure with selectively presented information on the chatbot’s expertise or weaknesses is able to mitigate this negative effect.

1. Introduction

Recent advancements in artificial intelligence combined with the rise in popularity of messaging apps fuel the development and deployment of chatbots in the service frontline. Chatbots are text-based user interfaces that build on natural language processing to emulate human-to-human conversation, possibly replacing computer-mediated conversations with human service providers fully in the near future [24].

Today, chatbots are increasingly capable of imitating human conversation [1]. This makes it challenging for users to correctly identify whether they are interacting with a machine or human when this information is not explicitly provided [3, 33]. As this

development gains traction, service providers have to decide whether to disclose the chatbot identity and, if so, whether to provide additional information about it. From an ethical viewpoint, withholding identity information does not prove tenable, as intransparency regarding the non-human chatbot identity may be perceived as deceptive and could be exploited by service providers.

This is why various courts initiated legal regulations that obligate service providers to disclose the non-human identity of chatbots. California’s now established “bot bill” originally arose from political motives. The bill ought to bring an end to political bots on twitter and other social media platforms being used to deceive voters or artificially create consensus on divisive issues [2]. The European Commission has also been discussing a legal obligation to conspicuously disclose chatbot identity to create higher transparency [9]. These regulations do not only impact political bots, but also commercial, interactive chatbots, not only in California or the EU, but worldwide. This implies that any firm using a chatbot in the service frontline is or will be obligated to provide information on the chatbot’s non-human identity to its users.

First empirical studies attend to this matter and consistently find negative effects of chatbot disclosure on both psychological and behavioural user reactions [12, 13, 17, 20, 21, 25, 26], as people tend to trust algorithms less than humans, despite equal or sometimes even superior service delivery [6]. This in turn is problematic for service providers, as they want to avoid negative user reactions, but will be obligated to disclose chatbot identity sooner or later, hence creating a chatbot disclosure dilemma [20]. As the question whether or not to disclose becomes obsolete due to legal restrictions, focus should be shifted from whether to *how* to disclose chatbot identity. The question arises if chatbot disclosure can be communicated in such a manner, that the loss of trust can be mitigated and the chatbot disclosure dilemma can be resolved.

The concept of selective self-presentation as a part of computer-mediated communication (CMC) theory offers a promising approach to solving this dilemma [31]. In CMC, i.e. communication that takes place through electronic media, there are fewer social cues available that allow for evaluation of the interaction partner than in face-to-face interactions. This can be leveraged, as it allows for highly malleable first impressions, which can be shaped to create trust. Service providers can benefit from this opportunity by transferring the insights from CMC to the issue of chatbot disclosure.

The goal of this study is to examine the impact of chatbot disclosure paired with selective-self presentation on user trust in the conversational partner. The study examines mechanisms from CMC theory, specifically the cues-filtered-out perspective, hyperpersonal model and signaling theory, and contributes to research on mediated conversations, specifically human-chatbot interactions by being the first to empirically investigate how to communicate chatbot identity information without creating a loss of trust.

2. Literature review

To embed our study into existing literature, we discuss related work on chatbot disclosure. Further, we highlight the importance of trust in human-chatbot interactions.

2.1. Related work on chatbot disclosure

Research on the repercussions of chatbot disclosure is still at a nascent stage. Pioneering empirical studies have focused on understanding the effect of disclosing vs. not disclosing the chatbot's identity to users and arrived at the conclusion that transparently communicating chatbot identity comes at the cost of negative user reactions: it may reduce customer retention [20], user acceptance [21], duration of interaction and purchase rate [17], efficiency of human-machine cooperation [13], perceived social presence and humanness [12], and persuasion efficiency [25]. These results are startling, as negative biases to disclosed bots emerge despite equal performance levels of disclosed and undisclosed bots and superiority of examined bots over humans.

Interestingly, all studies have so far only focused on examining *whether* chatbot disclosure yields certain effects on user responses, while the question of *how* to disclose chatbot identity remains unexamined. Notably, one study suggests that negative effects of chatbot disclosure can be mitigated using a late disclosure

strategy [17]. However, this finding may become negligible as laws on transparency forbid chatbot disclosure at later stages of a conversation.

Further, a common ground in explaining negative responses towards chatbots is the lack of trust in algorithms. In this research, we therefore focus on trust as the key response in human-chatbot interactions.

2.2. Trust in algorithms

The biased assessment of disclosed compared to undisclosed bots is rooted in users' inherent aversion towards algorithms. In fact, a broad literature stream has provided evidence that people tend to trust humans over algorithms, despite functional superiority of the latter [14]. Studies on chatbot disclosure show that this effect prevails not only when comparing algorithms to humans, but disclosed algorithms to undisclosed algorithms. This implies that not the actual, but perceived identity impacts trust [25].

We define trust as the trustor's willingness to rely on a trustee to fulfil their obligations, to act in the trustor's interest and to tell the truth [15, 18]. Trust between two exchange partners is a crucial antecedent of desirable user behaviour [11, 19].

Problematically, trust in computer-mediated environments is lower than in face-to-face interactions due to higher levels of uncertainty and scepticism in online environments as a result of fewer available social cues [8, 23]. When one's counterpart then reveals themselves as a chatbot, trust will stoop even lower due to users' aversion towards algorithms. In order for a user to assess trustworthiness of their exchange partner in CMC, information on their identity is essential [7]. Specifically, users assess whether and to what extent they will trust a chatbot as a result of how the chatbot presents itself and communicates its capabilities [10]. This suggests that trust does not have to diminish if the service provider manages to communicate the chatbot's identity in a favourable way. This so-called selective self-presentation is a mechanism in CMC theory that helps facilitate impressions and relationships online [30].

3. Theoretical framework

This section reviews relevant aspects of CMC theory which can be applied to resolving the chatbot disclosure dilemma. Based on this, we derive hypotheses for our study.

3.1. Selective self-presentation in computer-mediated communication

CMC is characterized by a limited social bandwidth due to lower number of social cues. This overarching theme of CMC is often described with the term cues-filtered-out [4], as social cues that are transmitted in face-to-face conversations can only be transmitted to a small extent in CMC. In CMC, mostly verbal cues are evaluated by a message receiver—compared to the additional non-verbal cues in face-to-face interactions [31]. Therefore, as there are fewer cues available, people tend to mistrust information that is communicated in CMC [7].

However, there are instances in which online communication is facilitated in such a manner, that it becomes as personal as or even surpasses the intimacy of human face-to-face communications [30]. This so-called hyperpersonal model of CMC therefore offers ways by which trust levels in computer-mediated interactions can even exceed that of human interactions [31]. The model allows a sender of a message communicative advantages compared to face-to-face interaction, as they are able to strategically develop and control self-presentation and thus allowing an optimized, selective presentation of themselves to others. Hence, due to this high controllability and malleability of first impressions, selective self-presentation can be an efficient tool if applied favourably [30].

This means that in a computer-mediated conversation, a sender of information can actively shape their conversational partner's perception of them by selectively presenting information about themselves [29]. By transmitting cues that portray themselves preferentially, favourable reactions are fostered. The sender can therefore not only provide information they wish the receiver to know, but additionally use mechanisms to create intimacy and eventually trust. For instance, this could include communications of motivations, personal beliefs, affiliations or competencies, as these will greatly affect how the receiver interprets the sender's statements. These cues act as signals, which will be evaluated according to their perceived reliability [7].

While the hyperpersonal model refers to behaviour of human conversational partners that communicate online, service providers can profit from the model's insights by applying its mechanisms to self-presentation of chatbots in order to address the issue of chatbot disclosure.

3.2. Hypotheses development

In our context, merely disclosing chatbot identity without enhancing it with additional information should lower trust compared to not disclosing identity due to higher scepticism towards algorithmic entities [5, 6]. An interaction with an undisclosed chatbot should have higher resemblance to a face-to-face interaction as the user will assume they are interacting with a human [17, 33]. However, if the chatbot's non-human identity is disclosed, perceived social bandwidth will be even smaller, therefore making it harder for feelings of trust to emerge. Thus, without any additional information that fosters a favourable perception of the conversational partner, according to the cues-filtered-out perspective of CMC, reaction to a disclosed chatbot should be negative. Therefore:

H₁: Merely disclosing chatbot identity reduces trust compared to not disclosing chatbot identity.

However, adding selective self-presentation mechanisms to chatbot disclosure can improve the user's perception of their conversational partner. Based on previous research on trust in human-chatbot interactions, we argue for two opposite perspectives of what may be a preferential self-presentation of chatbots. On the one hand, establishing expertise and competence is stated to be the most influential factor for human trust in chatbots [22]. Thus, according to the hyperpersonal model of CMC, if a chatbot focusses the communication of its identity on the selected cues that strengthen perceptions of expertise, trust should be enforced.

On the other hand, research has shown that there seems to be a mismatch of user expectations towards chatbot performance and actual performance [16]. From this perspective, it may be beneficial not to convey information of expertise, but quite contrary selectively present information that conveys the chatbot's limits. Actively communicating weaknesses can create trust as the users can adjust their expectations accordingly [10]. Especially in contexts where users hold negative predispositions in the first place, humility and acknowledgement of limitations can help build trust [28, 34].

As both approaches can shape chatbot identity in preferential ways, we postulate following hypotheses:

H₂: Communicating expertise when disclosing chatbot identity increases trust compared to merely disclosing chatbot identity.

H₃: Communicating weaknesses when disclosing chatbot identity increases trust compared to merely disclosing chatbot identity.

However, as it is reasonable to assume that users are aware of senders being able to manipulate the selectively presented information, varying information should be processed differently. This notion is informed by signaling theory, which states that users assess and evaluate the reliability of signals in the selected self-presentations [7]. Most identity claims cannot be quickly verified, resulting in higher wariness for certain claims than for others. Communicating weaknesses comes at a greater cost for the message sender than claiming expertise [8].

Due to this, a message receiver in CMC should assess a humble admission of limits as more reliable than an ostensibly overconfident communication of expertise. Therefore:

H₄: Communicating weaknesses when disclosing chatbot identity increases trust compared to communicating expertise.

4. Study

To test our hypotheses, we conducted an online scenario experiment. The following sections highlight the study design, sample and results.

4.1. Design and sample

Real-life online chats were evaluated prior to designing the experiment. The scenario style enabled us to create an interaction from which participants could not infer the identity of their conversational partner without disclosure, thus allowing to control for confounding influences and ensuring high internal validity [27].

As the disclosure of the non-human chatbot identity is the central manipulation of this study, a prestudy was

conducted to test whether the conversational partner was perceived as human when the bot identity was not disclosed. For this, we took a measure for perceived identity on a 7-point scale, anchored by 1 = chatbot and 7 = human. We found a significant negative effect of chatbot disclosure on perceived identity ($N = 12$; $M_{disclosed} = 3.833$, $SD = 1.951$; $M_{undisclosed} = 5.417$, $SD = 1.258$; $t = 1,459$, $p < 0.1$), showing that when chatbot identity was not disclosed, participants perceived their conversational partner as significantly more human.

For our main study, we recruited participants using distribution lists and social media. In the experiment, participants were instructed to imagine that they were planning to take out a new liability insurance and were about to contact the insurance company via their online chat. Then, participants were randomly assigned to one of four scenarios: no chatbot disclosure, mere chatbot disclosure, chatbot disclosure with communication of expertise and chatbot disclosure with communication of weaknesses. In the no chatbot disclosure scenario, the conversational partner merely introduced himself as “Michael” in the chat interface. In the three disclosure scenarios, “Michael” revealed himself as a chatbot. Additionally, the chatbot communicated its expertise (“Due to my high efficiency I am able to find the best offer for you”) or its weaknesses (“Please note that I’m only in use for a year now and am still learning”) respectively in the two selective self-presentation scenarios. For an overview of the scenarios see Figure 1.

The rest of the conversation was identical in all four scenarios. The scenarios were presented using mock-ups of an online chat. The user specified what kind of insurance they were looking for and was made a fitting offer by their conversational partner in the online chat. This ensured that actual performance level was held at the same level for all four scenarios, therefore allowing differences in trust only to originate from different forms of identity disclosure.

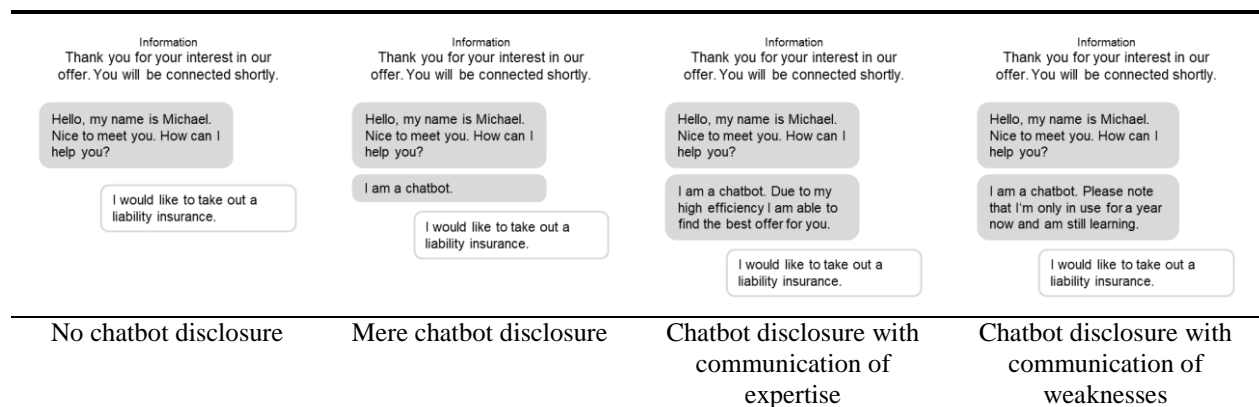


Figure 1. Manipulations of chatbot disclosure

Table 1. Measuring trust

Trust in conversational partner ($\alpha = 0.87$)		
Dimension	Item	Sources
Trust in competence ($\alpha = 0.76$)	The conversational partner is knowledgeable.	Adapted from [15, 32]
	The conversational partner is trustworthy.	
Trust in benevolence ($\alpha = 0.83$)	The conversational partner puts my interest first.	
	The conversational partner wants to understand my needs and wishes.	
Trust in integrity ($\alpha = 0.70$)	The conversational partner is honest.	
	The conversational partner provides unbiased recommendations.	

After reading through the interaction, participants were asked to report their trust in the conversational partner. Trust was measured by taking the mean of participants' statements regarding trust in competence, trust in benevolence and trust in integrity. All items were measured on 7-point-likert scales, anchored by 1 = strongly disagree and 7 = strongly agree. The items for trust were taken from [15] and [32] and adapted slightly to fit the context of our study while keeping chatbot identity undisclosed. For an overview of the items, see Table 1. As the reliability measure for overall trust (combining three dimensions) was higher than for each dimension separately, all dimensions were combined to one comprehensive variable trust.

The initial sample consisted of 346 participants. Those participants that did not complete the entire survey or did not pass attention checks were discarded from further analyses. The effective sample thus consists of 257 participants (58% female, $M_{age} = 26$ years). To make sure that participants were familiar with the context of our study, we set the required minimum age to 18 years. As a manipulation check, we asked participants to recite how their conversational partner had introduced himself to them. All remaining participants answered the manipulation check correctly. There were no significant differences in distribution of age, gender and experience with chatbots between the groups (all $p > 0.1$).

4.2. Results

To test the differences in trust across treatment conditions, we used analysis of variance (ANOVA) with the disclosure manipulations as independent variable and trust in the conversational partner as dependent variable. ANOVA results show an insignificant effect of chatbot disclosure on trust in the conversational partner ($F(3; 253) = 1.69, p = 0.169$). This suggests that at least two treatment conditions do not yield significant differences. Thus, to identify nuances in the effects of the four disclosure scenarios on trust, we used planned

contrasts. An overview of descriptive statistics can be seen in Table 2.

Table 2. Descriptive statistics

Scenario	<i>N</i>	<i>M</i>	<i>SE</i>
No disclosure	68	5.042	0.142
Mere chatbot disclosure	52	4.677	0.162
Chatbot disclosure with communication of expertise	74	5.054	0.136
Chatbot disclosure with communication of weaknesses	63	5.135	0.147

Results of planned contrasts show that merely disclosing chatbot identity reduced trust compared to not disclosing chatbot identity ($\Delta_{Trust} = -0.365, SE = 0.215, t = -1.70, p < 0.1$). This is in line with results of prior research [20] and therefore supports H_1 . As expected, communicating expertise when disclosing chatbot identity increases trust compared to merely disclosing chatbot identity ($\Delta_{Trust} = 0.378, SE = 0.211, t = 1.79, p < 0.1$), supporting H_2 . Furthermore, communicating weaknesses when disclosing chatbot identity also increases trust compared to merely disclosing chatbot identity ($\Delta_{Trust} = -0.365, SE = 0.215, t = -1.70, p < 0.05$), therefore supporting H_3 . Finally, we did not find a significant difference in trust when comparing communication of expertise with communication of weaknesses when disclosing chatbot identity ($\Delta_{Trust} = 0.08, SE = 0.200, t = 0.40, p > 0.1$), therefore not supporting H_4 . Furthermore, though not hypothesized, it is mentionable that there is no significant difference in trust between the selective self-presentation scenarios and the no chatbot disclosure scenario. An overview of results can be seen in Figure 2.

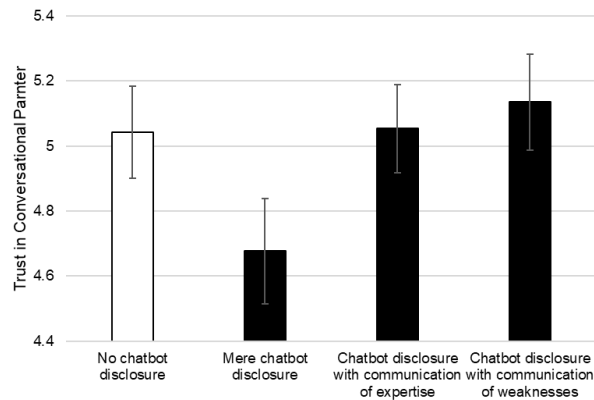


Figure 2. Trust in conversational partner across treatment conditions

5. Discussion

The goal of our study was to examine whether the chatbot disclosure dilemma—trading off transparency regarding chatbot identity and negative user responses—could be resolved by leveraging selective chatbot presentation. In fact, the results confirm the existence of this dilemma—disclosing chatbot identity comes at the cost of lower user trust. Interestingly, the negative effect of disclosure remains while holding the performance levels of service delivery of disclosed and undisclosed conversational partners constant, suggesting that solely the information that users are interacting with an algorithm instead of a live person causes the biased reaction. The negative reactions to chatbot identity disclosure are in line with the cues-filtered-out perspective of CMC, as chatbot identity is connected to a perceived lower capacity to transmit social cues. This problem gains relevance for service providers when providing information on the non-human identity of a chatbot becomes legally inevitable.

However, we find that if disclosing chatbot identity is paired with selectively presented information about the chatbot, the disclosure dilemma can be adequately addressed. Our results suggest that communicating the chatbot’s expertise, i.e. showing what it is capable of, as well as communicating the chatbot’s weaknesses, i.e. showing what it may not be capable of, can produce trust levels corresponding to that of undisclosed conversational partners. This shows that in a chatbot context, the application of the mechanism described in the hyperpersonal model of CMC can fully compensate the loss of trust caused by chatbot disclosure and mimic a situation as if chatbot identity would not have been disclosed. Notably, we find no proof that trust levels can exceed that of human interactions (or here: interactions with undisclosed chatbots), as the theory suggests. However, this does not necessarily contrast with the

theory, as our study only applied one of the approaches of the hyperpersonal model, specifically that approach which fits the context of this study.

Furthermore, we could not prove that one kind of selectively presented information was superior to the other. Signaling theory suggests identity claims will be evaluated differently, depending on whether they come at the cost of the sender [7]. According to this, communicating weaknesses should have resulted in higher trust than communicating expertise, as someone who will admit to their flaws is less likely to be perceived as deceptive than someone who points out their strengths. However, we could not find proof for this mechanism. This might suggest that pairing chatbot disclosure with any kind of further explanation might be sufficient in mitigating the negative effect of disclosure. In the presented scenario, the communication of expertise was justified, as the conversational partner was able to make an informed recommendation for a liability insurance in the interaction, therefore proving that the communication of expertise was not just an empty, deceptive claim, but actually of substance.

The results of our study contribute to existing literature on chatbot identity disclosure by studying it from a CMC theory perspective [12]. Existing studies on the repercussions of chatbot disclosure have thus far mainly focused on the negative effect of disclosing the non-human chatbot identity to users. We are the first to test the effect of different types of chatbot disclosure on user trust. Thereby, we manage to offer a feasible solution for service providers that are faced with the chatbot disclosure dilemma at present or in the imminent future.

6. Outlook and Limitations

With service providers being bound by law to disclose chatbots’ non-human identity, the challenge of creating the right disclosure strategy gains more and more relevance. We show that selectively presenting preferential information about the chatbot when disclosing its identity addresses the chatbot disclosure dilemma.

Future research should investigate whether the mitigation of the negative disclosure effect stems solely from adding an explanation, suggesting a mere exposure effect, or does actually depend on the type and content of the signal provided by the chatbot. This is specifically relevant, as we could not find a significant difference in trust between the expertise and weaknesses communications. This could have resulted from the fact that chatbot performance was held constant in the experiment. Further studies should focus on reactions to chatbot disclosure under consideration of different

performance levels. In addition to this, future studies should also consider testing other operationalisations of communication of expertise and weaknesses, as the phrasing of our manipulations may have had an effect on user trust.

Further, our study was conducted in the context of liability insurance. We chose this context, as it is a practically relevant application for chatbots. However, future studies should examine whether the effects prevail in other contexts, as trusting beliefs and desirability of a high social bandwidth might differ across contexts.

Finally, this study applied a scenario design using mock-ups to allow for systematic manipulation of chatbot disclosure without confounding factors. To add external validity, we plan to examine real-life chatbot interactions in following studies.

7. References

- [1] Adiwardana, D., M.-T. Luong, D.R. So, J. Hall, N. Fiedel, R. Thoppilan, Z. Yang, A. Kulshreshtha, G. Nemade, Y. Lu, and Q. Le V, Towards a Human-like Open-Domain Chatbot, Google Research, 1/27/2020.
- [2] California Legislative Information, “SB-1001 Bots: Disclosure”, 2018.
- [3] Candello, H., C. Pinhanez, and F. Figueiredo, “Typefaces and the Perception of Humanness in Natural Language Chatbots”, in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA. 2017. ACM Press.
- [4] Culnan, M.J. and M.L. Markus, “Information Technologies”, in Handbook of organizational communication: An interdisciplinary perspective, F.M. Jablin, L.L. Putnam, K.H. Roberts, and L.W. Porter, Editors. 1987. Sage: Newbury Park.
- [5] Dawes, R.M., “The robust beauty of improper linear models in decision making”, *American Psychologist*, 34(7), 1979, pp. 571–582.
- [6] Dietvorst, B.J., J.P. Simmons, and C. Massey, “Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err”, *Journal of experimental psychology. General*, 144(1), 2015, pp. 114–126.
- [7] Donath, J., “Identity and Deception in the Virtual Community”, in *Communities in Cyberspace*, P. Kollock and M. Smith, Editors. 1999. Routledge: London.
- [8] Donath, J., “Signals in Social Supernets”, *Journal of Computer-Mediated Communication*, 13(1), 2007, pp. 231–251.
- [9] European Commission, „Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen“, Brussels, 2018.
- [10] Følstad, A., C.B. Nordheim, and C.A. Bjørkli, “What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study”, in *Internet Science: 5th International Conference, INSCI 2018, St. Petersburg, Russia, October 24-26, 2018, Proceedings*, S.S. Bodrunova, Editor. 2018. Springer International Publishing: Cham.
- [11] Hart, C.W.L. and M.D. Johnson, “Growing the Trust Relationship”, *Marketing Management*, 8(1), 1999, pp. 9–19.
- [12] Hendriks, F., C. Ou, A.K. Amiri, and S. Bockting, “The power of computer-mediated communication theories in explaining the effect of chatbot introduction on user experience”, in *Proceedings of the 53rd Hawaii International Conference on System Sciences*. 2020.
- [13] Ishowo-Oloko, F., J.-F. Bonnefon, Z. Soroye, J. Crandall, I. Rahwan, and T. Rahwan, “Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation”, *Nature Machine Intelligence*, 1(11), 2019, pp. 517–521.
- [14] Jussupow, E., I. Benbasat, and A. Heinzl, “Why are we Averse Towards Algorithms? A Comprehensive Literature Review on Algorithm Aversion”, in *Proceedings of the 28th European Conference on Information Systems*. 2020.
- [15] Komiak and Benbasat, “The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents”, *MIS Quarterly*, 30(4), 2006, pp. 941–960.
- [16] Luger, E. and A. Sellen, ““Like Having a Really Bad PA”: The Gulf between User Expectation and Experience of Conversational Agents”, in *Proceedings of the 34th Annual CHI Conference on Human Factors in Computing Systems*. 2016.
- [17] Luo, X., S. Tong, Z. Fang, and Z. Qu, “Machines versus Humans: The Impact of AI Chatbot Disclosure on Customer Purchases”, *Marketing Science*, Forthcoming.
- [18] Moorman, C., R. Deshpandé, and G. Zaltman, “Factors Affecting Trust in Market Research Relationships”, *Journal of Marketing*, 57(1), 1993, pp. 81–101.
- [19] Morgan, R.M. and S.D. Hunt, “The Commitment-Trust Theory of Relationship Marketing”, *Journal of Marketing*, 58(3), 1994, pp. 20–38.
- [20] Mozafari, N., W.H. Weiger, and M. Hammerschmidt, “The Chatbot Disclosure Dilemma: Desirable and Undesirable Effects of Disclosing the Non-Human Identity of Chatbots”, in *Proceedings of the 41st International Conference on Information Systems*. 2020.
- [21] Murgia, A., D. Janssens, S. Demeyer, and B. Vasilescu, “Among the Machines: Human-Bot Interaction on Social Q&A Websites”, in *Proceedings of the 2016 CHI Conference*

on Human Factors in Computing Systems, San Jose, California, USA. 2016.

[22] Nordheim, C.B., A. Følstad, and C.A. Bjørkli, “An Initial Model of Trust in Chatbots for Customer Service—Findings from a Questionnaire Study”, *Interacting with Computers*, 31(3), 2019, pp. 317–335.

[23] Riedl, R., P. Mohr, P. Kenning, F. Davis, and H. Heekeren, “Trusting Humans and Avatars: Behavioral and Neutral Evidence”, in *Proceedings of the International Conference on Information Systems*, Shanghai, China. 2011.

[24] Schuetzler, R.M., J.S. Giboney, G.M. Grimes, and J.F. Nunamaker, JR., “The Influence of Conversational Agents on Socially Desirable Responding”, in *Proceedings of the 51st Hawaii International Conference on System Sciences*. 2018.

[25] Shi, W., X. Wang, Y.J. Oh, J. Zhang, S. Sahay, and Z. Yu, “Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies”, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Hawaii, USA. 2020.

[26] Skjuve, M., I.M. Haugstveit, A. Følstad, and P.B. Brandtzaeg, “Help! Is My Chatbot Falling into the Uncanny Valley? An Empirical Study of User Experience in Human-Chatbot Interaction”, *Human Technology*, 15(1), 2019, pp. 30–54.

[27] Slack, M.K. and J.R. Draugalis, “Establishing the internal and external validity of experimental studies”, *American Journal of Health-System Pharmacy*, 58(22), 2001, pp. 2173–2181.

[28] Tangney, J.P., “Humility: Theoretical Perspectives, Empirical Findings and Directions for Future Research”, *Journal of Social and Clinical Psychology*, 19(1), 2000, pp. 70–82.

[29] Walther, J.B., “A longitudinal experiment on relational tone in computer-mediated and face to face interaction”, in *Proceedings of the 25th Hawaii International Conference on System Sciences*. 1992.

[30] Walther, J.B., “Computer-Mediated Communication: Impersonal, Interpersonal and Hyperpersonal Interaction”, *Communication Research*, 23(1), 1996, pp. 3–43.

[31] Walther, J.B., “Theories of Computer-mediated Communication and Interpersonal Relations”, *The Handbook of Interpersonal Communication*, 4, 2011, pp. 443–479.

[32] Wang, W. and I. Benbasat, “Trust In and Adoption of Online Recommendation Agents”, *Journal of the Association for Information Systems*, 6(3), 2005, pp. 72–101.

[33] Wunderlich, N.V. and S. Paluch, “A Nice and Friendly Chat with a Bot: User Perceptions of AI-Based Service Agents”, in *38th International Conference on Information Systems*, Seoul, South Korea. 2017.

[34] Xie, Y. and S. Peng, “How to repair customer trust after negative publicity: The roles of competence, integrity, benevolence, and forgiveness”, *Psychology and Marketing*, 26(7), 2009, pp. 572–589.