

More Than a Bot? The Impact of Disclosing Human Involvement on Customer Interactions with Hybrid Service Agents

Ulrich Gnewuch,^{a,*} Stefan Morana,^b Oliver Hinz,^c Ralf Kellner,^d Alexander Maedche^a

^aKarlsruhe Institute of Technology, 76133 Karlsruhe, Germany; ^bSaarland University, 66123 Saarbrücken, Germany; ^cGoethe University Frankfurt, 60323 Frankfurt am Main, Germany; ^dUniversity of Passau, 94032 Passau, Germany

*Corresponding author

Contact: ulrich.gnewuch@kit.edu,  <https://orcid.org/0000-0003-1423-1777> (UG); stefan.morana@uni-saarland.de,  <https://orcid.org/0000-0002-2266-1286> (SM); ohinz@wiwi.uni-frankfurt.de,  <https://orcid.org/0000-0003-4757-0599> (OH); ralf.kellner@uni-passau.de (RK); alexander.maedche@kit.edu,  <https://orcid.org/0000-0001-6546-4816> (AM)

Received: February 23, 2022

Revised: November 30, 2022; April 25, 2023

Accepted: June 27, 2023

Published Online in Articles in Advance:
August 23, 2023

<https://doi.org/10.1287/isre.2022.0152>

Copyright: © 2023 The Author(s)

Abstract. The proliferation of hybrid service agents—combinations of artificial intelligence (AI) and human employees behind a single interface—further blurs the line between humans and technology in online service encounters. While much of the current debate focuses on disclosing the nonhuman identity of AI-based technologies (e.g., chatbots), the question of whether to also disclose the involvement of human employees working behind the scenes has received little attention. We address this gap by examining how such a disclosure affects customer interactions with a hybrid service agent consisting of an AI-based chatbot and human employees. Results from a randomized field experiment and a controlled online experiment show that disclosing human involvement before or during an interaction with the hybrid service agent leads customers to adopt a more human-oriented communication style. This effect is driven by impression management concerns that are activated when customers become aware of humans working in tandem with the chatbot. The more human-oriented communication style ultimately increases employee workload because fewer customer requests can be handled automatically by the chatbot and must be delegated to a human. These findings provide novel insights into how and why disclosing human involvement affects customer communication behavior, shed light on its negative consequences for employees working in tandem with a chatbot, and help managers understand the potential costs and benefits of providing transparency in customer–hybrid service agent interactions.

History: Karthik Kannan, Senior Editor; Jason Chan, Associate Editor.



Open Access Statement: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. You are free to download this work and share with others, but cannot change in any way or use commercially without permission, and you must attribute this work as “*Information Systems Research*. Copyright © 2023 The Author(s). <https://doi.org/10.1287/isre.2022.0152>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.”

Supplemental Material: The online appendices are available at <https://doi.org/10.1287/isre.2022.0152>.

Keywords: hybrid service agent • artificial intelligence • human involvement disclosure • communication style • chatbot • human–AI collaboration • customer service

1. Introduction

With the rapid proliferation of artificial intelligence (AI) in online service encounters, customers increasingly find themselves interacting with chatbots instead of human service employees (Schanke et al. 2021). However, despite advances in AI, chatbots frequently struggle with nonroutine questions and complex requests, causing frustration and poor customer experience (Schuetzler et al. 2021). To avoid these issues, firms have begun to employ hybrid service agents: combinations of AI agents (e.g., chatbots) and human agents (e.g., service employees) that function as an integrated unit with a single

interface to the customer (Rai et al. 2019, Schuetzler et al. 2021).¹ Their fundamental idea is to balance the complementary strengths and weaknesses of artificial and human intelligence by combining them such that common questions and requests are handled by AI, whereas the rest are delegated to a human (De Keyser et al. 2019).

Although hybrid service agents offer several advantages over service channels operated by either humans (e.g., live chat) or AI alone (e.g., chatbots), they further blur the line between humans and technology in online service encounters. Customers already struggle to determine whether they are interacting with a human or a

chatbot (Mozafari et al. 2022); this is likely to be exacerbated in interactions with hybrid service agents, which may involve a chatbot, a human employee, or both (Grimes et al. 2021). Not only does this cause confusion and annoyance to customers, but it has also become a serious concern for firms. Recent policy initiatives aiming to protect customers from counterfeit service encounters, such as California’s “BOT bill” (State of California 2018) and the European “Ethics Guidelines for Trustworthy AI” (AI HLEG 2019), have put pressure on firms to disclose that their chatbots are not real people (Robinson et al. 2020). However, above and beyond making the chatbot identity transparent, firms that use hybrid service agents (rather than fully automated chatbots) must also decide whether to disclose *human involvement* or avoid transparency about the behind-the-scenes employees who step in if the chatbot is unable to respond. Although anecdotal evidence suggests that not disclosing human involvement can cause backlash from the firm’s customer base and the general public (Forbes 2020), research on the impact of disclosing it is scarce. Prior studies have focused on the impact of chatbot identity disclosure in customer–chatbot interactions (Luo et al. 2019, Mozafari et al. 2022), with little attention being paid to hybrid service agents in general (Adam et al. 2022) and the disclosure of human involvement in customer–hybrid service agent interactions in particular. Hence, an important yet largely unanswered question is whether firms should disclose human involvement in customer–hybrid service agent interactions and, if so, how such a disclosure affects not only customers but also the employees who work in tandem with the chatbot.

Against this backdrop, our aim is to empirically examine the impact of human involvement disclosure (HID) on customer interactions with hybrid service agents. Specifically, we focus on understanding whether, how, and why customers communicate differently with a hybrid service agent when human involvement is disclosed (versus not disclosed). Although communication is a fundamental part of customer–firm interactions, how customers use language to express themselves is an often overlooked aspect of customer behavior (Berger et al. 2020). What customers say and how they say it (their communication style) cannot only provide marketing insights but also affect or even disrupt service operations (Altman et al. 2021). This is therefore a crucial concern for firms that use AI-based technologies designed to communicate with customers using natural language. We also examine an important downstream consequence: how customer communication style influences the workload of human employees. A higher workload could not only affect service operations but also imply that firms are unable to fully realize the benefits of employing hybrid service agents. To this end, we conducted two experiments in which customers interacted with a hybrid service agent via chat in a customer service

context. The results of our randomized field experiment indicate that customers exhibit a more human-oriented communication style when human involvement is disclosed (versus not disclosed), which in turn increases employee workload. Our controlled online experiment replicated these findings and, in addition, revealed that the effect of HID on customer communication style is driven by customers’ impression management concerns. That is, customers are more concerned about making a good impression when they know that humans are working in tandem with the chatbot, which leads them to adopt a more human-oriented communication style. We conducted various robustness checks to rule out alternative explanations for our results, such as customer perceptions of the chatbot’s capabilities and variation in employee language. Furthermore, our results were robust across different analytical approaches, across subsamples, and to the inclusion of several control variables. Finally, we performed a set of additional analyses to generate further insights into the impact of HID on other business-related outcomes, such as customer tendency to seek out human involvement and customer sentiment.

Our work offers three main contributions to information systems (IS) research. First, it contributes to the literature on the role of information technology (IT) in customer service encounters. While prior research has primarily focused on IT-based self-service (e.g., web portals, chatbots) and human-based service (e.g., phone, live chat), our study sheds light on a novel hybrid approach that combines humans and chatbot technology behind a single interface. In particular, our empirical investigation of customer–hybrid service agent interaction provides important insights into how customers respond to hybrid service interfaces that further blur the once clear lines between humans and technology in customer service encounters. Second, our research adds to the emerging literature on the use of AI in service automation by revealing how transparency about human involvement in predominantly AI-based service encounters can place additional workload on employees working in tandem with AI and ultimately undermine AI’s ability to free up employees from mundane customer service work. Third, our research extends human–computer interaction (HCI) literature by identifying impression management concerns as a key psychological mechanism in customer–hybrid service agent interactions.

2. Related Literature

Our work is informed by three research streams. First, it relates to the established stream of IS research on the role of IT in service encounters. Because of the transformation from face-to-face service encounters to IT-based self-service (e.g., web portals) and IT-mediated service (e.g., live chat), seeking to understand customer usage of online service channels, the drivers of channel choice,

and the interactions among different channels has a long tradition in IS research (Ba et al. 2010, Kumar and Telang 2012). This line of work also introduced the notion of “hybrid services” to describe combinations of IT-based self-service and human-based service (Xu et al. 2014). However, little research has examined hybrid service interfaces and how customers respond to them (Adam et al. 2022).

Second, particularly relevant to our work is the emerging stream of IS research on the use of AI to automate repetitive customer interactions and free up service employees for more value-adding tasks. One of the most prominent technologies in this domain is that of AI-based chatbots (Adam et al. 2020, Schanke et al. 2021, Han et al. 2022). Chatbots are software applications designed to interact with customers using natural language and answer their questions around the clock (Gnewuch et al. 2022). However, as chatbots often struggle with complex questions and nonroutine requests, many firms have recognized the need to keep a human in the loop (Schuetzler et al. 2021). The idea is that chatbots and employees work together in a symbiotic fashion (Jain et al. 2021), with simple questions and requests handled automatically by the chatbot while more difficult ones are delegated to a human. This has given rise to a class of human–AI hybrids (Rai et al. 2019) that we refer to as hybrid service agents: combinations of AI agents (e.g., chatbots) and human agents (e.g., service employees) that function as an integrated unit and serve customers via a single interface.

Third, our work is closely related to prior research in the field of HCI. Although little research has focused on human–AI hybrids, several studies have investigated how and why people interact differently with human versus nonhuman counterparts (e.g., chatbots). By comparing human–human and human–chatbot interactions, researchers have found that people invest greater time and effort into interactions with another human (Shechtman and Horowitz 2003), use a simpler vocabulary and more profanity when interacting with a chatbot (Hill et al. 2015), and demonstrate different personality traits depending on the identity of their counterparts (Mou and Xu 2017). The literature has identified several psychological mechanisms related to perceptions of humans versus chatbots that may help explain these differences. One key set of mechanisms is linked to the two major dimensions of social cognition: competence and warmth. Compared with chatbots, humans are generally perceived as more competent (e.g., having more expertise and knowledge) and warmer (e.g., being more empathetic, friendly, and trustworthy) (Go and Sundar 2019, Luo et al. 2019, Cheng et al. 2022, Mozafari et al. 2022). A related mechanism involves the expectations that people form before or during interactions with a human or chatbot. For example, people tend to have higher competence-related expectations of humans (e.g.,

regarding conversational engagement) than they do of chatbots (Grimes et al. 2021). In light of AI advances that make distinguishing between humans and chatbots increasingly difficult, recent research in this stream has focused on the impact of chatbot identity disclosure—that is, informing customers that they are interacting with a chatbot and not a human. On the one hand, studies have determined that such disclosure reduces purchase rates (Luo et al. 2019), customer trust (Mozafari et al. 2022), and service evaluation (Castelo et al. 2023). On the other hand, research has found that customers feel fooled when made to believe they were interacting with a human but find out that it was actually a chatbot (Castillo et al. 2021). Table 1 presents a summary of previous studies on disclosure in customer interactions with AI-based service agents (e.g., chatbots).

Despite the considerable amount of existing research, two important gaps remain. First, empirical investigation of service encounters has predominantly focused on service channels either operated by humans alone or enabled by technology alone (Ba et al. 2010, Kumar and Telang 2012, Schanke et al. 2021, Han et al. 2022). Therefore, we know little about hybrid approaches combining employees and technology not only within the same channel but also behind a single interface (Adam et al. 2022). The growing prevalence of hybrid service agents underscores the need to investigate customer interactions with them and better understand the potential consequences for employees who step in if the chatbot is unable to respond. Second, while prior research has demonstrated the largely negative impact of chatbot identity disclosure in service encounters (Go and Sundar 2019, Luo et al. 2019, Mozafari et al. 2022), little attention has been paid to the disclosure of human involvement. This is particularly important in the context of hybrid service agents, as the more elusive nature of their interactions with customers may involve a chatbot, a human employee, or both. As summarized in Table 1, our work addresses these gaps by investigating the impact of HID on customer interactions with hybrid service agents.

3. Theory Development

Our primary focus in this research is to understand whether, how, and why customers communicate differently with a hybrid service agent when human involvement is disclosed (versus not disclosed), and how these differences affect the workload of employees working in tandem with the chatbot. Drawing on research from interpersonal communication and human–AI collaboration, we focus our theorizing on the impact of HID on one direct customer outcome—communication style—and one indirect employee outcome—workload. Furthermore, we investigate customers’ impression management concerns as the psychological mechanism underlying the effect of HID on communication style. Finally, given that

Table 1. Related Empirical Studies on Disclosure in Customer–(Hybrid) Service Agent Interaction

Study	Agent type	Disclosure type	Method	Outcome variables	Psychological mechanisms	Main findings
Castelo et al. (2023)	Chatbot	Chatbot identity disclosure	Online experiment	Overall service evaluation	Cost-cutting attributions	Chatbot identity disclosure leads to negative service evaluations because customers attribute firms’ use of chatbots to cost-cutting motivations.
Cheng et al. (2022)	Chatbot	Chatbot identity disclosure	Online survey	Trust in chatbot, reliance on and resistance to chatbot	Perceived empathy and friendliness	Chatbot identity disclosure weakens the positive effect of perceived empathy on trust in the chatbot but strengthens the positive effect of perceived friendliness on trust. Trust leads to higher reliance and lower resistance to the chatbot.
Go and Sundar (2019)	Chatbot	Chatbot identity disclosure	Online experiment (Wizard of Oz)	Attitudes toward the website, revisit intention	Perceived homophily, expertise, and friendliness	Chatbot identity disclosure has negative effects on perceived homophily, attitudes toward the website, and revisit intentions for a highly anthropomorphized chatbot. The disclosure also reduces perceived expertise when the chatbot is highly interactive.
Grimes et al. (2021)	Chatbot	Chatbot identity disclosure	Online experiment	Perceived conversational engagement	Expected conversational engagement	Chatbot identity disclosure reduces expectations of conversational engagement, thereby increasing the likelihood of a chatbot exceeding, rather than falling short of, initial expectations.
Huang and Dootson (2022)	Chatbot	Employee availability disclosure	Online experiment (scenario-based)	Customer aggression	Coping strategies	Disclosing the option to be connected to an employee after, rather than before, a chatbot failure leads customers to engage in emotion-focused coping, resulting in customer aggression.
Luo et al. (2019)	Chatbot	Chatbot identity disclosure	Field experiment	Customer purchase rates	Perceived knowledge and empathy	Chatbot identity disclosure reduces customer purchase rates because customers perceive a disclosed chatbot as less knowledgeable and less empathetic.
Mozafari et al. (2022)	Chatbot	Chatbot identity disclosure	Online experiment (scenario-based)	Customer retention	Trust	Chatbot identity disclosure reduces customer retention through decreased trust in high service criticality situations but increases customer retention through increased trust when a service failure occurs.
This study	Hybrid (chatbot +human)	Human involvement disclosure	Field experiment, online experiment	Customer communication style, employee workload	Impression management concerns	Human involvement disclosure leads customers to adopt a more human-oriented communication style, which in turn increases employee workload. The effect of human involvement disclosure on customer communication style is driven by customers’ impression management concerns.

a disclosure can occur at different points in time (Luo et al. 2019), we differentiate between disclosing human involvement before the interaction starts (*up-front HID*) and during the interaction when an employee steps in (*step-in HID*). Figure 1 illustrates our research model.

3.1. Human Involvement Disclosure and Customer Communication Style

Communication style can be understood as one's pattern of communication when interacting with another person. In text-based communication, it specifically refers to how people form a message beyond its content (Brown et al. 2016). The same content can be expressed in different ways by, for example, using fewer or more words and less or more complex language. According to audience design theory (Bell 1984), people tailor their communication style to fit the audience. For example, when addressing a child, adults tend to avoid complex formulation and vocabulary. Importantly, people use such audience design strategies not only when communicating with other humans but also when interacting with technology. For example, when addressing a chatbot, people tend to write short messages with simple sentence structures (Shechtman and Horowitz 2003, Hill et al. 2015). Furthermore, their messages often consist of incomplete sentences or only keywords, resembling search engine queries rather than natural conversation (Castillo et al. 2021). Drawing on these empirical insights, we argue that without HID, customers interacting with a hybrid service agent would assume their counterpart is an automated chatbot² and therefore tailor their messages in a similar fashion by adopting a rather task-oriented, command-like communication style.

Against this backdrop, we draw on audience design theory to propose that a customer's communication style will differ from the one described above if human involvement in the interaction with a hybrid service agent is disclosed. Specifically, we argue that HID makes customers aware that their audience includes not only a chatbot but also an unseen employee working behind the scenes. When human involvement is disclosed *up front*—that is, before customers start interacting with a hybrid service agent—the disclosure typically indicates the possibility of human involvement during the course of the interaction but does not guarantee that an employee will step in. Nevertheless, revealing the presence of an employee working in tandem with the chatbot signals to customers that their audience includes not only a chatbot but also a human who might monitor, read, and eventually become involved in the interaction. According to audience design theory, people also tailor their communication style to so-called auditors—those who are not directly involved with but are present and listen to a conversation. For example, Youssef (1993) observed that the presence of an auditor, such as a child's mother, has a greater influence on how children speak than the

actual identity of their counterpart does. Similarly, research has shown that the mere presence of another human being can substantially influence customer behavior in both offline and online settings (Argo and Dahl 2020). Although research on audiences consisting of both humans and nonhumans is scarce, prior studies have revealed notable differences in how people communicate with another human versus a chatbot. More specifically, people tend to use longer, more vocabulary-rich messages with complete, grammatically correct sentences when interacting with a human (Hill et al. 2015). Based on these considerations, we argue that after an up-front HID, customers will tailor their communication style to take into account the employee's presence. Consequently, they will exhibit a more *human-oriented* communication style that is characterized by longer and more natural sentences (e.g., "Hello, I just received my monthly bill, but it must be wrong. Can you help me?") instead of simple keywords (e.g., "wrong bill"). We therefore pose the following hypothesis.

Hypothesis 1. *After an up-front human involvement disclosure (versus no up-front disclosure), customers exhibit a more human-oriented communication style in interacting with the hybrid service agent.*

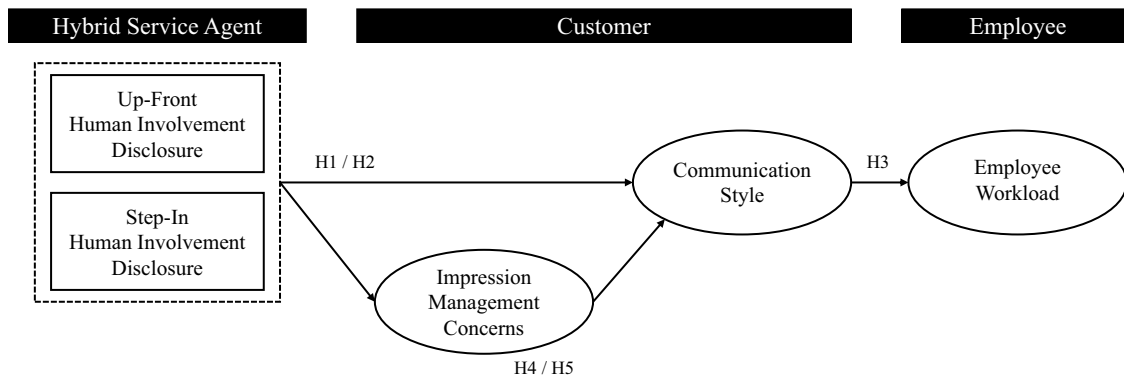
Besides up front, human involvement can also be disclosed during an interaction when an employee actually steps in for the chatbot. Such step-in HID may be particularly salient to customers as it signals that they are in fact engaging with two immediate addressees (a chatbot and an employee) rather than having a dyadic conversation with a chatbot. Audience design theory states that when audiences consist of more than one person, people use a communication style that appeals to all members of the audience. For example, in conversations with multiple participants, people tailor their communication style to take the perspectives of all addressees into consideration (Yoon and Brown-Schmidt 2019). Given that research has shown that people communicate differently with a human than with a chatbot (Hill et al. 2015), we argue that after a step-in HID, customers will tailor their communication style in a way that takes into account the employee who just stepped in and might continue to be involved in the interaction. As a result, their communication style will be more human-oriented than when there is no step-in HID. Therefore, we propose a second hypothesis.

Hypothesis 2. *After a step-in human involvement disclosure (versus no step-in disclosure), customers exhibit a more human-oriented communication style in interacting with the hybrid service agent.*

3.2. Customer Communication Style and Employee Workload

Human-AI collaboration can occur not only on the job level but also on the level of tasks and task instances

Figure 1. Research Model



(Fügener et al. 2022). To distribute work between humans and AI at the task-instance level, human–AI collaborative environments require effective delegation mechanisms (Baird and Maruping 2021, Fügener et al. 2022). Hybrid service agents build on this idea by leveraging a chatbot to handle common questions and requests and delegating more difficult ones to an employee. The underlying delegation mechanism typically follows a simple rule: If the chatbot is unable to respond, delegate the message to a human. Consequently, the workload of employees working in tandem with a chatbot is not fixed but depends heavily on the delegation rate of customer messages.

Against this backdrop, we argue that all else being equal, the style in which customers communicate with a hybrid service agent influences how much work will be delegated to an employee. Although the field of natural language processing (NLP) has made significant progress in enabling AI-based chatbots to understand natural language input, research has shown that how customers formulate their questions and requests can negatively affect NLP performance (Beaver et al. 2020). For instance, even small amounts of noise in a message can make it more difficult for a chatbot to understand what exactly the customer seeks (Beaver et al. 2020). Based on these observations, we argue that when customers adopt a more human-oriented communication style characterized by longer and more natural sentences instead of simpler keyword requests, the chatbot is less well equipped to deal with these messages and more likely to delegate them to a human. In contrast, when customer messages resemble short and simple information requests similar to search engine queries, the chatbot is more likely to handle them automatically without human involvement. Hence, we expect that a more human-oriented communication style plays to the weaknesses of a chatbot and thus leads to higher delegation rates. Employees’ workloads then increase, as they need to step in for the chatbot more often and spend more time responding to customer messages themselves. This leads to our third hypothesis.

Hypothesis 3. *A more human-oriented communication style of customers in interacting with a hybrid service agent increases employee workload.*

3.3. Mediating Role of Impression Management Concerns

Up to this point, we have focused on how customers adapt their communication style depending on whether human involvement is disclosed in the interaction with a hybrid service agent. To theorize the underlying psychological mechanism at work, we draw on social psychology research that suggests that tailoring one’s communication style is an important component of impression management. As we elaborate below, we propose that disclosing human involvement activates *impression management concerns*—that is, concerns about making a good impression on others—which in turn lead customers to adopt a more human-oriented communication style.

A wealth of research in social psychology has shown that people are deeply concerned with how others perceive them (Leary and Kowalski 1990). Even when no immediate or future outcomes depend on the impressions they make, people have a strong desire to be viewed in a positive light and do their best to make a good impression. As a result, people follow social standards and norms (i.e., act more desirably) when others are present (Leary and Kowalski 1990). Compared with human–human interactions, impression management concerns are less likely to arise when people interact with machines such as robots or chatbots (Glikson and Woolley 2020). Although this offers a clear advantage in healthcare contexts where patients can open up about sensitive topics, reduced impression management concerns have also proven to be beneficial in customer service. For example, Følstad et al. (2018) found that people were less concerned when seeking answers from a customer service chatbot instead of a human employee because they did not feel judged for their potentially naive questions. Taken together, this evidence suggests

that impression management concerns are less pronounced when communicating with a chatbot.

Against this backdrop, we argue that disclosing (versus not disclosing) human involvement increases customers' impression management concerns when interacting with a hybrid service agent, which in turn influences their communication style. More specifically, when customers are made aware of the employee working in tandem with the chatbot, they become concerned about making a good impression and presenting themselves in a positive light to the unseen employee. Following this line of thought, we argue that this effect occurs not only when an employee actually steps in during an interaction (step-in HID) but also when there is no guarantee that an employee will do so (up-front HID); just the existence of another human who might read and monitor their conversation with the chatbot should be enough to activate impression management concerns. This argument is in line with prior research indicating that the mere presence of another human, for example in a store, triggers impression management concerns (Argo and Dahl 2020). As social psychology research has shown that such concerns lead people to act in socially desirable ways, we further argue that customers concerned about making a good impression will adopt a human-oriented communication style that is more in line with social standards and norms of communication. Although an in-depth discussion of communication norms is beyond our scope, it is reasonable to assume that talking to someone as if they were a search engine (e.g., with simple keyword requests instead of natural sentences) would not leave a good impression but rather be considered awkward and rude. Hence, customers who are concerned about making a good impression due to a HID in their interaction with the hybrid service agent will exhibit a more human-oriented communication style than they would if they believed their counterpart to be just a chatbot.

Hypothesis 4. *The effect of an up-front human involvement disclosure on customer communication style is mediated by impression management concerns.*

Hypothesis 5. *The effect of a step-in human involvement disclosure on customer communication style is mediated by impression management concerns.*

4. Randomized Field Experiment: Impact of Human Involvement Disclosure on Customer Communication Style and Employee Workload

To test our hypotheses on how up-front and step-in HID affect customer communication style (Hypotheses 1 and 2) and employee workload (Hypothesis 3), we conducted a randomized field experiment in cooperation with a multinational telecommunications company. The

company had implemented a hybrid service agent on their website to answer common questions (e.g., contract extension, Internet connection breakdown) and provide information about their products and services (e.g., mobile phone plans). The hybrid service agent could be accessed via a chat interface and introduced itself as “Lisa, an automated chatbot” so that customers knew from the start that they were not interacting with a real person. A team of human employees worked behind the scenes and stepped in when customer messages were not handled automatically. Before our experiment, the hybrid service agent had not disclosed any human involvement.

4.1. Method

4.1.1. Experimental Design and Treatments. The field experiment consisted of a 2 (up-front HID: disclosure versus no disclosure) \times 2 (step-in HID: disclosure versus no disclosure) between-subjects factorial design, resulting in four separate experimental conditions. In the conditions with up-front HID, possible human involvement was disclosed at the outset of the interaction with the customer. This was operationalized by inserting the statement “If I don’t know the answer to your question, my human colleague will read your message and give you an answer” at the end of the hybrid service agent’s welcome message to indicate that, under certain conditions, a human employee could step in. In the conditions without up-front HID, this statement was not included.

In the conditions with step-in HID, human involvement was disclosed at the point in time when a human employee stepped in during the customer’s interaction with the hybrid service agent. This was operationalized by sending the message “One moment, please. Unfortunately, I don’t know the answer to your question. I will pass it on to my human colleague who will give you an answer shortly” when the chatbot was unable to respond automatically. In the conditions without step-in HID, the hybrid service agent did not disclose the actual human involvement and customers received the message “One moment, please. Currently looking for an answer to your question” instead. Online Appendix A.1 provides screenshots of both treatments.

4.1.2. Manipulation Check. Following prior studies (Schanke et al. 2021), we conducted a separate manipulation check to confirm the effectiveness of our treatments. Based on the scenario of a customer receiving a higher-than-usual mobile phone bill, we designed a typical interaction between a customer and the hybrid service agent of a fictitious telecommunications company (see Online Appendix A.2 for details). We then created four different versions of a video playing out the interaction based on the four experimental conditions. To carry

out the manipulation check, we recruited 120 participants via the online platform Clickworker who received €0.50 in exchange for their participation. They were randomly assigned to watch one of the four video-recorded interactions and then report the extent to which they believed that a human had been involved in the interaction, using a seven-point Likert scale (1 = strongly disagree to 7 = strongly agree). A one-way analysis of variance (ANOVA) revealed significant variation in perceived human involvement across conditions ($F(3, 116) = 64.00, p < 0.001$). Planned contrasts confirmed that perceived human involvement was significantly lower in the condition without disclosure (mean (M) = 1.46, standard deviation (SD) = 0.88) than in the conditions with up-front HID ($M = 3.76, SD = 2.29; t(116) = 5.73, p < 0.001$), step-in HID ($M = 6.17, SD = 1.12; t(116) = 11.86, p < 0.001$), or both up-front and step-in HID ($M = 6.06, SD = 1.37; t(116) = 11.85, p < 0.001$). Hence, our experimental manipulations performed as intended.

4.1.3. Procedure and Hybrid Service Agent Design. The field experiment proceeded as follows. When customers clicked on the “Chat” button on the company’s website, a chat window opened, and they were randomly assigned to one of the four experimental conditions. In the window, customers were greeted by the hybrid service agent in accordance with their designated experimental condition and then entered their questions or requests. During the interaction, each customer message was processed by the chatbot’s NLP algorithm to identify the customer’s intent (i.e., the action or goal the customer hoped to achieve by sending the message). For example, the message “I want to cancel my contract” resulted in the intent “churn.” If an intent was recognized with a confidence score greater than 0.95, the chatbot automatically sent a predefined response mapped to the intent. For the intent “churn,” this was “We’re sorry to hear that you want to cancel your contract. Please select which product or service you want to cancel.” The intents and the predefined responses had been refined by the company over several months and remained the same during the experiment. If no intent was recognized with a confidence score greater than 0.95, the message was delegated to a human. In such cases, employees received a notification in the employee user interface of the hybrid service agent indicating that they needed to step in for the chatbot in a customer interaction (see Online Appendix A.1 for a screenshot). Employees could then either (1) confirm a response suggested by the chatbot (with a confidence score of less than 0.95), (2) edit the suggested response, or (3) enter a completely new response. When a response was finalized, the employee would send it to the customer. The delegation mechanism was the same for all customer messages.

4.1.4. Data Collection and Measures. Data collection took place over a two-week period in February 2019.

During this period, a total of 8,966 customers were randomly assigned to one of the four experimental conditions, resulting in a roughly equal number of customers across conditions. For each customer–hybrid service agent interaction (hereafter referred to as chat), we collected chat data (e.g., text content of all messages, timestamps, sender), basic customer characteristics (e.g., device, location), and information about the activity of employees working in tandem with the chatbot (e.g., number of customer messages handled by an employee, response times). Sensitive customer data, such as names, phone numbers, and addresses, were anonymized before the analysis. To assess the efficacy of our randomization procedure, we compared the experimental groups on two baseline customer characteristics (device and location) and assessed the number of customers in each group across different time periods (day of the week and part of the day). The results showed that customers were randomly assigned to the experimental groups based on device and location and across weekdays and within a day (see Online Appendix A.3 for details). On average, the chats in our sample lasted 7.82 minutes ($SD = 57.77$) and included 4.71 customer messages ($SD = 2.37$), 6.06 chatbot messages ($SD = 2.16$), and 1.02 employee messages ($SD = 1.28$). Human employees stepped in for the chatbot in 60.8% of chats.

Our focal constructs were customer communication style and employee workload. Both were modeled as latent constructs using three well-selected indicators. To operationalize customer communication style and specifically measure the degree of human orientation, we drew on research that examined how people communicate with a chatbot as opposed to with another human (Shechtman and Horowitz 2003, Hill et al. 2015, Knijnenburg and Willemsen 2016). Based on these studies, we identified three key structural characteristics of communication style that differ between human–chatbot and human–human communication: verbosity, complexity, and density. First, verbosity can be understood as the use of many or too many words. Prior studies have found that people use significantly fewer words when interacting with a chatbot as opposed to with another human (Shechtman and Horowitz 2003, Hill et al. 2015). We calculated verbosity as the average number of words per message from a customer in a chat. Second, complexity (also called linguistic complexity or readability) refers to the ease or difficulty with which people can read and understand a message. Prior studies have found that when interacting with a chatbot, people tend to avoid complex sentences and prefer command-style language with only keywords (Knijnenburg and Willemsen 2016, Castillo et al. 2021). We calculated the average complexity of customers’ messages in a chat using the readability formula of Coleman (1971). When we use an alternative complexity measure, the results of our main analyses are qualitatively similar. Third, density (also

called functional density) is a characteristic that differentiates between function words (e.g., conjunctions, pronouns, prepositions) and content words (e.g., nouns, verbs, adjectives). Whereas content words carry meaning and information, function words have little meaning and only serve to make sentences grammatically correct by binding content words together. Prior studies have found that people use more grammatically correct sentences when interacting with another human compared with when they interact with a chatbot (Hill et al. 2015, Knijnenburg and Willemsen 2016). We calculated density as the average ratio of function words to the total number of words in each of a customer's messages in a chat. As in previous research, we used the linguistic inquiry and word count (LIWC) text analysis program to classify words into these predefined categories (Pennebaker et al. 2015). Taken together, we modeled customer communication style as a latent construct with three structural characteristics—verbosity, complexity, and density—as reflective indicators. Online Appendix A.4 provides further details.

Employee workload in hybrid service agents is inextricably linked to the chatbot's ability to handle customer messages automatically rather than delegating them to a human. From our discussions with the company, we learned that employees' overall workloads depend not only on how often they are required to step in (frequency) but also on how much time they spend responding to a customer's messages (duration) and whether they can simply confirm or edit the chatbot's suggested response or must enter a completely new one (intensity). We therefore adopted these factors as three key indicators of employee workload. First, we measured frequency as the total number of responses sent by an employee instead of the chatbot in a chat. Second, we calculated duration as the total time that employees spent responding to customer messages in a chat. We created this variable by computing the elapsed time between each customer message and the employee response (Altman et al. 2021) and taking the sum of these response times. Third, we calculated intensity using a weighted average of the specific actions that employees took when they stepped in. Naturally, confirming a suggested response required less work than editing it before sending, which in turn required less work than entering a new response. Taken together, we modeled employee workload as a latent construct with three reflective indicators: frequency, duration, and intensity. Online Appendix A.4 provides further details. Table 2 presents descriptive statistics for the indicators of both constructs in each experimental condition.

4.2. Results

4.2.1. Measurement Model Evaluation. We first conducted a confirmatory factor analysis using the lavaan package in R (Rosseel 2012) to assess the validity and

reliability of our latent constructs (1) customer communication style and (2) employee workload. The results indicate that the proposed measurement model exhibits a good fit to the data (comparative fit index (CFI) = 0.99, Tucker–Lewis index (TLI) = 0.98, root mean square error of approximation (RMSEA) = 0.04, standardized root mean square residual (SRMR) = 0.02), providing support for construct validity. We further assessed convergent and discriminant validity by examining factor loadings, composite reliability scores, interconstruct correlations, and average variances extracted (AVE) for each construct. All indicators loaded strongly on their constructs with loadings ranging from 0.68 to 0.90. Moreover, all composite reliability scores were above the recommended level of 0.70, all AVE values exceeded 0.50, and the square root of each construct's AVE was greater than the interconstruct correlations, thus indicating convergent and discriminant validity. Online Appendix A.5 provides an overview of correlations and psychometric properties of both constructs.

4.2.2. Structural Model and Hypothesis Testing. To estimate the structural model and test our hypotheses, we used structural equation modeling (SEM) using the lavaan package in R. We created two dummy variables for up-front and step-in HID (0 = without disclosure, 1 = with disclosure) and used them as independent variables in our model. Furthermore, we included customer communication style as an (unobserved) latent construct with three (observed) indicators: verbosity, complexity, and density. Employee workload was also modeled as a latent construct with three indicators: frequency, duration, and intensity. In addition, we included a set of variables to control for the type of service (e.g., sales, churn) and customer characteristics (device and location; see Online Appendix A.3). Finally, we added employee fixed effects to account for unobserved heterogeneity across individual employees and weekday and part-of-day fixed effects to control for time-specific trends (Altman et al. 2021). We used full information maximum likelihood estimation to use all data available. The overall fit indices of the model indicated an adequate fit to the data (CFI = 0.91, TLI = 0.86, RMSEA = 0.05, SRMR = 0.02).

The results in Table 3 show that the effect of up-front HID on customer communication style is significant ($b = 0.143$, $p < 0.001$), supporting Hypothesis 1. We also find a significant effect of step-in HID on customer communication style ($b = 0.066$, $p = 0.010$), providing support for Hypothesis 2. Unlike the two main effects, the interaction effect is not significant ($b = 0.018$, $p = 0.553$), indicating that combining up-front and step-in HID does not yield a synergistic effect that is greater than the sum of their independent effects. Comparing the standardized coefficients of the two treatments reveals that up-front HID has a stronger effect on communication style than step-in HID does. Finally, the results show a significant

Table 2. Descriptive Statistics (Field Experiment)

	No disclosure		Up-front HID only		Step-in HID only		Both up-front and step-in HID	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Customer communication style								
Verbosity	7.12	(5.43)	8.75	(5.83)	7.21	(5.12)	8.90	(5.81)
Complexity	19.35	(32.12)	21.84	(27.30)	18.31	(31.10)	21.20	(26.12)
Density	34.97	(23.28)	39.90	(20.88)	35.72	(23.30)	39.38	(20.42)
Employee workload								
Frequency	0.74	(1.00)	1.03	(1.11)	1.02	(1.43)	1.26	(1.46)
Duration ^a	0.64	(1.06)	0.88	(1.27)	0.90	(1.74)	1.05	(1.63)
Intensity	1.76	(2.61)	2.30	(3.01)	2.44	(3.64)	3.02	(4.14)
N	2,179		2,326		2,110		2,351	

Notes. $N = 8,966$. HID, human involvement disclosure.

^aAmount of time in minutes.

effect of customer communication style on employee workload ($b = 0.380, p < 0.001$), supporting Hypothesis 3. All reported effects are consistent in magnitude, direction, and significance in both the model with controls (Model 1) and the one without (Model 2). To provide a fairer comparison, we also analyzed the structural model with only the subsample of customers who had a human step in during their chat (Model 3), which yielded similar results. Finally, to examine the causal chain in our theoretical model, we performed a SEM-based mediation analysis (Cheung and Lau 2008) to formally test the indirect effects of up-front and step-in HID on employee workload through customer communication style. The results, presented in Online Appendix A.6, show that both indirect effects are significant, providing further support for our theoretical model. Taken together, our results provide evidence that customers exhibit a more human-oriented communication style when human involvement is disclosed (versus not disclosed), which in turn increases employee workload.

4.2.3. Robustness Checks. We conducted a series of robustness checks to validate our findings. Specifically, we repeated our analysis using a different analytical approach and carried out various analyses to rule out possible alternative explanations. First, to verify the robustness of our SEM-based findings, we reanalyzed the data using an econometric regression-based approach. In this analysis, we also accounted for potential estimation biases caused by the fact that a proportion of customers assigned to receive the step-in HID treatment did not actually receive it (38.5%), as an employee never had to step in during their chat. Therefore, we used the local average treatment effects (LATE) model (Imbens and Angrist 1994) to correct for the potential biases arising from this form of one-sided treatment noncompliance. Online Appendix A.7 provides details on our model specification and estimation. Consistent with our SEM-based findings and in support of Hypotheses 1–3, the results in

Tables A7 and A8 show that both up-front and step-in HID have significant positive effects on customer communication style, which in turn has a significant positive effect on employee workload. Hence, we can conclude that our findings on the impact of HID are robust and not due to the particular analytical approach chosen.

Second, a potential alternative explanation for our findings could be that it is not the disclosure of human involvement in itself that caused the observed effects on customer communication style; instead, the results could be due to the inclusion of text signaling the chatbot’s possible inferiority to humans in the treatment conditions (e.g., “If I don’t know the answer”). To rule this out, we conducted a follow-up controlled online experiment ($N = 410$) in which we added treatment conditions with neutral wording (e.g., “If the situation requires”) and compared them to our original treatment conditions. Online Appendix A.8 provides details on the experimental design and analyses. Overall, the results in Table A10 indicate no significant differences in customer communication styles between original and neutral treatment versions (all $p > 0.415$). Furthermore, we can fully replicate the results from our field experiment using our original HID treatments and the neutral versions in the additional online experiment. Collectively, these findings suggest that the effects observed in the field experiment are indeed caused by the disclosure of human involvement in itself and not by the inclusion of text signaling that the chatbot might be inferior to humans.

Third, another plausible alternative explanation for the observed results is that the differences in customer communication style were driven by differences in employee behavior when stepping in for the chatbot. With respect to this, it is important to note that all employees had received the same training and were provided with the same broad set of canned responses that were in fact the same responses that the chatbot used. Nonetheless, as a robustness check, we analyzed employees’ actions and language across experimental conditions

Table 3. Results of Structural Equation Modeling Analysis (Field Experiment)

	(1) Full sample			(2) Full sample (without control variables)			(3) Subsample with actual human involvement		
	Customer communication style	Employee workload	Standard error	Customer communication style	Employee workload	Standard error	Customer communication style	Employee workload	Standard error
<i>Up-front HID</i>	0.143*** (0.020)		(0.119)	0.138*** (0.022)		(0.134)	0.091*** (0.023)		(0.023)
<i>Step-in HID</i>	0.066* (0.026)		(0.064)	0.067* (0.028)		(0.072)	0.049* (0.026)		(0.026)
<i>Up-front HID × Step-in HID</i>	0.018 (0.030)		(0.030)	0.005 (0.034)		(0.065)	−0.022 (0.031)		(0.031)
<i>Customer communication style</i>									
Control variables									
<i>Type of service^a</i>									
<i>Service</i>	0.276* (0.119)	0.040 (0.134)							
<i>Sales</i>	0.156* (0.064)	0.019 (0.072)					0.102 (0.169)	−0.015 (0.332)	
<i>Churn</i>	0.067 (0.057)	−0.053 (0.065)					0.046 (0.094)	−0.054 (0.184)	
<i>Extension</i>	0.058 (0.049)	−0.014 (0.056)					0.049 (0.063)	−0.036 (0.122)	
<i>Contact</i>	0.069 (0.064)	0.001 (0.071)					−0.005 (0.062)	−0.026 (0.119)	
<i>Small talk</i>	0.128 (0.073)	0.196* (0.080)					0.028 (0.080)	−0.015 (0.155)	
<i>None</i>	0.393*** (0.069)	−0.020 (0.078)					−0.036 (0.114)	0.120 (0.222)	
<i>Customer device^b</i>	0.023 (0.013)	−0.050*** (0.013)					0.169 (0.113)	−0.127 (0.220)	
<i>Customer location^c</i>	0.044*** (0.012)	0.047*** (0.010)					0.015 (0.017)	−0.043* (0.017)	
Employee fixed effects	Yes	Yes					−0.003 (0.048)	0.053*** (0.014)	
Weekday fixed effects	Yes	Yes					Yes	Yes	Yes
Time-of-day fixed effects	Yes	Yes					Yes	Yes	Yes
N	8,966	8,966		8,966	8,966		5,454	5,454	5,454
R ²	0.116	0.229		0.025	0.129		0.037	0.070	

Note. HID, human involvement disclosure.

^aBaseline = other.^b1 = mobile.^c1 = Germany.* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

and individuals and found no systematic differences in their behavior (see Online Appendix A.9 for details). In addition, we created two new control variables—measuring the similarity of the text between messages from the bot and from employees and between individual employees for each chat—to account for potential variation in employees’ language in our analysis. After adding both controls to our main analysis, the results remained qualitatively similar. Taking this evidence together, it seems unlikely that our findings could be explained by variation in the language used by employees.

Fourth, a final potential alternative explanation is that when human involvement was not disclosed, customers could have been more likely to end the chat without their problems being solved. If this were the case, the differences in customer communication style could be explained by the fact that getting a problem solved in the chat might have required a more human-like communication style with longer, more complex, and more natural messages. To rule this out, we manually coded each chat to score the extent to which the customer’s problem was resolved through the interaction on a five-point scale from 1 = not at all resolved to 5 = completely resolved (see Online Appendix A.10 for details). We then compared problem resolution scores across conditions and found no significant difference (all $p > 0.468$), suggesting that problem resolution did not depend on whether human involvement was disclosed. Furthermore, we added the problem resolution score as a control variable in our main analysis and the results remained qualitatively similar. In sum, these findings allow us to rule out the possibility that the observed effects are due to customers ending the chat without their problems being solved when human involvement was not disclosed.

Finally, in our analysis, we included the step-in HID treatment as a dichotomous variable in line with our two-level treatment design (0 = without step-in HID, 1 = with step-in HID). However, treated customers could receive the step-in HID treatment more than once because it was shown every time a human employee stepped in for the chatbot. We therefore repeated our main SEM analysis using step-in HID treatment intensity—operationalized as the number of times a customer received the treatment—instead of the dichotomous treatment level. The results, presented in Online Appendix A.11, show that the impact of step-in HID on customer communication style increases with the number of times the disclosure is shown to customers ($b = 0.280$, $p < 0.001$), further corroborating our main findings.

4.2.4. Additional Analyses. Although the focus of our field experiment was to investigate the impact of HID on customer communication style and employee workload, we supplemented our core results with additional analyses exploring several other important business-related outcomes. First, an important question to examine is

whether HID increased customers’ tendency to actively seek out human involvement. To explore this, we manually reviewed all chats to identify customer messages with an explicit request for an employee and then compared the occurrence of such a request across conditions (see Online Appendix A.13 for details). Although there is no widespread tendency to seek out human involvement (96.7% of chats contain no such request), our results do suggest that customers are more likely to ask for an employee when human involvement is disclosed up front ($p < 0.001$). This finding, while preliminary, may indicate that customers are less willing to interact with a chatbot when they know from the beginning that humans are also around to help. It would be interesting for future research to study whether this aspect of customer behavior is another example of algorithm aversion (Dietvorst et al. 2015) and, if so, how a customer’s preference for humans over chatbots could be overcome when human involvement is disclosed.

Second, we extended our analysis to investigate how HID affects overall communication length. Prior literature has suggested that communication is longer when human involvement is disclosed (versus not disclosed; Luo et al. 2019). At the same time, greater efficiency in terms of addressing customer requests as quickly as possible is generally desired in customer service operations. The findings of our additional analysis on communication length, presented in Online Appendix A.14, suggest that customer–hybrid service agent interactions tend to be one minute longer on average when there is an up-front or step-in HID (both $p < 0.001$). One plausible explanation—consistent with our findings on the mediating role of customers’ impression management concerns (see next section)—is that customers may spend more time crafting messages and reading responses when they feel they are being taken care of by a human instead of a bot.

Third, customer satisfaction with a service encounter is of particular importance to firms. Because our data set did not include explicit information on customer satisfaction, we used an automated sentiment analysis approach to assess customers’ evaluations of their interactions with the hybrid service agent. The results of our additional analysis on customer sentiment, presented in Online Appendix A.15, suggest a significant negative effect of up-front HID on customer sentiment ($b = -0.011$, $p = 0.018$). There are several explanations for this finding. For example, it could be that customers are generally more likely to vent their frustration and negative emotions to a human working in tandem with a bot than to a fully automated chatbot or, alternatively, that they become frustrated more easily when being served by the chatbot while knowing that humans are also around to help. Although our data do not allow us to uncover the underlying reasons, we believe that these alternative explanations call for future research to study customer

satisfaction regarding service encounters with hybrid service agents that involve (un)disclosed employees.

Finally, to provide additional managerial insights, we analyzed the downstream financial impact of the additional workload caused by disclosing human involvement. Although exploratory in nature, our results suggest that the additional workload increases costs associated with actual human involvement in a hybrid service agent by 38–70%—potentially even more if firms need to hire and train additional employees (see Online Appendix A.16 for details).

5. Controlled Online Experiment: Mediating Role of Impression Management Concerns

Our field experiment provides real-world evidence that HID influences customer communication style in the interaction with a hybrid service agent. To reveal the underlying psychological mechanism and assess the robustness of our findings, we conducted a controlled online experiment. The objective of this experiment was threefold: (1) replicate the findings of the field experiment in a controlled setting, (2) examine impression management concerns as a mediator to the effects of up-front and step-in HID on communication style (Hypotheses 4 and 5), and (3) rule out potential alternative explanations.

5.1. Method

5.1.1. Experimental Design and Treatments. The online experiment resembled the field experiment in applying a 2 (up-front HID: disclosure versus no disclosure) \times 2 (step-in HID: disclosure versus no disclosure) between-subjects factorial design. We used the same verbal statements as in the field experiment for the up-front HID (“If I don’t know the answer to your question, my human colleague will read your message and give you an answer”) and step-in HID treatments (“One moment, please. Unfortunately, I don’t know the answer to your question. I will pass it on to my human colleague who will give you an answer shortly”). To keep the text length equal across all groups, we formulated statements of the same length for the corresponding conditions without disclosure (see Online Appendix B.1).

5.1.2. Procedure and Hybrid Service Agent Design. To enhance experimental realism, we framed the experiment as a real-world test of the new customer service platform of a telecommunications company. More specifically, we told participants that they were invited to test and provide feedback on the new platform before it would be released to the company’s customers. To make the experiment appear authentic, we explained to participants that they would receive a randomly selected service issue and that they should use this information as a starting point to contact customer service. In reality,

however, all participants received the same service issue (an unexpectedly high bill) to ensure a high level of comparability across their interactions with the hybrid service agent.

In the experiment, we first asked participants to imagine that they had just received their monthly mobile phone bill. The bill appeared to be higher than usual but did not provide an explanation for the higher costs (see Online Appendix B.1 for a full description). We chose this scenario to offer better comparison with the results of the field experiment, as billing questions and complaints are typical customer service issues for telecommunications companies. After reading the scenario description, participants summarized the problem at hand in their own words in an open-ended text box. In addition to serving as a control variable for customers’ baseline communication styles, this question also functioned as an attention check, allowing us to filter out participants who showed clear signs of inattentiveness (e.g., nonsensical answers). Next, we told participants to contact customer service to find out the reason for their unexpectedly high bill.

Subsequently, participants were randomly assigned to one of the four experimental conditions. In each condition, they interacted with a hybrid service agent in the form of a custom chatbot that was developed for this experiment using the Microsoft Bot Framework and pre-tested extensively (see Online Appendix B.2 for details). We did not involve human confederates in the experiment, aiming to avoid any potential confounding effects caused by their behavior. This helped ensure high internal validity and identical interactions across conditions. Participants entered and sent their own messages to reflect a natural and realistic interaction, also allowing us to measure their communication styles as done in the field experiment based on the text content of their messages. The chatbot used NLP to analyze text input (i.e., detect keywords in customer messages) and automatically provide predetermined responses. To ensure that chats were comparable across participants, we implemented a guided dialog that participants entered once they had clearly described their reason for contacting customer service. This dialog contained six messages with general information about the causes of an unusually high bill and two follow-up questions on details of the bill (e.g., “Which item on your bill seems to be incorrect?”). At the end of the dialog, participants learned that their unexpectedly high bill was likely due to an expired discount and the chat ended.

As in the field experiment, the welcome message at the beginning of the chat contained the up-front HID. The step-in HID was displayed once during the chat before the start of the guided dialog. To make it appear as if there were actual human involvement at this point, we delayed the subsequent message by 50 seconds, which corresponds to the average response time of

employees in the field experiment. After the experiment, participants completed a survey with measures, demographic questions, and manipulation checks. At the end of the survey, we included an open-ended question asking participants to share their opinions about HID (e.g., “Do you want to know if human employees are involved when you talk to a chatbot?”). Finally, we debriefed participants, including about the fact that they interacted only with a chatbot and that they took part in a research study.

5.1.3. Participants. We recruited participants via the online platform Clickworker. We chose this platform because it is widely used by companies to conduct usability testing and user research, making it ideal for our experiment framed as a real-world test of a new customer service platform. Of the 300 participants who completed both the chat and the survey, we excluded 12 for failing the attention check (i.e., entering a nonsensical response when asked to summarize the problem at hand). The analysis then included responses from 288 participants (61% male, $M_{\text{age}} = 40$; see Online Appendix B.3 for sample characteristics) who received €2.50 for their time, which was 15 minutes on average.

5.1.4. Measures. To assess customer communication style, we used the same measurement as in the field experiment—that is, we calculated its three indicators verbosity, complexity, and density from the text messages entered by participants during the chat. In the survey, we measured impression management concerns using four items adapted from the public self-consciousness scales of Fenigstein et al. (1975) and Govern and Marsch (2001) on a seven-point Likert scale (1 = strongly disagree to 7 = strongly agree). To control for individual differences, we collected demographic information (gender, age, education), assessed participants’ prior chatbot experience (i.e., how often they contact customer service via a chatbot) and need for human interaction (Dabholkar 1996) and calculated their baseline communication styles from their summary of the problem in the open-ended question before the chat. Online Appendix B.3 presents all measurement items and reliabilities.

5.1.5. Manipulation Check. We included two manipulation check questions at the end of the survey. First, we asked participants to indicate whether they thought that a human employee was involved in their interaction with customer service (seven-point Likert scale: 1 = strongly disagree to 7 = strongly agree). A one-way ANOVA revealed significant variation in perceived human involvement across conditions ($F(3, 284) = 51.39$, $p < 0.001$). Planned contrasts confirmed that perceived human involvement was significantly lower in the condition without HID ($M = 2.01$, $SD = 1.54$) than in the

conditions with up-front HID ($M = 3.50$, $SD = 2.22$; $t(284) = 4.74$, $p < 0.001$), step-in HID ($M = 5.36$, $SD = 1.79$; $t(284) = 10.76$, $p < 0.001$), or both up-front and step-in HID ($M = 5.23$, $SD = 1.90$; $t(284) = 10.18$, $p < 0.001$). The second manipulation check was to verify that participants knew that their direct counterpart was a chatbot and did not falsely assume they were chatting with a human masquerading as a bot. Therefore, we asked participants to indicate whether they thought that their direct counterpart was a chatbot or a human (seven-point Likert-type scale: 1 = automated chatbot to 7 = human employee; adapted from Mozafari et al. 2022). There were significant differences across the conditions ($F(3, 284) = 6.13$, $p < 0.001$). Consistent with the first manipulation check, planned contrasts showed that participants in the condition without HID perceived their counterpart significantly more as an automated chatbot ($M = 2.24$, $SD = 1.61$) than did participants in the conditions with step-in HID ($M = 3.16$, $SD = 1.55$; $t(284) = 3.28$, $p = 0.001$) or both up-front and step-in HID ($M = 3.27$, $SD = 1.90$; $t(284) = 3.62$, $p < 0.001$). The difference in the up-front HID condition was nonsignificant ($M = 2.53$, $SD = 1.75$; $t(284) = 1.03$, $p = 0.306$). Most importantly, however, all four scores were significantly lower than the scale midpoint (all $p < 0.01$), indicating that participants assumed their direct counterpart to be a chatbot (Crolic et al. 2022). In sum, these results demonstrate that our manipulations worked as intended.

5.2. Results

As in the field experiment, we employed SEM using the lavaan package in R to analyze the data and test our hypotheses. We first conducted a confirmatory factor analysis, which showed that the measurement model exhibited a good fit to the data (CFI = 0.96, TLI = 0.94, RMSEA = 0.06, SRMR = 0.06). All indicators loaded strongly on their constructs with loadings ranging from 0.60 to 0.98. Convergent validity was satisfactory, as the composite reliabilities and AVE for each construct exceeded the suggested thresholds (>0.70 and >0.50 , respectively). The square root of each AVE value was greater than all individual correlations, supporting discriminant validity. Table 4 presents descriptive statistics for each condition.

5.2.1. Effects of Human Involvement Disclosure on Customer Communication Style. To replicate the findings of the field experiment, we first constructed a baseline model. We again modeled customer communication style as a latent construct with verbosity, complexity, and density as observed indicators. We entered up-front and step-in HID as dummy-coded independent variables (0 = without disclosure, 1 = with disclosure) and included chatbot experience, need for human interaction, age, gender, and baseline communication style as control variables. The model indicated a good fit to the data

Table 4. Descriptive Statistics (Controlled Online Experiment)

	No disclosure		Up-front HID only		Step-in HID only		Both up-front and step-in HID	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Impression management concerns ^a	2.74	(1.23)	3.66	(1.58)	3.51	(1.47)	3.54	(1.35)
Customer communication style								
Verbosity	7.11	(4.08)	9.02	(5.55)	10.61	(5.70)	12.26	(7.62)
Complexity	9.73	(16.40)	15.27	(15.04)	16.17	(15.81)	16.42	(15.81)
Density	34.41	(14.18)	38.11	(12.47)	38.11	(13.36)	40.88	(13.26)
N	72		72		74		70	

Notes. N = 288. HID, human involvement disclosure.

^aMeasured on a seven-point Likert scale.

(CFI = 0.94, TLI = 0.92, RMSEA = 0.06, SRMR = 0.06). Furthermore, the results showed significant effects of up-front HID ($b = 0.177$, $p = 0.027$) and step-in HID ($b = 0.222$, $p = 0.005$) on customer communication style. As in the field experiment, the interaction effect was not significant ($b = -0.099$, $p = 0.314$). All results were qualitatively similar when control variables were not included (see Online Appendix B.4). By replicating the direct effects of HID observed in the field experiment, these results demonstrate the robustness of our findings.

5.2.2. Mediating Effect of Impression Management Concerns. To examine the mediating role of impression management concerns as proposed in Hypotheses 4 and 5, we performed a SEM-based mediation analysis using the bias-corrected bootstrapping procedure with 5,000 samples (Cheung and Lau 2008). We constructed a mediation model with communication style as the dependent variable, up-front and step-in HID as independent variables, impression management concerns as the mediator, and the same set of control variables. Overall, the model indicated a good fit to the data (CFI = 0.94, TLI = 0.92, RMSEA = 0.06, SRMR = 0.06). The results in Table 5 show that both up-front and step-in HID have significant positive effects on impression management concerns ($b = 0.310$, $p < 0.001$ and $b = 0.299$, $p < 0.001$, respectively). Moreover, the cumulative effect of up-front and step-in HID is less than the sum of their independent effects, as the interaction is negative ($b = -0.291$, $p < 0.001$). This suggests that both disclosures together do not lead to greater impression management concerns than either alone. Furthermore, impression management concerns positively influence customer communication style ($b = 0.209$, $p = 0.004$). Most importantly, the indirect effects of up-front HID ($b = 0.065$, $SE = 0.030$) and step-in HID ($b = 0.063$, $SE = 0.028$) on customer communication style through impression management concerns are significant, as the confidence intervals (CIs) of these effects do not include zero (up-front HID: 95% CI [0.019, 0.140]; step-in HID: 95% CI [0.018, 0.131]). Furthermore, the direct effect of up-front HID on communication style

becomes nonsignificant when the mediator is included ($p = 0.162$, indicating full mediation), whereas the direct effect of step-in HID remains significant ($p = 0.044$, indicating partial mediation; Zhao et al. 2010). One possible explanation for this is that once a (disclosed) human employee steps in for the chatbot (i.e., step-in HID), other mechanisms may also come into play (e.g., the employee's helpfulness). All results were qualitatively similar when control variables were not included (see Online Appendix B.4). In sum, these results provide evidence for the mediating effect of impression management concerns, supporting Hypotheses 4 and 5.

5.2.3. Additional Analyses. We conducted several additional analyses to examine potential alternative mechanisms. However, none of these mechanisms, including customers' tendencies to actively seek out human involvement and social cognition perceptions (i.e., perceived competence and warmth), could explain our findings (see Online Appendix B.5). This provides further support for the role of impression management concerns as the psychological mechanism underlying the effect of HID. We also performed a content analysis of participants' responses to the open-ended question on whether they would want to know about employees working in tandem with a chatbot (see Online Appendix B.6); most participants indeed indicated that they would want to be informed about human involvement. Furthermore, participants reported that they would feel fooled by and lose trust in a firm if they found out later that they were deliberately kept in the dark.

6. Discussion

The proliferation of hybrid service agents adds a new layer of complexity to the already blurred line between humans and technology in online service encounters. While ethical and legal reasons have pushed most firms to disclose their chatbots' nonhuman identities, it is less clear whether firms should also be transparent about behind-the-scenes employees who step in if a chatbot is unable to respond. Against this backdrop,

Table 5. Results of Structural Equation Modeling Analysis (Controlled Online Experiment)

	Impression management concerns		Customer communication style	
	Coefficient	Standard error	Coefficient	Standard error
<i>Up-front HID</i>	0.310***	(0.083)	0.116	(0.083)
<i>Step-in HID</i>	0.299***	(0.082)	0.165*	(0.082)
<i>Up-front HID × Step-in HID</i>	−0.291**	(0.101)	−0.040	(0.099)
<i>Impression management concerns</i>			0.209**	(0.077)
Control variables				
<i>Baseline communication style</i>	−0.056	(0.060)	0.029	(0.057)
<i>Chatbot experience</i>	0.186**	(0.060)	−0.183**	(0.058)
<i>Need for human interaction</i>	0.148	(0.063)	0.005	(0.060)
<i>Age</i>	−0.253***	(0.059)	0.212***	(0.059)
<i>Gender^a</i>	−0.045	(0.061)	0.169**	(0.056)
<i>R²</i>	0.215		0.158	
Hypothesized indirect effects			Effect (standard error)	Bootstrapped 95% confidence interval
Up-front HID → impression management concerns → customer communication style			0.065 (0.030)	[0.019, 0.140]
Step-in HID → impression management concerns → customer communication style			0.063 (0.028)	[0.018, 0.131]
				Hypotheses
				H4: supported
				H5: supported

Notes. *N* = 288. HID, human involvement disclosure.

^a1 = female.

p* < 0.05; *p* < 0.01; ****p* < 0.001.

we investigated the impact of HID on customer interactions with hybrid service agents. Consistent evidence from a randomized field experiment and a controlled online experiment suggests that HID substantially influences how customers communicate with a hybrid service agent. In line with audience design theory, we find that when human involvement is disclosed, customers exhibit a more human-oriented communication style, characterized by longer, more complex, and more natural messages rather than simple keyword-style queries. Interestingly, our results suggest that a disclosure before the interaction (up-front HID) has a similar or even stronger effect than a disclosure during the interaction when an employee steps in (step-in HID), even though the up-front HID does not guarantee that an employee will eventually become involved. Furthermore, we reveal that the effect of HID on customer communication style is driven by customers' impression management concerns. That is, customers are more concerned about making a good impression when they know that humans are working in tandem with the chatbot, which then affects how they communicate with the hybrid service agent. Finally, our results demonstrate that these changes in communication style have negative downstream consequences, as they increase the workload of employees. In addition to our main results, we also find that when human involvement is disclosed before the interaction, customers are more likely to actively seek out human involvement and

tend to express more negative sentiment. Furthermore, customer–hybrid service agent interactions tend to be longer when human involvement is disclosed. Taken together, our findings highlight the major impact that HID has on customer–hybrid service agent interactions.

6.1. Implications for Research

Our work makes three contributions to research. First, we contribute to the literature on the role of IT in customer service encounters. While prior research has primarily focused on IT-based self-service (e.g., web portals, chatbots) and human-based service (e.g., phone, live chat) (Ba et al. 2010, Kumar and Telang 2012, Schanke et al. 2021, Han et al. 2022), our study sheds light on a novel approach that combines humans and chatbot technology behind a single service interface: hybrid service agents. In particular, our empirical investigation of customer–hybrid service agent interaction adds to prior literature by revealing notable differences in customer behavior when customers know that their counterpart is not just a chatbot but that there are also humans around to help. Even when there is no guarantee that a human will eventually step in for the chatbot, the mere disclosure of potential human involvement can alter customer behavior. Our research thus offers important insights into how customers respond to hybrid service interfaces that blur the once clear lines between humans and technology in customer service encounters.

Second, we contribute to the emerging stream of IS research on the use of AI in service automation. While current research highlights the various benefits of using AI to augment, rather than replace, service employees (Jain et al. 2021, Schuetzler et al. 2021), there is limited understanding of the obstacles that can hamper effective human–AI collaboration (Jussupow et al. 2021). Against this backdrop, our study sheds light on how unexpected changes in customer behavior can affect AI-to-human delegation mechanisms, place additional workload on human employees working in tandem with AI, and create additional costs for firms. More broadly, these insights contribute to IS delegation research (Baird and Maruping 2021, Fügener et al. 2022) by showing that delegation mechanisms are not only shaped by the human and the AI agent themselves but can also be deliberately or accidentally influenced by external actors (here, customers). Overall, our key addition to the literature is the notion that being transparent about human involvement in predominantly AI-based service encounters may ultimately lead to a lower degree of automation and therefore undermine AI’s ability to free up employees from mundane and repetitive customer service work.

Third, we contribute to the HCI literature on understanding the psychological mechanisms involved in customer interactions with human, nonhuman, and hybrid agents. Previous research has primarily focused on competence- and warmth-related perceptions and expectations as key drivers of customer behavior (Go and Sundar 2019, Luo et al. 2019, Grimes et al. 2021, Mozafari et al. 2022). However, these mechanisms were unable to explain our results. Instead, our study reveals a different psychological mechanism behind customer behavior: impression management concerns. Although such concerns are well known from traditional human–human interactions (Argo and Dahl 2020), our study provides novel insights into how similar concerns arise when customers become aware of human employees working in tandem with a chatbot. These insights contribute to HCI literature by revealing that customers’ behavior may sometimes depend more on how they perceive themselves than how they perceive their counterparts (e.g., a chatbot, human, or both).

6.2. Implications for Practice

Our research also provides several implications for managers and policymakers. First, our findings suggest that firms face somewhat of a dilemma when deciding whether to disclose human involvement in customer–hybrid service agent interactions. On the one hand, a disclosure triggers undesirable changes in customer behavior and adversely affects employees’ workloads. This not only undermines the hybrid service agent’s effectiveness in automating routine customer interactions and creates additional costs (see Online Appendix A.16) but

could also disrupt customer service operations if too much work is delegated to a human. On the other hand, our qualitative findings indicate that customers demand transparency from firms employing hybrid service agents. Many customers want to know about the involvement of human employees and some even report feeling betrayed and losing trust in a firm if they only find out at a later stage (see Online Appendix B.6). These findings resonate with criticism surrounding the use of so-called pseudo-AI in the popular press (Forbes 2020) and suggest that firms should be aware of the negative backlash that may stem from deliberately avoiding being transparent about human involvement. Given that disclosing human involvement may follow chatbot identity disclosure to become law at some point, we suggest two actions that managers can take to mitigate its negative consequences. Initially, managers should improve the ability of their hybrid service agents to handle a human-oriented customer communication style by training their chatbots to better understand the nuances of human language, such as semantics and pragmatics. Furthermore, before going live with the disclosure, they should temporarily increase the headcount of employees who can work with the core team to absorb higher workloads during the transition phase. As these solutions may not always be feasible or economically viable, managers could also consider some easy-to-implement strategies. For instance, including a statement such as “Don’t worry, my human colleagues don’t mind if you communicate with them like a robot” in the welcome message of the hybrid service agent could prevent impression management concerns from arising. Another strategy could be to limit the length of customer messages to a maximum number of characters or words, encouraging customers to use simpler keyword queries instead of more human-oriented language.

In addition to these managerial insights, our findings carry implications for policymakers. As the previous discussion suggests, firms must currently decide for themselves what is right and wrong in terms of disclosure when using hybrid service agents. Will they disclose human involvement even if it leads to higher employee workload and costs? Maybe not. What is missing is a regulatory framework that balances economic interests and ethical concerns and provides guidance for firms on what is and is not allowed. Although research has emphasized that the blurring line between humans and AI creates counterfeit service encounters and raises serious transparency issues (Robinson et al. 2020), most policy initiatives to date exist only as nonlegally binding principles and guidelines. The one notable exception (California’s BOT bill) only requires the disclosure of chatbot identity and thus does not regulate whether, when, and to what extent firms should disclose that humans are working in tandem with their chatbot. There is therefore a need for policymakers to review and extend existing regulatory frameworks to protect

customers from counterfeit service encounters while also minimizing potential negative impacts for firms (e.g., additional costs). Given the clear parallels to mandatory call recording disclosures (“This call may be monitored or recorded”), existing telephone call recording laws may serve as a source of inspiration for future policy initiatives.

6.3. Limitations and Future Research

This work has certain limitations that offer several opportunities for future research. First, we primarily focused on understanding hybrid service agent interactions from the customer perspective. Although we examined how customer behavior affects the workload of employees working in tandem with the chatbot, future research should place greater emphasis on the employee perspective. It would be particularly valuable to better understand how employees feel about the (non)disclosure of their involvement, as their work may remain completely hidden from customers. In addition, future research should investigate the nature and consequences of human–AI collaboration in hybrid service agents on a more general level. For example, such studies could draw from emerging research on algorithmic management (Möhlmann et al. 2021) and AI-to-human delegation (Fügner et al. 2022) to examine how employees cope with the loss of control when an AI algorithm decides whether and when they have to step in during customer interactions.

A second fruitful avenue for further research could be to extend our work with a broader investigation of customer communication behavior. As our analysis focused on structural characteristics of communication style in text-based customer service interactions (i.e., verbosity, complexity, and density), additional research is needed to investigate other relevant language aspects (e.g., level of concreteness), service contexts (e.g., medical consultations), and communication channels (e.g., voice). For instance, it could be interesting to explore whether different types of service interactions (e.g., enquiry-based versus complaint-based conversations) activate stronger impression management concerns and consequently also lead to a more human-oriented communication style. In addition, as chatbots are often enriched with social cues, such as human names and humanlike appearances (Feine et al. 2019), it would be valuable to explore how these cues affect customer communication behavior.

Third, our research opens the door to further investigation of customer perceptions of the service encounter, the firm, and the hybrid service agent itself. Although we provide a detailed understanding of customer communication style and some additional insights into other customer outcomes (e.g., tendency to seek out human involvement, sentiment), there remains a need for research on how customers evaluate service encounters with hybrid service agents and how the disclosure of human involvement can affect their evaluations. For example, our results could serve as a starting point for

investigating whether such disclosure causes or exacerbates algorithm aversion (Dietvorst et al. 2015) and whether customers are even more frustrated by failures during an interaction when they know that humans are working in tandem with the chatbot. Moreover, as our qualitative insights suggest that customers want to know about human involvement (see Online Appendix B.6), future research could more systematically investigate the role of nontransparency and how customers react when they find out that a firm has deliberately kept them in the dark.

Finally, it is important to discuss our findings in light of the continuous advances in AI. As the technology behind chatbots is constantly improving, there may be a point in time when human involvement is no longer needed to provide quality customer service. Although the chatbot in our field experiment was based on a commercial product from one of the market leaders in this field, used state-of-the-art intent recognition algorithms, and was extensively trained and continuously optimized by the telecommunications company, we acknowledge that the observed effects of customer communication style on employee workload may fade as the technology improves. With the gradual emergence of chatbots based on large language models such as ChatGPT or Bard, it would be interesting to explore how they might reduce the need for human involvement in certain areas while increasing it in others. For example, as large language model-based chatbots may generate incorrect but plausible-looking answers (“hallucinations”), future research could investigate how and when employees should step in to review an answer and verify the accuracy of the information before sending it to the customer.

In conclusion, our research takes a first step toward understanding customer–hybrid service agent interactions and the impact of disclosing human involvement to customers. We hope it offers thought-provoking insights for researchers, managers, and policymakers alike and can stimulate more work in this promising area of research.

Acknowledgments

The authors thank the senior editor, associate editor, and three anonymous reviewers for constructive comments and guidance throughout the review process. The authors are also grateful to Matthias Gnewuch and seminar participants at HEC Montréal for helpful feedback on earlier versions of this paper. The authors also thank their industry partner for support. Any errors are the responsibility of the authors.

Endnotes

¹ We use the term *hybrid service agent* to refer to the interface that customers directly interact with in their service encounters. Furthermore, we use the term *employees* to refer to the (un)disclosed human service employees who work in tandem with the AI-based chatbot to serve customers.

² The results of the manipulation checks in the controlled online experiment show that customers strongly believed they were interacting with an automated chatbot when human involvement was not disclosed.

References

- Adam M, Roethke K, Benlian A (2022) Human vs. automated sales agents: How and why customer responses shift across sales stages. *Inform. Systems Res.*, ePub ahead of print November 10, <https://doi.org/10.1287/isre.2022.1171>.
- Adam M, Wessel M, Benlian A (2020) AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets* 9(2):1–19.
- AI HLEG (2019) Ethics guidelines for trustworthy AI. Retrieved January 10, 2021, <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines>.
- Altman D, Yom-Tov GB, Olivares M, Ashtar S, Rafaeli A (2021) Do customer emotions affect agent speed? An empirical study of emotional load in online customer contact centers. *Manufacturing Service Oper. Management* 23(4):854–875.
- Argo JJ, Dahl DW (2020) Social influence in the retail context: A contemporary review of the literature. *J. Retailing* 96(1):25–39.
- Ba S, Stallaert J, Zhang Z (2010) Balancing IT with the human touch: Optimal investment in IT-based customer service. *Inform. Systems Res.* 21(3):423–442.
- Baird A, Maruping LM (2021) The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *Management Inform. Systems Quart.* 45(1):315–341.
- Beaver I, Freeman C, Mueen A (2020) Toward awareness of human relational strategies in virtual agents. *Proc. AAAI Conf. on Artificial Intelligence* (AAAI Press, Palo Alto, CA), 2602–2610.
- Bell A (1984) Language style as audience design. *Language Soc.* 13(2):145–204.
- Berger J, Humphreys A, Ludwig S, Moe W, Netzer O, Schweidel D (2020) Uniting the tribes: Using text for marketing insight. *J. Marketing* 84(1):1–25.
- Brown SA, Fuller R, Thatcher S (2016) Impression formation and durability in mediated communication. *J. Assoc. Inform. Systems* 17(9):614–647.
- Castelo N, Boegershausen J, Hildebrand C, Henkel AP (2023) Understanding and improving consumer reactions to service bots. *J. Consumer Res.* Forthcoming. <https://doi.org/10.1093/jcr/ucad023>.
- Castillo D, Canhoto AI, Said E (2021) The dark side of AI-powered service interactions: Exploring the process of co-destruction from the customer perspective. *Service Industry J.* 41(13–14):900–925.
- 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 Res.* 32(2):496–517.
- Cheung GW, Lau RS (2008) Testing mediation and suppression effects of latent variables. *Organ. Res. Methods* 11(2):296–325.
- Coleman EB (1971) Developing a technology of written instruction: Some determiners of the complexity of prose. Rothkopf E, Johnson P, eds. *Verbal Learning Research and the Technology of Written Instruction* (Teachers College Press, Columbia University, New York), 155–204.
- Crolic C, Thomaz F, Hadi R, Stephen AT (2022) Blame the bot: Anthropomorphism and anger in customer-chatbot interactions. *J. Marketing* 86(1):132–148.
- Dabholkar PA (1996) Consumer evaluations of new technology-based self-service options: An investigation of alternative models of service quality. *Internat. J. Res. Marketing* 13(1):29–51.
- De Keyser A, Köcher S, Alkire L, Verbeeck C, Kandampully J (2019) Frontline service technology infusion: Conceptual archetypes and future research directions. *J. Service Management* 30(1):156–183.
- Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. *J. Experiment. Psych. General* 144(1):114–126.
- Feine J, Gnewuch U, Morana S, Maedche A (2019) A taxonomy of social cues for conversational agents. *Internat. J. Human Comput. Stud.* 132:138–161.
- Fenigstein A, Scheier MF, Buss AH (1975) Public and private self-consciousness: Assessment and theory. *J. Consulting Clinical Psych.* 43(4):522–527.
- Følstad A, Nordheim CB, Bjørkli CA (2018) What makes users trust a chatbot for customer service? An exploratory interview study. *Proc. 5th Internat. Conf. on Internet Sci.* (Springer, Cham, Switzerland), 194–208.
- Forbes (2020) Artificial or human intelligence? Companies faking AI. Retrieved January 10, 2021, <https://www.forbes.com/sites/cognitiveworld/2020/04/04/artificial-or-human-intelligence-companies-faking-ai>.
- Fügener A, Grahl J, Gupta A, Ketter W (2022) Cognitive challenges in human-artificial intelligence collaboration: Investigating the path toward productive delegation. *Inform. Systems Res.* 33(2):678–696.
- Glikson E, Woolley AW (2020) Human trust in artificial intelligence: Review of empirical research. *Acad. Management Ann.* 14(2):627–660.
- Gnewuch U, Morana S, Adam MTP, Maedche A (2022) Opposing effects of response time in human-chatbot interaction. *Bus. Inform. Systems Engrg.* 64(6):773–791.
- Go E, Sundar SS (2019) Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Comput. Human Behav.* 97:304–316.
- Govern JM, Marsch LA (2001) Development and validation of the situational self-awareness scale. *Conscious Cognition* 10(3):366–378.
- Grimes GM, Schuetzler RM, Giboney JS (2021) Mental models and expectation violations in conversational AI interactions. *Decision Support Systems* 144:113515.
- Han E, Yin D, Zhang H (2022) Bots with feelings: Should AI agents express positive emotion in customer service? *Inform. Systems Res.*, ePub ahead of print December 2, <https://doi.org/10.1287/isre.2022.1179>.
- Hill J, Ford WR, Farreras I (2015) Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Comput. Human Behav.* 49:245–250.
- Huang YS, Dootson P (2022) Chatbots and service failure: When does it lead to customer aggression. *J. Retailing Consumer Services* 68:103044.
- Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. *Econometrica* 62(2):467–475.
- Jain H, Padmanabhan B, Pavlou PA, Raghu TS (2021) Editorial for the special section on humans, algorithms, and augmented intelligence: The future of work, organizations, and society. *Inform. Systems Res.* 32(3):675–687.
- Jussupow E, Spohrer K, Heinzl A, Gawlitza J (2021) Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Inform. Systems Res.* 32(3):713–735.
- Knijnenburg BP, Willemsen MC (2016) Inferring capabilities of intelligent agents from their external traits. *ACM Trans. Interactive Intelligence Systems* 6(4):1–25.
- Kumar A, Telang R (2012) Does the web reduce customer service cost? Empirical evidence from a call center. *Inform. Systems Res.* 23(3):721–737.
- Leary MR, Kowalski RM (1990) Impression management: A literature review and two-component model. *Psych. Bull.* 107(1):34–47.
- Luo X, Tong S, Fang Z, Qu Z (2019) Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Sci.* 38(6):913–1084.

- Möhlmann M, Zalmanson L, Henfridsson O, Gregory RW (2021) Algorithmic management of work on online labor platforms: When matching meets control. *Management Inform. Systems Quart.* 45(4):1999–2022.
- Mou Y, Xu K (2017) The media inequality: Comparing the initial human-human and human-AI social interactions. *Comput. Human Behav.* 72:432–440.
- Mozafari N, Weiger WH, Hammerschmidt M (2022) Trust me, I’m a bot: Repercussions of chatbot disclosure in different service frontline settings. *J. Service Management* 33(2):221–245.
- Pennebaker JW, Booth RJ, Boyd RL, Francis ME (2015) Linguistic inquiry and word count. Accessed August 21, 2020, www.LIWC.net.
- Rai A, Constantinides P, Sarker S (2019) Editor’s comments: Next-generation digital platforms: Toward human–AI hybrids. *Management Inform. Systems Quart.* 43(1):iii–ix.
- Robinson S, Orsingher C, Alkire L, De Keyser A, Giebelhausen M, Papamichail K, Shams P, et al. (2020) Frontline encounters of the AI kind: An evolved service encounter framework. *J. Bus. Res.* 116:366–376.
- Rosseel Y (2012) lavaan: An R package for structural equation modeling. *J. Statist. Software* 48(2):1–36.
- Schanke S, Burtch G, Ray G (2021) Estimating the impact of “humanizing” customer service chatbots. *Inform. Systems Res.* 32(3):736–751.
- Schuetzler RM, Grimes GM, Rosser HK (2021) Deciding whether and how to deploy chatbots. *MIS Q. Executive* 20(1):1–15.
- Shechtman N, Horowitz LM (2003) Media inequality in conversation: How people behave differently when interacting with computers and people. *Proc. CHI Conf. on Human Factors in Comput. Systems* (Association for Computing Machinery, New York), 281–288.
- State of California (2018) Senate Bill No. 1001. Chapter 89. Retrieved January 10, 2021, https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=20170180SB1001.
- Xu JD, Benbasat I, Cenfetelli RT (2014) The influences of online service technologies and task complexity on efficiency and personalization. *Inform. Systems Res.* 25(2):420–436.
- Yoon SO, Brown-Schmidt S (2019) Audience design in multiparty conversation. *Cognition Sci.* 43(8):1–28.
- Youssef V (1993) Children’s linguistic choices: Audience design and societal norms. *Language Soc.* 22(2):257–274.
- Zhao X, Lynch J, Chen Q (2010) Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *J. Consumer Res.* 37(2):197–206.