

IN PRESS, JOURNAL OF THE ASSOCIATION FOR CONSUMER RESEARCH

Automation in Marketing and Consumption- Volume 9, Issue 3

Imperfectly Human:

The Humanizing Potential of (Corrected) Errors in Text-Based Communication

Shirley Bluvstein*, Xuan Zhao*, Alixandra Barasch, Juliana Schroeder

Shirley Bluvstein* - *Assistant Professor*, Yeshiva University, Sy Syms School of Business
(shirley.bluvstein@yu.edu)

Xuan Zhao* - *Research Scientist*, Stanford University, Department of Psychology
(xuanzhao@stanford.edu)

Alixandra Barasch- *Associate Professor*, University of Colorado Boulder, Leeds School of Business (alix.barasch@colorado.edu)

Juliana Schroeder- *Associate Professor*, University of California Berkeley Haas School of Business (jschroeder@haas.berkeley.edu)

Authors' note:

* The first two authors contributed equally. Correspondence concerning this manuscript should be addressed to Xuan Zhao (xuanzhao@stanford.edu).

ABSTRACT

Today more than ever before, online text-based interactions have become a common means of communication between consumers and companies. But with the advent of AI-powered chatbots, customers sometimes struggle to ascertain the humanness of their online interaction partners (e.g., customer service agents). The current research investigates the humanizing potential of one common feature in text communication—typographical errors (“typos”). Across five experiments reported in the main text, two supplemental experiments, and pilot data (total N = 3,399), participants perceived customer service agents who made and subsequently corrected a typo to be more human—and more helpful—than agents who made no typos or made but did not correct a typo. These findings provide novel insights into how conversational features influence customers’ perceptions of online agents. In an era where AI frequently surpasses human performance in a variety of domains, consumers may perceive the act of making (and correcting) errors to be a hallmark of humanness.

Keywords: humanization, chatbot, artificial intelligence, error, communication, customer service

Imperfectly Human:

The Humanizing Potential of (Corrected) Errors in Text-Based Communication

Over the past few decades, two technological inventions have fundamentally reshaped how consumers connect with companies: The Internet has enabled people to communicate with other humans via text on screens instead of by looking directly into each other's eyes, and humanlike chatbots and conversational Artificial Intelligence (AI) technologies have enabled people to communicate with machine algorithms instead of only with other humans. Due to these two inventions, many companies are now able to provide online chat options to customers who want to obtain information or assistance in real time. Although automating customer service via chatbots promises efficiency and scalability, potentially yielding faster service for customers and financial returns for companies (Reddy 2017), consumers may prefer to receive service from human agents because they assume that a human can better empathize with them, understand their issues, and ultimately solve their problems (LivePerson 2019).

While *being* human is a biological fact, *perceiving* humanness is a psychological process. Previous research has examined how the aesthetics or content of text can influence perceived humanness. However, little is known on how the *process* of writing — that is, the way in which written communication dynamically unfolds in real time — can impact perception. This paper seeks to fill this gap and demonstrate a novel heuristic for perceiving humanness that arises from the act of writing: making, and subsequently correcting, typographical errors (“typos”). Our work thus highlights a historical inflection point in the relationship between machines and humans: For decades, the frontier of developing more humanlike machine intelligence was to

make machines “smarter” and *less* prone to errors; but now, with increasingly sophisticated AI, it may be that *more* error-prone machines can seem more human in certain contexts.

Perceiving Humanness

Because online conversations are often written, they inherently lack the “richness” of other types of human interactions, meaning that consumers have reduced ability to assess their chat partners’ human (or non-human) identities and mental capacities (Kiesler, Siegel, and McGuire 1984; Pinker and Bloom 1990; Schroeder and Epley 2015; Schroeder, Kardas, and Epley 2017). Prior research has examined humanizing cues in social interaction such as appearance, speed of movement, voice, and identity (Looser and Wheatley 2010; Morewedge, Preston, and Wegner 2007; Schroeder and Epley 2016; Zhao, Phillips, and Malle 2019), all of which are largely absent from text.

Within the scope of text-based communication, prior work has focused primarily on two sets of humanizing cues. First, the semantic content of what a person writes—such as the words they choose to use—can provide signals of humanness. This includes both the extent to which a communicator expresses coherent, sophisticated, and relevant thought (Christian 2012; Lortie and Guitton 2011), as well as the linguistic features and social desirability of what they express (McCoy and Ullman 2018). For instance, when guessing whether content was generated by humans or AI, people tend to associate first-person pronouns, spontaneous self-expression, or family topics with humans (Jakesch et al. 2023). Second, people infer humanness by considering the style in which content is communicated. For instance, textual paralinguistics—written manifestations of audible, tactile, and visual elements that mimic nonverbal cues in face-to-face interaction (e.g., exclamation points, emojis, handwritten-like typefaces, and vocalizations)—has

been shown to humanize communicators or products (Candello, Pinhanez, and Figueiredo 2017; Luangrath, Peck, and Barger 2017; Schroll, Schnurr, and Grewal 2018).

Notably, the majority of this prior work has focused on how static text — the final product of a communication effort—influences perception of humanness of the communicator. What is missing in this research is an understanding of how the very process of generating text communication can humanize a communicator. Indeed, most written conversations with customer service agents are now synchronous and instant (e.g., online chat) rather than asynchronous (e.g., email). This provides unprecedented access into the mind of the communicator, because consumers can watch each line of conversation unfold and get insight into how the communicator’s thoughts are getting translated into language. This opens a new and important research space of investigating how the *act* of writing can reveal (or conceal) humanness.

To Err is Human? To Correct, Divine

Making errors is not always seen as a positive, humanizing behavior. Although common wisdom suggests that “to err is human,” mistakes are often associated with “stupid” and “ignorant” AI systems not seen as measuring up to human standards. Examining how these seemingly contradictory intuitions are manifested in consumers’ perception of online chat agents promises both theoretical clarity and practical relevance.

On the one hand, it is possible that errors will be humanizing and create positive social perceptions. For instance, previous research findings (dubbed the “Pratfall Effect”) have shown that making mistakes can, under certain conditions, increase a person’s likeability (e.g., Aronson, Willerman, and Floyd 1966; Helmreich, Aronson, and LeFan 1970). Moreover, consumers sometimes prefer products that were made by mistake to products made intentionally (Reich,

Kupor, and Smith 2017). Akin to how human (vs. machine) labor can seem more attractive because it is perceived to contain sentimental value (Fuchs et al. 2015) and to be unique (Granulo et al. 2021), errors may be viewed fondly as humanizing pratfalls in certain contexts. For instance, hitting a wrong button on a keyboard seems to be a common human experience to which most people can relate (Norman 2004). Capitalizing on this intuition, some programmers inserted typos in their programs in an attempt to make their chatbots appear more humanlike in the annual Turing Test (i.e., the Loebner Prize; Christian 2012).

However, despite its intuitive appeal, the potentially humanizing effect of typos so far lacks empirical support. In fact, other research has found that typos consistently lead to negative social perceptions. For example, in one set of studies, people perceived a communicator who made typos to be less socially attractive and less intelligent than a communicator who did not make typos (Westerman, Cross, and Lindmark 2018). In other studies, participants rated individuals who produced spelling and grammatical errors in emails or essays as less competent, less trustworthy, and as having weaker mental capacities (Figueredo and Varnhagen 2005; Kreiner et al. 2002; Lea and Spears 1992; Vignovic and Thompson 2010).

In light of these opposing findings, we contend that an error is neither inherently humanizing nor dehumanizing, but what happens *after* the error is key. In experiments documenting the negative impact of errors, the errors were presented as part of a final product—such as an email or an essay—and were never corrected (e.g., Figueredo and Varnhagen 2005; Lea and Spears 1992; Vignovic and Thompson 2010); by contrast, studies showing positive effects of errors almost always presented errors within a stream of actions, wherein the actors either corrected or at least acknowledged their blunders as the interaction unfolded (e.g., Aronson et al., 1966; Mirnig et al., 2017). Inspired by this critical difference, we propose that

written typos during a text-based interaction may humanize a communicator when they are *subsequently corrected* by the communicator.

We believe that correcting an error in written text can signal humanness for at least two reasons. First, correcting one's mistake signals that the communicator is actively monitoring, assessing, and reacting to each moment of the ongoing conversation. Even though the initial occurrence of an error may result from an absent or inattentive mind, correcting one's error shows the mind is engaged and intentional. Second, correcting one's mistake also signals that one cares about how one is being perceived. Therefore, communicators who correct their own errors signal that they have social awareness and a desire to fix their errors so as to avoid being perceived as mentally "inferior." By contrast, people tend to believe that machines lack the capacity for such self-awareness or conscious intent (Gray, Gray, and Wegner 2007) and do not care how they will be perceived by others (Boden 2018). Drawing on this reasoning, we hypothesize that people will perceive an online chat agent who makes and subsequently corrects their typographical errors as more human than an agent who does not make errors, and also as more human than an agent who makes but does not correct their errors.

This novel hypothesized process for perceiving humanness suggests a modern "reversal" of the Turing Test (Turing 1950), such that today's computer programs are becoming so sophisticated, they may now appear more human by simulating humanlike fallibility rather than by being perfect. However, our intent in conducting this research is *not* to provide companies with a tool to manipulate customers, but instead to facilitate new scientific insight about the heuristics consumers use to assess humanness. We hope such understanding can be valuable for consumers and policymakers in an era where company interactions increasingly occur in a virtual space (see General Discussion).

OVERVIEW OF STUDIES

We test our predictions across five experiments and two supplemental experiments with a variety of research paradigms. Studies 1A – 1C each test our primary hypothesis that corrected typos, when observed in text-based online chats that unfold over time, will humanize a customer service agent (compared to no typo). To test the critical role of correcting one’s error in humanizing the agent, Study 2 includes an additional, typo-only condition in which the chat agent’s typos were left unaddressed. Finally, Study 3 investigates the moderating role of agent identity: by either keeping the agent’s identity ambiguous or declaring it as a chatbot or a human, this study reveals that the humanizing effect of corrected typos is stronger when the agent’s identity is ambiguous, but still persists when its identity is known. Critically, this finding suggests that it is possible for firms to implement these humanizing cues while still being transparent to consumers.

For all studies, we report all data exclusions, manipulations, and measures. Due to space constraints, we report some of our preregistered analyses in the Web Appendix (but still list all measured variables within our first study). All studies except for Supplemental Study 2 were preregistered. Stimuli, measures, data, and preregistration forms are available on the Open Science Framework (<https://osf.io/6s8wm>).

PILOT DATA: LAY BELIEFS ABOUT TYPOS

Before testing our primary hypotheses, we wondered whether people would intuitively recognize the value of typos in customer service interactions. Given that typos can be seen as negative or undesirable features in human-to-human interactions, we predicted that people may

hold the *expectation* that typos will harm customer service interactions and fail to recognize their potential humanizing value. A pilot study supported this prediction (see Web Appendix for full description of procedure and results), showing that participants ($N = 403$) indeed assumed that customers would be less interested in engaging with customer service agents who made and corrected typos ($M = 3.80$, $SD = 1.53$), than those who did not make typos ($M = 5.76$, $SD = 1.03$; $F(1, 393) = 233.89$, $p < .001$, $\eta^2 = .373$). Moreover, only 18% of the participants reported that they could think of a situation where typos might be beneficial, which was significantly lower than the chance level of 50%, $\chi^2 = p < .001$.

STUDY 1A: CAN TYPOS HUMANIZE A CHATBOT?

In Study 1A, participants interacted with a pre-programmed chatbot during a live conversation. We created this dynamic chatbot paradigm to enable a prolonged, real-time engagement between participants and the chat agent, ensuring the psychological realism of our method and the ecological validity of our findings.

Method

Participants and design. This experiment employed a between-subjects design with two conditions (corrected-typo, uncorrected-typo). Four hundred participants on Prolific completed the study for \$2.00 each. According to our pre-registered exclusion criteria, we excluded 14 participants who either provided nonsensical and irrelevant responses or inadvertently asked questions that revealed the program's technological limitations, and one participant due to suspicious spam activity, resulting in 385 participants in the final analysis ($M_{\text{age}} = 33.90$, 53.2% female).

Chat platform. Instead of deploying marketplace AI-based chatbots, we developed our own rule-based chatbot so that we could precisely manipulate the presence or absence of typos. Learning from experts who wrote scripts for commercially successful personal assistant products (e.g., Siri and Alexa), we created a script for the chatbot in the role of a research assistant named Angela (selected because it was a common female name), who interviewed the participant with a list of personal questions. The program detected keywords in participants' responses and provided minimal, pre-determined responses to convey listening and understanding (see example chat in Figure 1; full script and computer codes available in our OSF folder).¹

Procedure. We informed participants that this study was part of a nationwide project on participant well-being during COVID-19, and that they would be connected to a “research agent” via a chat platform who would collect their information. Next, participants saw a loading animation and the sentence, “You will be matched with a research agent in a moment.” Participants were randomly assigned to either the typo or the no-typo condition. In both conditions, they were first greeted by the chat agent, who introduced herself. In the typo condition, the agent made and corrected a typo (“helo /“*help””; see Figure 1), whereas in the no-typo condition, Angela did not make a typo. Next, Angela provided a brief task instruction and confirmed that the participant was ready to start. To strengthen the typo manipulation, we included a second typo and correction in Angela's first question (“talking” / “*taking””; see Figure 1). No further typos were present for the rest of the chat. After receiving a response, Angela continued with four questions designed to elicit self-disclosure (e.g., “What's one big goal in your life right now?”). Finally, Angela concluded the chat by indicating that the time was up, thanked the participant, and instructed them to proceed to the next page.

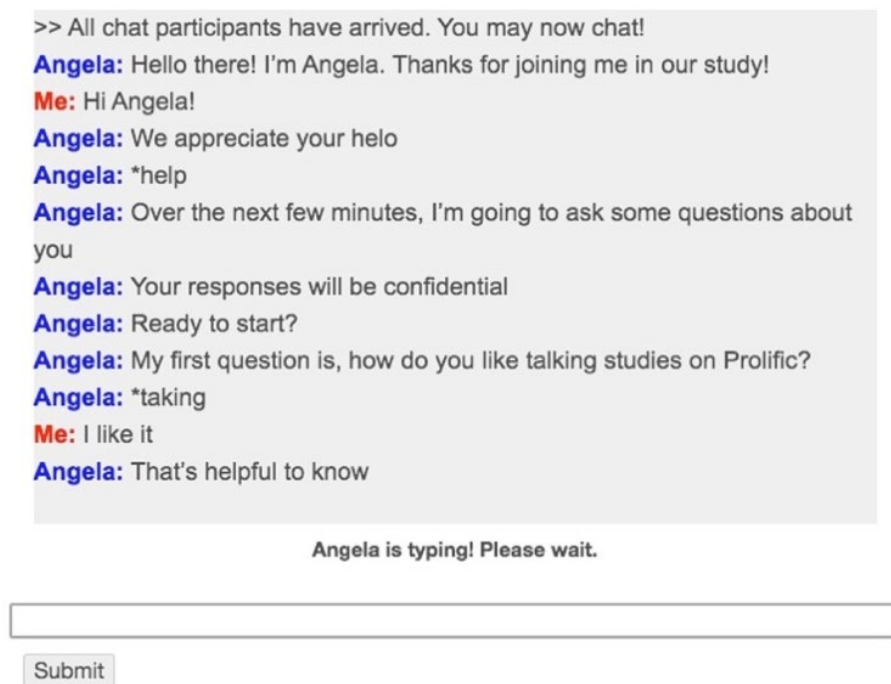


Figure 1. Example of the automated chatbot in Study 1A.

Measures. Following the chat, participants filled out a survey measuring their evaluation of the agent and the conversation. Participants reported their *perception of the agent's humanness* on a four-item composite measure that was developed based on prior literature and administered across all studies. The first two items focused on the uncertainty commonly experienced by online users (Adams 2018), thus providing face validity to the humanness perception construct: “To what extent did the customer service agent seem human on the previous screen?” (1 – *not at all human*; 7 – *extremely human*); “In your opinion, how likely is it that the customer service agent is a bot versus a human?” (0 – *very likely a bot*; 100 – *very likely a human*; we linearly transformed the scale ratings from 1 to 7 for our subsequent analyses). The next two items were developed to capture people’s lay belief that AI systems are unable to experience emotions or bodily sensations: “If you told a joke, do you think the agent would get

it?” and “Do you think this customer service agent ever gets tired when working?” (1 – *not at all likely*; 7 – *extremely likely*). These four items were highly correlated ($\alpha = .80$) and were combined to form our primary measure of *perceived humanness*. For all statistical analyses in this and subsequent studies, results hold when examining the single, most direct humanness item alone (i.e., “To what extent did the customer service agent seem human on the previous screen?”).

Participants reported their *perceptions of the agent’s helpfulness* (“How likely do you think it is that this agent could help you solve your problem?” and “How likely do you think it is that this agent would be able to understand you?”; $r = .87$; answered on Likert scales from 1 – *not at all* to 7 – *very much*).²

Finally, participants elaborated on their humanness judgment in an open-ended question, indicated what typo the agent made (if any) in a manipulation check question, and answered demographic questions on their age, gender, highest education level, and race/ethnicity.

Results

Manipulation check. Among all participants, 94.3% (169 out of 191 in the typo condition, 194 out of 194 in the no-typo condition) identified their condition correctly. Results remained the same when we analyzed the data including or excluding participants who failed the manipulation checks. Hence, below we report results including all participants.

Perceived humanness. As predicted, participants engaging in a live chat with a service agent perceived the agent who made and corrected their typos to be significantly more human ($M = 3.99$, $SD = 1.72$) than one who made no typos ($M = 3.27$, $SD = 1.34$; $F(1, 383) = 20.96$, $p < .001$, $\eta^2 = .052$).

Perceived helpfulness. Participants in both conditions believed that the agent would be somewhat able to understand and help them solve their issues, with directionally but non-significantly greater perceptions of helpfulness in the typo condition ($M = 4.91$, $SD = 1.62$) than in the no-typo condition ($M = 4.70$, $SD = 1.46$; $F(1, 383) = 1.86$, $p = .17$, $\eta^2 = .005$). While there was no direct effect of condition on helpfulness, there was an indirect effect through perceived humanness: Perceived humanness mediated the effect of typo on the agent's perceived helpfulness ($b = .51$, $SE = .11$, $95\% CI = [.28, .74]$).

Discussion

Study 1A employed an interactive paradigm wherein participants engaged in a live interaction with an ambiguous communicator powered by our pre-programmed chatbot. We observed that making and correcting typos led the chat agent to be perceived as more human. Contrary to our predictions, we did not observe an effect of condition on perceptions of the agent's helpfulness. We suspect that this might have at least partly resulted from the fact that the reflective conversation questions were subjective and open-ended in nature, so participants may not readily see the chatbot as "helpful" given they were asking questions and not providing solutions. In future studies, we examine chatbots in more typical service contexts, where helpfulness is an expected outcome of the conversation.

SUMMARY OF STUDIES 1B AND 1C

Study 1B examined the humanizing effect of typos when customers actually interacted with another human service agent. (A full description of the procedure and results can be found in the Web Appendix.) A team of research assistants interacted with 200 participants ($M_{age} = 33.60$, 45% female) via a private online chatroom. Research assistants in the typo condition

made and corrected two typos during the conversation (as in Study 1A), whereas those in the control condition made no typos. As predicted, a chat agent who made and corrected typos was perceived to be more human ($M = 4.31$, $SD = 1.51$) than an agent who did not make any typos ($M = 3.58$, $SD = 1.58$; $F(1, 188) = 10.48$, $p = .001$, $\eta^2 = .053$).

In Study 1C, we sought to test how corrected typos impacted perceived humanness of a chat agent across multiple customer service contexts. A total of 301 participants ($M_{\text{age}} = 34.19$, 47.1% female) were randomly assigned to imagine that they were experiencing a billing, shipping, or product issue with their cellphone company and that they logged into the company's website and started to chat with a customer service agent. Participants then saw a screenshot displaying a customer service agent's greeting message where the agent either misspelled "else" as "esle" (and corrected it in a subsequent message) or did not make a typo. As predicted, when the chat agent corrected a typo, participants perceived the agent to be more human ($M = 5.15$, $SD = 1.78$) compared to when the agent did not make a typo ($M = 2.73$, $SD = 1.56$; $F(1, 294) = 153.25$, $p < .001$, $\eta^2 = .345$). Furthermore, participants in the corrected-typo condition expected the agent to be more helpful ($M = 5.12$, $SD = 1.25$) than those in the no-typo condition ($M = 4.02$, $SD = 1.31$; $F(1, 295) = 55.06$, $p < .001$, $\eta^2 = .157$).³ As in Study 1A, perceived humanness of the agent mediated the difference in perceived helpfulness between the corrected-typo condition and the no-typo condition ($b = -1.02$, $SE = .12$; 95% CI = [-1.27, -.78]).

To examine whether corrected typos humanize communicators more than other potentially humanizing features, we conducted Supplemental Study 1 ($N = 815$; see Web Appendix for details). In this study, we compared the perceived humanness from a corrected typo with three other cues often associated with humans in text-based communication: the agent's name, gender, and a human photo. Correcting a typo had the strongest effect on

perceived humanness ($F(1, 798) = 227.33, p < .001, \eta^2 = .22$) compared to the other three tested cues.

STUDY 2: CORRECTED VERSUS UNCORRECTED TYPO

To investigate whether it is correcting a typo, rather than just making a typo, that humanizes the chat agent, Study 2 included three experimental conditions: one in which the chat agent made and corrected typos, another in which the agent just made the typos without correcting them, and a third in which the agent made no typo at all.

Method

Participants and design. This experiment employed a between-subjects design with three conditions: corrected-typo, uncorrected-typo, and no-typo. A total of 603 participants completed the study for \$2.00 each. According to our pre-registered exclusion criteria, we excluded participants whose responses to the questions were unintelligible or lacked minimal effort ($N = 28$), whose chat responses exposed the mechanics of the chatbot ($N = 51$)⁴, and who experienced technical errors during the chat ($N = 6$), resulting in a final sample of 518 participants for final analysis ($M_{\text{age}} = 36.94$; 42.5% female).

Procedure. Prior to the study, participants learned that we were “constructing a new participant database to study consumer purchases” and were randomly assigned to the no-typo condition, the corrected-typo condition, or the uncorrected-typo condition prior to the conversation. After entering the virtual chatroom, participants were greeted by a chat agent. In both the corrected-typo and the uncorrected-typo conditions, the agent made two typos during the conversation (i.e., “*helo*” / “**help*”, “*talking*”/ “**taking*”). However, in the uncorrected-typo condition, the agent continued the conversation without addressing the typos. During the

conversation, the agent asked a total of ten personal questions (e.g., “What is your date of birth?”, “What is your zip code?”), and participants were allowed to skip questions by simply pressing a skip button on their interface.

Measures. Participants reported their *perceptions of the agent’s humanness* ($\alpha = .92$) using the same items as in previous studies. Participants also indicated what typo the agent made, if any, in a manipulation check question.

Results

Manipulation check. Among all participants, 17.5% (89 participants) failed the manipulation check (66 out of 173 in the uncorrected-typo condition, 23 out of 156 in the corrected-typo condition, and 1 out of 180 in the no-typo condition). Results remained the same when we analyzed the data including or excluding participants who failed the manipulation checks. Therefore, below we report results including all participants.

Perceived humanness. A one-way ANOVA revealed a significant effect of the typo manipulation on perceived humanness of the chat agent ($F(2, 515) = 6.20, p < .001, \eta^2 = .024$). Specifically, when the chat agent corrected the typos, participants perceived the agent to be significantly more human ($M = 4.07, SD = 1.76$) compared to when the agent did not make a typo ($M = 3.45, SD = 1.67; t(340) = 3.32, p = .001, d = 0.36$) or when the agent made but did not correct the typos ($M = 3.50, SD = 1.85; t(331) = 2.83, p = .005, d = 0.31$). In fact, when the agent did not make any typos or did not correct the typos it did make, participants perceived similar levels of humanness in the agent ($t(359) = 0.30, p = .764, d = 0.02$).

Discussion

Using a live chat paradigm, Study 2 revealed that it is correcting a typo, rather than merely making a typo, that leads to greater perceptions of humanness. In other words, the

presence of errors *per se* does not necessarily humanize a communication, but the act of *correcting* errors seems to convey a mindful agent. We replicated this pattern in Supplemental Study 2 (N = 391) using a different set of typos (“*meat*” → “*meet*”; “*plajiarisim*” → “*plagiarism*”) with a button-based chatbot interface (see Web Appendix for details).

STUDY 3: AMBIGUOUS VERSUS KNOWN CHAT AGENT IDENTITY

Across earlier studies, the chat agent’s identity was kept ambiguous and left up to participants’ inferences. However, in many customer service contexts, consumers *know* that they are talking to a chatbot or to a human (e.g., because the company discloses this information). Thus, an important remaining question is whether corrected typos also increase subjective perceptions of humanness even when the identity of the agent is unambiguous. In Study 3, we tested this by directly manipulating the identity of the agent (ambiguous, chatbot, or human) and examining whether the humanizing effect of corrected typos would be moderated by knowledge of the agent’s identity. We predicted that corrected typos would have a stronger humanizing effect when the agent’s identity was ambiguous because people would be more sensitive to cues that can potentially reveal an agent’s identity when that identity isn’t already known. However, in light of people’s tendency to anthropomorphize objects that merely resemble human appearances or actions (e.g., Aggarwal and McGill 2007; Zhao et al. 2019), we also predicted that correcting typos would still humanize a chat agent that is known to be a chatbot. Finally, given the dehumanizing nature of written communication for human conversations, we were curious to examine whether correcting typos could even further humanize an agent known to be human.

Method

Participants and design. This experiment employed a 2 (error: typo, no-typo) × 3 (agent identity: ambiguous, chatbot, human) between-subjects design. A total of 386 participants were recruited via Prolific and completed the study for \$0.50 each ($M_{\text{age}} = 34.47$; 54.9% female).

Chat interface. We created a dynamic online platform, showing a chat agent sending one message at a time, and asked participants to evaluate the agent. We employed this paradigm—instead of a live chat—to keep the conversation identical and to manipulate only participants’ knowledge of the agent’s identity across conditions while holding all else constant.

Procedure. Participants learned that they would view an online chat conducted on a cellular company’s customer service platform. Depending on their condition assignment, participants were either told that this company was training “their new AI-powered chatbot that they call Angela” (chatbot condition) or “their new employee, named Angela” (human condition) to “use an interactive chat platform to answer customer service questions,” or that the cellular company was “developing an interactive chat platform to answer customer service questions” (ambiguous condition). Participants in all conditions were then asked to watch and evaluate a dynamic online chat conducted on this platform. On the next page, participants saw the chat agent send greeting messages line-by-line. The agent in the typo condition misspelled “else” as “esle” and corrected it in a subsequent message, whereas the agent in the no-typo condition spelled everything correctly.

Measures. After seeing the chat page, participants responded to three *perceived humanness* questions identical to those in previous studies ($\alpha = .79$; we removed “How likely is it that the customer service agent is a bot versus a human?” because the agent’s identity was known in some conditions). Participants also reported their *perceptions of the agent’s helpfulness* ($r = .85$). We counterbalanced whether perceived humanness was measured before or after

perceived helpfulness to ensure that the main effects and mediation results were robust to measurement order. Finally, participants completed manipulation checks regarding the agent's identity ("a human employee," "an AI-powered chatbot," or "it wasn't mentioned") and what typo the agent made (if any), before reporting their gender and age.

Results

Manipulation checks. For the typo manipulation check, 95.8% of the participants correctly identified their condition. For the identity manipulation check, 127 out of 127 (100%) participants in the chatbot condition and 104 out of 129 in the human condition (80.6%) correctly reported the agent's identity; by contrast, only 86 out of 130 in the ambiguous condition (66.2%) correctly indicated that the agent's identity was not mentioned, whereas others believed it was an AI-powered chatbot (18.5%; 24 out of 130) or a human employee (15.4%; 20 out of 130). Including or excluding those who failed the identity manipulation check led to largely similar results; below, we report results including all participants (results excluding participants who failed the agent identity manipulation are reported in the Web Appendix).

Perceived humanness. As predicted, a two-way ANOVA on perceived humanness revealed a significant main effect of the corrected-typo manipulation ($F(1, 380) = 102.11, p < .001, \eta^2_p = .21$), such that participants perceived the agent to be more human when it made a corrected typo ($M = 4.49, SD = 1.56$) than when it did not ($M = 3.15, SD = 1.43$). It also revealed a significant main effect of agent identity ($F(2, 380) = 53.06, p < .001, \eta^2_p = .22$) such that participants perceived a known human agent to be more humanlike ($M = 4.57, SD = 1.57$) than an ambiguous agent ($M = 3.97, SD = 1.63; t(257) = 3.01, p = .001, d = .37$), which was more human than a known chatbot agent ($M = 2.90, SD = 1.24; t(255) = 5.86, p < .001, d = .73$). Finally, there was a significant interaction effect ($F(2, 380) = 6.01, p = .003, \eta^2_p = .031$; see

Figure 2). Examining the effect of the typo manipulation on each agent identity condition separately, we found that the humanizing effect of a corrected typo was larger when the agent’s identity was ambiguous (corrected-typo vs. no typo: $M_s = 4.97$ vs. 3.14 , $SD_s = 1.34, 1.38$; $t(128) = 7.62, p < .001, d = 1.34$) or was known to be a human ($M_s = 5.30$ vs. 3.78 , $SD_s = 1.31, 1.45$; $t(127) = 6.25, p < .001, d = 1.10$) than when it was known to be a chatbot ($M_s = 3.25$ vs. 2.53 , $SD_s = 1.18, 1.22$; $t(125) = 3.40, p < .001, d = 0.60$).

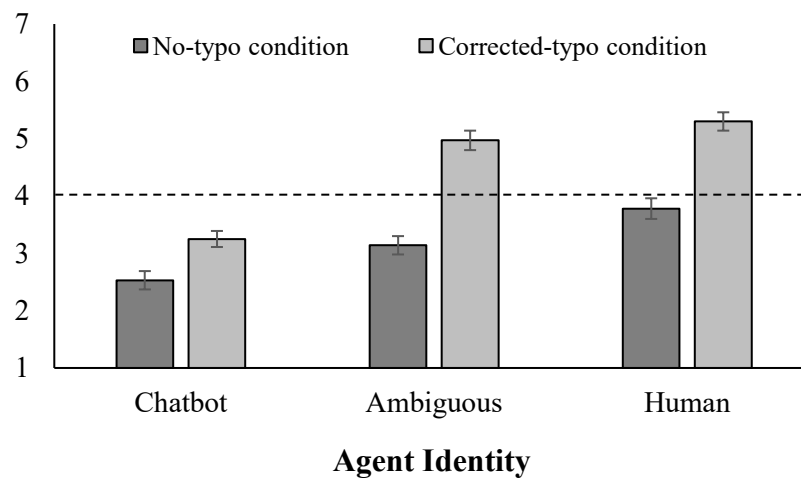


Figure 2. Results on perceived humanness in Study 3. The dotted horizontal line represents the neutral rating of 4. Error bars represent ± 1 standard error around the means.

Perceived helpfulness. A two-way ANOVA revealed a main effect of the typo manipulation ($F(1, 380) = 11.83, p < .001, \eta^2_p = .030$), such that participants perceived the agent to be more helpful when it made a corrected typo ($M = 4.59, SD = 1.49$) than when it did not ($M = 4.10, SD = 1.60$).⁵ It also revealed a main effect of agent identity ($F(2, 380) = 25.15, p < .001, \eta^2_p = .12$), such that participants believed that a human employee ($M = 4.85, SD = 1.43$) would be marginally more helpful than an ambiguous agent ($M = 4.54, SD = 1.43$; $t(257) = 1.75, p$

= .08, $d = .22$), and an ambiguous agent would be more helpful ($M = 4.54$, $SD = 1.43$) than a chatbot ($M = 3.61$, $SD = 1.58$; $t(255) = 4.95$, $p < .001$, $d = .62$). There was no interaction effect ($F(2, 380) = .528$, $p = .59$, $\eta^2_p = .003$). Finally, perceived humanness of the agent mediated the difference between the no-typo condition and the corrected-typo condition ($b = .84$, $SE = .11$; 95% CI = [.63, 1.06]).

Discussion

Consistent with our prediction, knowledge about a chat agent's identity moderated the humanizing effect of typos. In particular, the humanizing effect of the typo was larger when the agent's identity was ambiguous than when the agent was known to be a chatbot. But importantly, correcting a typo still humanized the communicator even when people knew that the chat agent was a chatbot or a human. Furthermore, seeing an agent correct a typo again led people to expect the agent to be more helpful than seeing an agent that made no typo.

GENERAL DISCUSSION

Online text-based communication is an essential means for companies to provide efficient and scalable customer service experience. As customer service agents in such conversations are increasingly pre-programmed AI chatbots (rather than humans), it is important to understand how consumers perceive the humanness of chat agents and the consequences of these perceptions for consumer judgments and decisions. Across five experiments (reported in the main text) and two supplemental experiments (reported in the Web Appendix) using a variety of stimuli and paradigms—from static or dynamic message displays, to live chats with a programmed chatbot, to live chats with a human agent—our research consistently demonstrates that making and correcting one's typos in a conversation can humanize a service agent (see

summary results in Figure 3). Furthermore, correcting typos can elicit favorable perceptions from customers, such as making those agents appear more helpful. Crucially, we found that it is the correcting of one’s typo, rather than merely making a typo, that humanizes an agent.

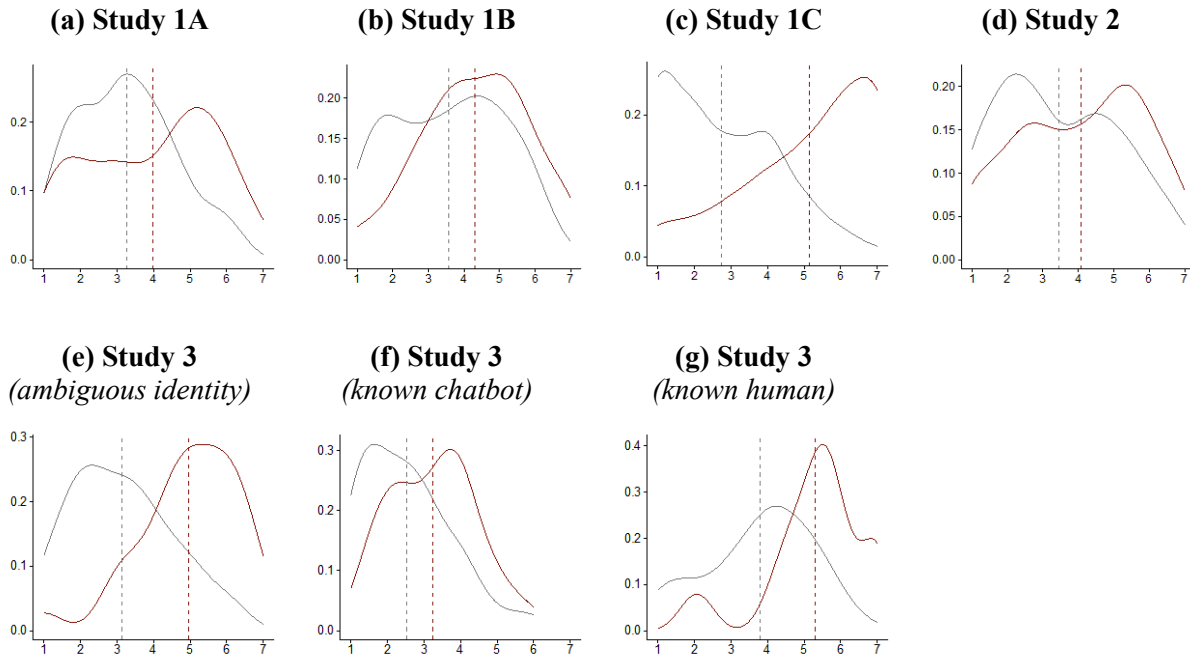


Figure 3. Distribution of perceived humanness in the no-typo (in grey) versus corrected-typo (in red) conditions across all studies.

Theoretical Contributions

By emphasizing the dynamic process of how text is generated, our research fills an important gap in the literature on mind perception. Different from the majority of prior work, which focuses on the outcome of a communication (e.g., typefaces, emojis; Candello et al. 2017; Jakesch et al. 2023; Luangrath et al. 2017; Schroll et al. 2018) or on verbal and nonverbal cues of the communicator (e.g., speed of movement, voice; Morewedge et al. 2007; Schroeder and Epley

2016), our research highlights how the very way in which text is written and appears on a screen can provide a unique window into the “mind” of an online agent.

Moreover, our research expands the empirical literature on the interpersonal consequences of making mistakes. Specifically, we help to reconcile an intriguing discrepancy between research suggesting that small errors can increase the perceived likeability of people and products (e.g., Aronson et al. 1966) and a separate literature showing that errors can lead to negative impressions of an actor (e.g., Figueredo and Varnhagen 2005). We show that acknowledging and correcting one’s error, as opposed to simply making an error, is a key factor for increasing perceived humanness. We theorize that this is because addressing one’s error signals an engaged and caring mind—it requires some degree of intentionality (e.g., desire to avoid being seen as careless) and meta-cognitive capacities (e.g., self-awareness), which are generally considered to be lacking in machines, algorithms, and other artificial intelligence.

Practical Implications

Knowledge about what humanizes a communicator can be a double-edged sword. Although it might help service agents and consumers convey their humanness in text-based communication, this knowledge might also tempt companies to trick users into believing they are interacting with a human when it is really a bot. Concerns regarding harm to consumers illustrate why policymakers should provide more guidance to, and constraints on, companies seeking to deploy increasingly humanlike chatbots. Indeed, recent policy efforts in the United States include some states requiring bots to disclose their identities when interacting with consumers or voters and proposed solutions like watermarking AI-generated content.

As attempts to deceive consumers within text-based conversations will likely result in negative consequences for the agent and the company, we believe companies should focus their

efforts on authentic and transparent humanization of customer service agents to maximize potential benefits (e.g., appearing helpful and warm) and minimize potential backlash (e.g., appearing manipulative and untrustworthy). As we found in Study 3, even when explicitly telling people that an agent is a human or chatbot, corrected typos *still* had an additional humanizing impact, suggesting that companies can be honest while simultaneously enjoying the benefits of enhanced humanization of their agents. Regardless of whether the company employs a human representative or a chatbot, finding ways to humanize the company's service agents can signal the company's dedication to connecting with customers, potentially offsetting the impersonal and dehumanizing nature of text-based interactions. Relatedly, our research provides the reassuring news that, contrary to the intuitions of many people, making a few typos when interacting with consumers is unlikely to create permanent damage in customer relationships (and may even have positive effects on a variety of other downstream consequences, such as perceptions of an agent's warmth, willingness to reward an agent or take their recommendation, and interest in interacting with an agent again in the future).

For consumers, our research highlights a cognitive heuristic affecting the way they perceive communicators. Shedding light on how people respond to their conversational counterparts in online settings can help them understand how dynamic cues in text communications affect their consumption experiences. This is especially important since, based on our pilot study, consumers are not always aware of how these forces impact their preferences. In the new age of digital communication, where consumers sometimes struggle to discern the identity of their online interaction partners, providing them with knowledge and tools to distinguish humans from bots may help them gain a greater sense of control over their digital environment.

Limitations and Future Directions

Our results elicit several questions regarding generalizability. First, the current investigation used typos that involved adjacent keys on the keyboard (e.g., mistakenly typing “p” as “o” in “help”), which may be uniquely humanizing because they imply “motoric failures” caused by physical embodiment that only humans have. It would be interesting to examine whether the humanizing effect of corrected typos would be stronger in situations where the motoric failure is more obvious, such as when it involves a categorical change (e.g., typing “p” as “[”], or weaker for typos that seem unrelated to motoric failures (e.g., typing “p” as “h”).

Second, typos are relatively harmless errors that everyone makes and can easily relate to, which might be why they were *not* perceived as diagnostic for the agent’s general competence (see Web Appendix for these results). It is unclear, however, whether the type of errors (e.g., grammatical errors, incorrect information), the frequency of errors (i.e., number of errors the communicator makes), or the customer’s *goal* within the interaction (e.g., seeking therapy vs. seeking scientific facts) might moderate how the errors affect consumer perceptions. Moreover, relationships among the perception of humanness, helpfulness, and general competence may be nuanced and context-dependent— for instance, in domains where machines are known to outperform humans (e.g., information retrieval), consumers might perceive a more machine-like agent to be more helpful.

Third, even though we limited our current investigation to correcting typos as a humanizing cue, our theoretical argument suggests that other changes in text that unfold over time can signify a thoughtful, engaged, and intentional mind. From pressing backspace to delete one’s writing, to changing one’s ideas as conversation unfolds, future research could examine

how such traces of a fallible mind can be uniquely humanizing. Moreover, given that other characteristics of written communication (e.g., emojis and informal language) have also been shown to humanize a writer, future research can compare the humanizing power of the dynamic process of writing versus the final product of writing.

Fourth, it is worth investigating what other downstream consequences may result from observing typos or errors more broadly. Our Web Appendix reports the results of the corrected-typo manipulation on several other possible consequences, some of which showed relatively consistent effects (e.g., such as perceptions of an agent's warmth, willingness to reward an agent or take their recommendation, and interest in interacting with an agent again in the future), but others which were inconsistent (e.g., customers' willingness to disclose their personal information to agents). In light of these early and inconclusive results, future researchers might want to investigate how humanness perception can impact self-disclosure and other conversation behaviors, as well as the underlying psychological mechanisms. Such findings will have important implications for online privacy in numerous contexts where consumers interact with chatbots, from online counseling to conversations on social media.

REFERENCES

- Adams, Tim (2018), "The charge of the chatbots: How do you tell who's human online?" *The Guardian* (November 18).
- Aggarwal, Pankaj and Ann L. McGill (2007), "Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products," *Journal of Consumer Research*, 34 (4), 468-479.
- Aronson, Elliot, Ben Willerman, and Joanne Floyd (1966), "The effect of a pratfall on increasing interpersonal attractiveness," *Psychonomic Science*, 4, 227-228.
- Boden, Margaret (2018), "The robots won't take over because they couldn't care less," *Aeon Essays*, (August 13), <https://aeon.co/essays/the-robots-wont-take-over-because-they-couldnt-care-less>.
- Candello, Heloisa, Claudio Pinhanez, and Flavio Figueriredo (2017), "Typefaces and the perception of humanness in natural language chatbots," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, New York, NY: Association for Computing Machinery, 3476–3487.
- Christian, Brian (2012), *The Most Human Human: What artificial intelligence teaches us about being alive*, New York, NY: Anchor Books.
- Figueredo, Lauren and Connie K. Varnhagen (2005), "Didn't you run the spell checker? Effects of type of spelling error and use of a spell checker on perceptions of the author," *Reading Psychology*, 26, 441-458.
- Fuchs, Christoph, Martin Schreier, and Stijn MJ Van Osselaer. "The handmade effect: What's love got to do with it?" *Journal of marketing* 79, no. 2 (2015): 98-110.
- Gray, Heather M., Kurt Gray and Daniel M. Wegner (2007), "Dimensions of mind perception," *Science*, 315, 619.
- Granulo, Armin, Christoph Fuchs, and Stefano Puntoni. "Preference for human (vs. robotic) labor is stronger in symbolic consumption contexts." *Journal of Consumer Psychology* 31, no. 1 (2021): 72-80.
- Helmreich, Robert, Eliot Aronson and Jim LeFan (1970), "To err is humanizing sometimes: Effects of self-esteem, competence, and a pratfall on interpersonal attraction" *Journal of Personality and Social Psychology*, 16, 259.

- Jakesch, Maurice, Jeffrey T. Hancock, and Mor Naaman (2023). "Human heuristics for AI-generated language are flawed," *Proceedings of the National Academy of Sciences*, 120(11).
- Kiesler, Sara, Jane Siegel and Timothy W. McGuire (1984), "Social psychological aspects of computer-mediated communication," *American Psychologist*, 39, 1123.
- Kreiner, David S., Summer D. Schnakenberg, Angela G. Green, Michael J. Costello and Anis F. McClin (2002), "Effects of spelling errors on the perception of writers," *The Journal of General Psychology*, 129, 5-17.
- Lea, Martin and Russell Spears (1992), "Paralanguage and social perception in computer-mediated communication," *Journal of Organizational Computing and Electronic Commerce*, 2, 321-341.
- LivePerson (2019), "How consumers view conversational commerce in 2019," <https://www.liveperson.com/resources-reports/consumer-preferences-2019>.
- Looser, Christine E. and Thalia P. Wheatley (2010), "The tipping point of animacy: How, when, and where we perceive life in a face," *Psychological Science*, 21, 1854-1862.
- Lortie, Catherine L. and Matthieu J. Guitton (2011), "Judgment of the humanness of an interlocutor is in the eye of the beholder," *PLOS One*, 6 (9), e0025085.
- Luangrath, Andrea W., Joann Peck and Victor A. Barger (2017), "Textual paralanguage and its implications for marketing communications," *Journal of Consumer Psychology*, 27, 98-107.
- McCoy, John P. and Tomer D. Ullman (2018), "A minimal Turing Test," *Journal of Experimental Social Psychology*, 79, 1-8.
- Mirrig, Nicole, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi (2017), "To err is robot: How humans assess and act toward an erroneous social robot," *Frontiers in Robotics and AI*, 4, 21.
- Morewedge, Carey K., Jesse Preston and Daniel M. Wegner (2007), "Timescale bias in the attribution of mind," *Journal of Personality and Social Psychology*, 93, 1-11.
- Norman, Donald A. (2004), "Everyday Things". Basic books.
- Pinker, Steven and Paul Bloom (1990), "Natural language and natural selection," *Behavioral & Brain Sciences*, 13, 707-784.
- Reddy, Trips. (2017), "How chatbots can help reduce customer service costs by 30%," *IBM*.

- Reich, Taly. Daniella Kupor and Rosanna Smith (2017) ,"Made By Mistake: When Mistakes Increase Product Preference " , in NA - Advances in Consumer Research Volume 45, eds. Ayelet Gneezy, Vladas Griskevicius, and Patti Williams, Duluth, MN : *Association for Consumer Research*, Pages: 834-838.
- Schroeder, Juliana and Nicholas Epley (2015), "The sound of intellect: Speech reveals a thoughtful mind, increasing a job candidate's appeal," *Psychological Science*, 26, 877-891.
- Schroeder, Juliana and Nicholas Epley (2016), "Mistaking minds and machines: How speech affects dehumanization and anthropomorphism," *Journal of Experimental Psychology: General*, 145, 1427-1437.
- Schroeder, Juliana, Michael Kardas and Nicholas Epley (2017), "The humanizing voice: Speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement," *Psychological Science*, 28, 1745-1762.
- Schroll, Roland, Benedikt Schnurr and Dhruv Grewal (2018), "Humanizing products with handwritten typefaces," *Journal of Consumer Research*, 45(3), 648-672.
- Turing, Alan Mathison (1950), "Mind." *Mind*, 59(236), 433-460.
- Vignovic, Jane A. and Lori F. Thompson (2010), "Computer-mediated cross-cultural collaboration: Attributing communication errors to the person versus the situation," *Journal of Applied Psychology*, 95, 265.
- Westerman, David, Aaron C. Cross and Peter G. Lindmark (2018), "I believe in a thing called bot: Perceptions of the humanness of 'chatbots'," *Communication Studies*, 1-18.
- Zhao, Xuan, Elizabeth Phillips and Bertram F. Malle (2019), "How people infer a humanlike mind from a robot body," forthcoming, <https://doi.org/10.31234/osf.io/w6r24>

Web Appendix

Imperfectly Human:

The Humanizing Potential of (Corrected) Errors in Text-Based Communication

This PDF file includes:

Pilot Study: Lay Beliefs about Typos

Supplemental Methods and Results for Studies 1–3

Supplemental Study S1: Typo vs. Other Humanizing Cues

Supplemental Study S2: Erring Without Correction Using a Button-Based Chatbot

Supplemental References

Auxiliary materials for this manuscript are deposited on OSF and include:

Materials and measures for all studies

Preregistration forms

Datasets

OSF depository URL (anonymized for reviewer):

https://osf.io/6s8wm/?view_only=c2e1d7bb0efb455fa97deb11a844f6d9

Pilot Study: Lay Beliefs about Typos

To investigate both general and managerial intuitions about typos in customer service contexts, we asked participants to take the perspective of a customer or a company manager and measured their expectations about customers' responses to corrected typos in a service agent's text. We predicted that participants would fail to recognize the potential humanizing power of corrected typos, regardless of whether they take the perspective of a customer or a manager.

Method

Participants and design. We recruited a total of 413 adults ($M_{age} = 37.64$, $SD_{age} = 13.04$; 47.3% female) from Prolific who completed our survey in exchange for \$0.28. We aimed to collect 50 participants in each of the eight experimental conditions, thus 400 participants across conditions. Ten participants were excluded from the analysis—one declined the consent form and thus never started the study, and nine started but did not complete the study—leaving a sample of 403 participants for data analysis.

This study employed a 2 (error: no-typo, typo) \times 2 (perspective: customer, manager) \times 2 (service context: billing issue, service issue) between-subjects design. Participants were randomly assigned to one of the eight conditions.

Procedure. Our primary manipulation of interest was whether participants were told that the customer service agent made no typos (no-typo condition) or made and corrected several typos (corrected-typo condition) when interacting with customers. We added two additional manipulations for enhanced generalizability. First, we manipulated participants' perspective by asking them to imagine either that they wanted to chat with a customer service agent over text to ask for assistance (customer perspective) or that they oversaw customer service agents who

assisted customers on such issues (manager perspective). Second, we manipulated whether the customer had a billing issue with their phone plan or a service issue with their Internet plan.

Measures. After reading their respective scenarios, participants reported their prediction of their own intention (customer perspective) or their prediction of their customers' intention (manager perspective) *to interact with the agent* on a composite measure with two items (e.g., “How much [do you/will customers] want to work with this customer service agent?” and “Upon being assigned to this customer service agent, how likely are [you/customers] to log out and try to get a different agent?” (reverse-scored); $r = .71$). Ratings on both items ranged from 1 (*not at all*) to 7 (*very much/likely*).

To measure general preferences for seeing typos, the last page of the survey asked participants to indicate whether they could think of a situation in which they would prefer a customer service agent who made typos in the text (yes/no) and to describe that situation (if they chose “yes”) before reporting their age and gender.

Results

A three-way ANOVA on participants' prediction of customers' intention to interact with the agent revealed the expected significant main effect of typo, $F(1, 393) = 233.89, p < .001, \eta^2 = .373$, such that participants reported (or predicted) lower intention to interact with the agent when the agent made a typo ($M = 3.80, SD = 1.53$) than when it did not ($M = 5.76, SD = 1.03$). Not only did participants see little value in using typos in customer service interactions, but they also believed that typos would actually *harm* customer engagement. The main effect of perspective was also significant, $F(1, 393) = 7.84, p = .005, \eta^2 = .020$, such that participants who took the perspective of a customer predicted greater intention to work with the agent ($M = 4.98,$

SD = 1.51) compared to those who took the perspective of a manager (M = 4.60, SD = 1.72). The effect of service context was not significant ($F(1, 393) = .32, p = .569, \eta^2 = .001$).

Among all interaction terms, only the interaction between typo and perspective was significant ($F(1, 393) = 9.07, p = .003, \eta^2 = .023$). Specifically, those who took the perspective of a manager expected customers to be even less willing to work with an agent with typos than an agent without typos (typo: M = 3.43, SD = 1.40; no-typo: M = 5.78, SD = 1.09), $t(198) = -13.13, p < .001, d = 1.87$, compared to those who took the perspective of a consumer (typo: M = 4.18, SD = 1.57; no-typo: M = 5.75, SD = .96), $t(199) = -8.56, p < .001, d = 1.20$.

Finally, we examined the proportions of participants who indicated that they could (versus could not) think of a situation where they would prefer a customer service agent who made typos in their text. Overall, only 18% of the participants reported that they could think of a situation where typos might be preferable, which was significantly lower than the chance level of 50%, $\chi^2 = p < .001$.

Discussion

The majority of participants in this study believed that agents who make typos will harm customer service interactions. Whether taking the perspective of a consumer or a manager in a customer service interaction, participants reported or predicted that they/customers would be less interested in engaging with customer service agents who made and corrected typos compared to those who did not make typos. Moreover, most participants struggled to come up with *any* situation in which it would be preferable for customer service agents to make typos. These results suggest that people do not have the intuition that typos could be desirable in online customer service—if anything, people’s general belief is that typos should be avoided.

Supplemental Method and Results for Study 1A

Measures not reported in main text

Perceived warmth and competence. Participants reported their *perceptions of the agent's warmth* (friendly/good-natured/warm/sincere; $\alpha = .90$) and *competence* (confident/competent/independent/intelligent; $\alpha = .90$; scale adapted from Fiske et al. 2007) on scales from 1 (*not at all*) to 7 (*very much*).

Willingness to reward the agent. We tested a potential behavioral consequence—participants' *willingness to reward the agent* (“To what extent do you think that this agent should be considered to receive a reward (full week of Starbucks coffee supply) from the company?”) —on a scale from 1 (*not at all*) to 7 (*very much*).

Information disclosure. Based on findings that people self-disclose more information to human-like virtual therapists (Kang and Gratch 2014) and feel a stronger sense of mutual awareness when answering personal questions from a more humanlike communicator (Appel et al. 2012), we decided to additionally explore whether people are more willing to disclose personal information when a chat agent makes and corrects typos than when it does not make typos, and whether this effect is mediated by perceptions of humanness. We measured participants' willingness to disclose personal information as a potential consequence. Specifically, we created an *information-disclosure index* that included six items (full name, home address, date of birth, email address, phone number, and social media profile picture; $\alpha = .86$), which represented the type of personal information most likely to be known by a friend or a colleague as they start to develop a personal or professional relationship.

Other potential consequences: Satisfaction, interpersonal closeness, and company impressions. Participants reported their *satisfaction* with the conversation and their *interpersonal*

closeness to the agent. To explore other potential consequences of humanizing a chat agent, we also measured participants' *impressions of the company* on three items ($\alpha = .90$): "This research team cares about their participants"; "This research team has invested a lot of effort toward engaging with their participants"; "I hold a positive view of this research team" (1 – *strongly disagree*; 7 – *strongly agree*). Prior research shows that consumers believe that the primary reason for companies to deploy many new technologies is to reduce costs (Larivière et al. 2017; LivePerson 2017). Given people's negative perceptions about the motives behind companies' usage of non-human agents, we expected that consumers would form more favorable impressions of businesses when they perceived their service agents to be more humanlike.

Results on measures not reported in main text

Perceived warmth and competence. Participants' perceptions of the agent's warmth did not differ between conditions (typo: $M = 5.39$, $SD = 1.27$; no typo: $M = 5.28$, $SD = 1.23$), $F(1, 383) = 0.73$, $p = .39$, $\eta^2 < .001$, nor did perceptions of the agent's competence (typo: $M = 5.24$, $SD = 1.35$; no typo: $M = 5.12$, $SD = 1.20$), $F(1, 383) = 0.92$, $p = .34$, $\eta^2 < .001$. However, perceived humanness mediated the relationship between the typo manipulation and perceptions of agent warmth ($b = 0.34$, $SE = .08$, 95% $CI = [.19, .51]$) and agent competence ($b = 0.34$, $SE = .08$, 95% $CI = [.19, .50]$).

Willingness to reward the agent. Participants who saw the agent make and correct typos were significantly more willing to reward the agent ($M = 5.34$, $SD = 1.81$) than those who saw an agent that made no typos ($M = 4.76$, $SD = 1.84$; $F(1, 383) = 9.63$, $p = .002$, $\eta^2 = .035$). Perceived humanness mediated the effect of typo on reward endorsement ($b = .51$, $SE = .12$, 95% $CI = [.28, .75]$).

Information disclosure. Measuring participants' information-disclosure intention with the 6-item information-disclosure index, we found that participants did not report significantly different intentions between the two conditions (typo: $M = 1.63$, $SD = 1.65$; no-typo: $M = 1.82$, $SD = 1.86$), $F(1, 383) = 1.20$, $p = .27$, $\eta^2 = .003$. On average, most participants were only willing to disclose one or two personal information items. However, perceived humanness of the chat agent mediated the effect of typo on information-disclosure index, $b = .24$, $SE = .07$, 95% CI = [.12, .39].

Other potential consequences. Participants in the typo condition did not report significantly greater satisfaction ($M = 5.48$, $SD = 1.19$) than those in the no-typo condition ($M = 5.35$, $SD = 1.15$), $F(1,383) = 5.42$, $p = .021$, $\eta^2 = .028$; however, perceived humanness mediated the effect of typo on people's satisfaction, $B = 0.26$, $SE = 0.06$; 95% CI = [0.14, 0.39]. Participants in the typo condition also did not report feeling significantly closer to the chat agent than those in the no-typo condition ($M_s = 4.10$ and 3.90 , $SD_s = 1.60$ and 1.43 , respectively), $F(1,383) = 1.66$, $p = .20$; however, perceived humanness mediated the effect of typo on closeness to the agent, $B = 0.48$, $SE = 0.11$; 95% CI = [0.27, 0.71].

Finally, participants formed a marginally more favorable opinion of the company when the chat agent made a typo ($M = 5.53$, $SD = 1.20$) than when the agent did not ($M = 5.32$, $SD = 1.16$); $F(1, 383) = 3.08$, $p = .080$, $\eta^2 = .008$. Moreover, perceived humanness mediated the effect of typo on people's favorable impression of the company ($b = .26$, $SE = .06$, 95% CI = [.14, .38]).

Supplemental Method and Results for Study 1B

Method

Participants and design. We employed a between-subjects design with two experimental conditions—typo and no-typo—and aimed to recruit 100 participants in each condition.¹ In total, 200 participants were recruited from either a university campus or from a metropolitan area with a more diverse population.² Participants completed this experiment in exchange for \$4.00.

Following our predetermined exclusion criteria, we excluded a total of 10 participants because they experienced prolonged disconnection from the chat due to Internet disruption or experimenter error ($N = 4$), lacked sufficient English proficiency to have an online chat ($N = 2$), provided inattentive responses ($N = 1$), or because the human chat agent made unintended errors other than the prescribed typos ($N = 3$). This resulted in a final sample of 190 participants (campus-based sample: $N = 74$, $M_{\text{age}} = 24.5$, $SD_{\text{age}} = 7.7$, 52.7% female; downtown city-based sample: $N = 126$, $M_{\text{age}} = 42.7$, $SD_{\text{age}} = 14.5$, 37.3% female).

Human confederates. Six experienced research assistants served as human chat agents and received rigorous standardized training based on a predetermined script to ensure identical conversations across all participants (see OSF folder for the confederate training protocol and script). The script was pretested to eliminate potential ambiguity in the agent's language and to keep participants on the same course of conversation. We trained the human confederates to time their responses to convey the impression of reading and typing.

¹ According to a power analysis, this sample size was sufficient to detect a small-to-medium effect size of $d = .40$ given power of .80 and an alpha level of .05. We chose the minimum sample size we would need due to the sheer amount of resources required to coordinate a live interaction paradigm that involved recruiting participants to a laboratory to interact with live human confederates.

² Additional analyses confirmed that recruitment location did not interact with any of our key dependent variables.

Conversation interface. To enable real-time conversations with human chat agents, we employed the ChatPlat platform (<https://www.chatplat.com/>), a web application that resembles an actual online chatroom and allows multiple users to send and receive messages in a private chatroom in real time. This application has been used in previous research to study human social interactions (e.g., Huang et al 2017; Logg et al 2019). ChatPlat also displays a prompt indicating “someone else is typing” when it detects activity in the other user’s entry box.

Procedure. We informed participants that this study was part of “a nationwide project to compare research practices and participant populations across behavioral labs,” and that they would be interacting with a “research agent” who could be either a human or a chatbot (when in fact the agent was a human across conditions). Next, we randomly assigned participants to either the typo or the no-typo condition before matching them with an agent to begin the chat. Participants then saw a loading animation and the sentence, “You will be matched with a research agent in a moment.” In both conditions, they were first greeted by the chat agent, who introduced herself as “Angela”. In the typo condition, Angela made and corrected a typo (“helo /“*help””; see Figure S1), whereas in the no-typo condition, Angela did not make a typo. Next, Angela provided a brief task instruction and confirmed that the participant was ready to start before proceeding to the first question. To strengthen the manipulation, we then included a second typo and correction in Angela’s first question in the typo condition (“talking” /“*taking””; see Figure S1). No further typos were presented for the rest of the chat.

The agent asked a total of eight personal questions that were intended to elicit factual personal information from this sample of participants (e.g., “What is the most sensitive issue that you ever encountered at the [lab name]?”, “What is your primary source of income?”, and “What religion do you believe in, if any?”). Participants could skip any question that they did not wish

to respond. When a participant asked a clarifying question, Angela provided either a brief answer (if the question was a simple yes/no question) or a pre-scripted response like “*Thanks for asking, but I can’t comment on that.*” Finally, Angela concluded the chat by indicating that the time was up, thanked the participant, and instructed participants to proceed to the next page.

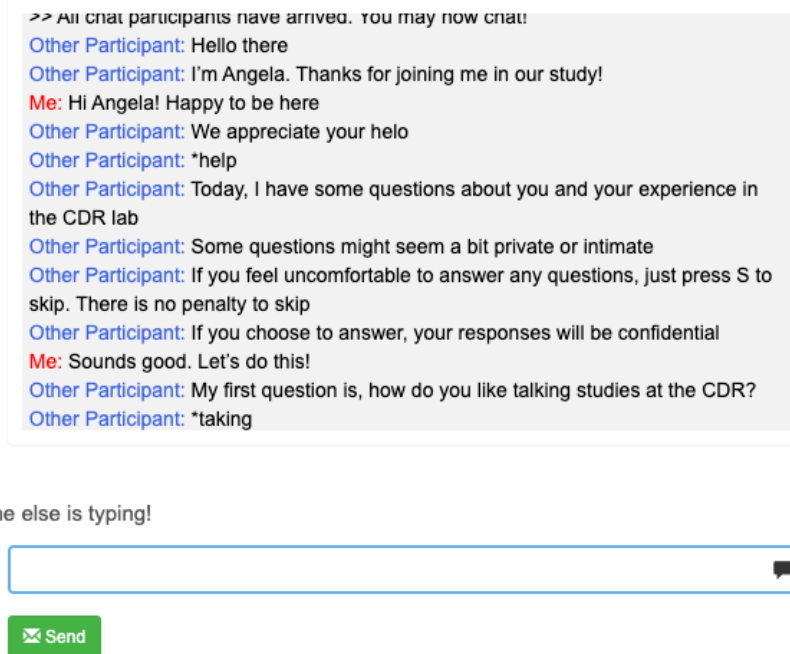


Figure S1. Example of the interactive chat platform in Study 1B (Chatplat).

Measures. First, participants reported their *perception of the agent’s humanness* on the same four-item composite measure described in Study 1A. Participants then reported their *perceptions of the agent’s warmth* and *competence* on the same measures described in the Study 1A Supplemental Materials.

Information disclosure. In this study, the chat agent asked questions that focused on eliciting factual, private information about the participants such as annual income and mortgage payment, and participants were allowed to skip a question if they did not wish to respond. This

setup provided an opportunity to explore participants' actual information disclosure behavior, and we analyzed the number of questions they chose to skip during the chat.

Given that the content of personal questions were limited by IRB regulations in a research lab, we then measured participants' willingness to share more personal and sensitive information in the self-reported survey afterwards, where participants indicated whether they would be willing to share private information with the same agent on an 18-item *information-disclosure index* (e.g., full name, home address, date of birth, email address, phone number, and social media profile picture; sharing medical records, tax documents, and Internet browsing history; $\alpha = .86$; see research materials in the OSF folder).

Other potential consequences: Satisfaction and interpersonal closeness. To explore other potential consequences of humanizing a chat agent, participants indicated their overall *satisfaction* of the conversation on two items ($\alpha = .90$): "How satisfied are you with your interaction with the agent?" (1 - *not at all*; 7 - *very much*); "How would you rate your interaction with the agent?" (1 - *very sad face*; 7 - *very happy face*). And they indicated their *interpersonal closeness* to the agent on three items ($\alpha = .91$): "To what extent did you feel connected to the agent?"; "To what extent did you feel close to the agent?"; "How much do you like the agent?" (1 - *not at all*; 7 - *very much*).

Control variable: Privacy concerns. We included part of the "Internet Users' Information Privacy Concerns" survey (Malhotra et al. 2004) to explore whether individual differences in self-reported privacy concern would moderate any effects on information disclosure.

Results

Manipulation check. Among all participants, 86.32% passed the manipulation check (92 out of 93 in the no-typo condition, 72 out of 97 in the corrected-typo condition). Results

remained the same when we analyzed the data including or excluding participants who failed the manipulation checks. As a more conservative test of our hypotheses, we report results including all participants regardless of whether they passed manipulation checks in this and subsequent studies.

Perceived humanness. As predicted, a chat agent who made and corrected typos was perceived to be more human ($M = 4.31$, $SD = 1.51$) than an agent who did not make any typos ($M = 3.58$, $SD = 1.58$; $F(1, 188) = 10.48$, $p = .001$, $\eta^2 = .053$).

Perceived warmth and competence. Participants in the typo condition perceived the chat agent to be warmer ($M = 5.07$, $SD = 1.59$) than those in the no-typo condition ($M = 4.52$, $SD = 1.73$; $F(1, 188) = 5.10$, $p = .025$, $\eta^2 = .027$). Participants in the typo and the no-typo conditions perceived the agent to be similarly competent ($M_s = 5.10$ and 4.87 , $SD_s = 1.50$ and 1.49 , respectively; $F(1, 188) = 1.10$, $p = .30$, $\eta^2 = .006$). Perceived humanness mediated the effect of typo on perceived warmth ($b = .41$, $SE = .14$; 95% $CI = [.16, .70]$).

Information disclosure. Because the number of questions participants answered during the chat was heavily skewed towards the maximum value, we conducted a non-parametric Mann-Whitney U-test as specified in our preregistration. Participants in the typo condition ($M = 7.57$, $SD = 0.99$) answered marginally more personal questions than those in the no-typo condition ($M = 7.30$, $SD = 1.28$), $W = 5004.5$, $p = .099$, and results were similar when we forced a parametric test (one-way ANOVA): $F(1,188) = 2.59$, $p = .109$, $\eta^2 = .014$. Furthermore, for information-disclosure intention measured by the 18-item index, participants in the typo condition reported that they would be willing to share more private information items in the future than those in the no-typo condition ($M_s = 6.60$ and 4.44 , $SD_s = 4.05$ and 3.63 , respectively), $F(1,188) = 14.9$, $p < .001$, $\eta^2 = .073$.

Mediation analyses showed that perceived humanness of the agent mediated the relationship between the typo manipulation and participants' information-disclosure behavior during the current chat, $b = .08$, $SE = .05$; 95% CI = [.01, .20], as well as their intention to disclose private information in a future chat, $b = .39$, $SE = .18$; 95% CI = [.08, .81].

Other potential consequences: Satisfaction and interpersonal closeness. Participants in the typo condition reported greater satisfaction with the conversation ($M = 5.27$, $SD = 1.32$) than those in the no-typo condition ($M = 4.79$, $SD = 1.50$), $F(1,188) = 5.42$, $p = .021$, $\eta^2 = .028$. Perceived humanness mediated the effect of the typo manipulation on satisfaction, $B = 0.36$, $SE = 0.12$; 95% CI = [0.14, 0.61].

Participants in the typo condition also reported feeling closer to the chat agent than those in the no-typo condition ($Ms = 3.78$ and 3.23 , $SDs = 1.59$ and 1.62 , respectively), $F(1,188) = 5.62$, $p = .018$, $\eta^2 = .029$. Perceived humanness again mediated the effect of typo on how close participants felt to the agent, $B = 0.44$, $SE = 0.14$; 95% CI = [0.17, 0.74].

Control variable: Privacy concerns. Finally, we analyzed whether participants' individual differences in privacy concerns moderated their information-sharing behavior and intentions by constructing two linear mixed models with privacy concerns, typo manipulation, and their interaction term as predictor variables and 1) participants' information-sharing behavior in the actual chat, and 2) their responses to the information-sharing index as dependent variables. Although we preregistered our intention to analyze the data with ANCOVA, we later learned that mixed models provide a more appropriate test because they allow us to test whether privacy concerns moderate the relationship between typo manipulation and information disclosure. We found that people's privacy concerns did not have a significant impact on their information disclosure behavior, $t(185) = -.13$, $p = .90$, $\eta^2 = .007$, nor did it interact with the typo

manipulation, $t(185) = -.93, p = .35, \eta^2 = .052$, whereas typos still had a marginally significant impact on information-sharing behavior, $t(185) = 1.79, p = .076, \eta^2 = .015$.

Regarding people's intention to disclose personal information on the information-sharing index, we found that privacy concern also did not show a significant impact, $t(185) = -1.21, p = .23, \eta^2 = .052$, nor did it interact with the typo manipulation, $t(185) = -1.36, p = .17, \eta^2 = .010$, whereas typos again had a significant impact on information-sharing index, $t(185) = 3.58, p < .001, \eta^2 = .077$.

Supplemental Method and Results for Study 1C

We conducted Study 1C to fulfill three goals. First, we sought to conceptually replicate how typos impact perceived humanness and helpfulness of a chat agent in different customer service contexts. Second, we explored additional behavioral consequences particularly relevant to customer service. Third, we counterbalanced the order of humanness and other consequence measures to rule out the possibility that explicitly asking participants to first consider the humanness of the chat agent might have led them to form more positive reactions about the typo-correcting agent afterwards. We expected that participants would have more favorable perceptions of the agent in the typo condition than in the no-typo condition regardless of whether perceptions of humanness were measured first or last.

Method

Participants and design. This study employed a 2 (error: typo, no-typo) \times 3 (service context: billing, product, shipping) between-subjects design. We aimed to recruit 50 participants in each of the six experimental conditions, thus 300 participants in total. A total of 310 participants were recruited in exchange for \$0.28 each ($M_{\text{age}} = 34.19, SD_{\text{age}} = 12.16, 47.1\%$

female). Among them, nine participants did not complete the survey and were thus excluded, resulting in 301 participants in the final analysis.

Procedure. Participants were randomly assigned to one of six conditions. For the context manipulation, participants imagined that they were experiencing a billing, shipping, or product issue with their cellphone company and that they logged into the company's website and started to chat with a customer service agent. Participants then saw a screenshot displaying a customer service agent's greeting message. The typo condition included a typo in one of the messages (misspelling "else" as "esle") that was corrected in a subsequent message by the same agent, whereas the no-typo condition included identical messages without typo or correction. After viewing the agent's greeting messages, participants were asked to answer a few questions.

Measures. To shorten this survey, we shortened our measure of the agent's *perceived humanness*, using a single item with clear face validity: "To what extent does this agent seem human?" (from 1 – *not at all human* to 7 – *very human*). Importantly, we counterbalanced the order of the *perceived humanness* item with other consequence measures, such that half of the participants answered the humanness item first, and the other half answered the following consequence measures first.

To examine potential consequences of the perception of humanness, we first measured participants' *perceptions of agent helpfulness* with the same items as in previous studies ($r = .84$). This was followed by a new item measuring participants' *interest in future interactions with the agent* (i.e., "How much would you like to work with this agent again in the future?"). Next, participants were asked to imagine that this agent told them about a new unlimited high-speed wi-fi service and reported their likelihood of *considering the agent's recommendation* on a single item, "How likely would you be to look into this special offer?" All ratings ranged from 1

(*not at all*) to 7 (*very much/likely*). Finally, participants indicated what typo the agent made (if any) as a manipulation check question, and then reported their age and gender.

Results

Manipulation check. Among all participants, 92.5% (282 participants) passed the manipulation check (130 out of 152 in the corrected-typo condition, 152 out of 153 in the no-typo condition). Results remained the same when we analyzed the data including or excluding participants who failed the manipulation check. Therefore, the results reported below include all participants.

Perceived humanness. A two-way ANOVA revealed a significant effect of the typo manipulation on perceived humanness of the chat agent ($F(1, 295) = 155.18, p < .001, \eta^2 = .345$). Specifically, when the chat agent corrected the typos, participants perceived the agent to be significantly more human ($M = 5.15, SD = 1.78$) compared to when the agent did not make a typo ($M = 2.73, SD = 1.56$). The main effect of scenario was not significant ($F(2, 295) = .394, p = .674, \eta^2 = .003$), nor was the interaction effect ($F(2, 295) = .211, p = .810, \eta^2 = .001$).

Perceived helpfulness. A two-way ANOVA revealed a significant main effect of typo ($F(1, 295) = 55.06, p < .001, \eta^2 = .157$). Specifically, participants in the corrected-typo condition expected the agent to be more helpful ($M = 5.12, SD = 1.25$) than those in the no-typo condition ($M = 4.02, SD = 1.31$). The main effect of scenario was not significant ($F(2, 295) = .569, p = .567, \eta^2 = .004$), nor was the interaction effect ($F(2, 295) = .371, p = .690, \eta^2 = .003$).

Consideration of agent's recommendation. A two-way ANOVA on customers' likelihood to consider the agent's recommendation revealed a significant main effect of typo ($F(1, 295) = 4.99, p = .026, \eta^2 = .017$). Specifically, participants in the corrected-typo condition reported a higher likelihood of looking into the agent's recommendation ($M = 3.88, SD = 1.77$) than those

in the no-typo condition ($M = 3.41$, $SD = 1.85$). The main effect of scenario was not significant ($F(2, 295) = .062$, $p = .940$, $\eta^2 < .001$), nor was the interaction ($F(2, 295) = .551$, $p = .577$, $\eta^2 = .004$).

Interest in future interactions. A two-way ANOVA on the willingness to use the agent in future interactions revealed a significant main effect of typo ($F(1, 295) = 32.51$, $p < .001$, $\eta^2 = .099$). Specifically, participants in the corrected-typo condition were more interested in interacting with the same agent again in the future ($M = 4.66$, $SD = 1.35$) than those in the no-typo condition ($M = 3.72$, $SD = 1.48$). The main effect of scenario was not significant ($F(2, 295) = .758$, $p = .470$, $\eta^2 = .005$), nor was the interaction ($F(2, 295) = .264$, $p = .768$, $\eta^2 = .002$).

Order effects. We examined whether explicitly considering the agent's humanness impacted the humanizing effect of corrected typos and other downstream consequences. Across all outcome variables, no significant interactions emerged between the typo manipulation and the order manipulation ($ps > .070$), suggesting that measurement order did not moderate the effects of corrected typos.

Mediation effects. Perceived humanness of the agent mediated the difference in perceived helpfulness between the corrected-typo condition and the no-typo condition ($b = -1.02$, $SE = .12$; 95% CI = [-1.27, -.78]). This mediation was significant both when humanness was measured first ($b = -1.22$, $SE = .18$; 95% CI = [-1.60, -.87]) and when it was measured second ($b = -.81$, $SE = .15$; 95% CI = [-1.13, -.53]).

Finally, perceived humanness of the agent mediated the difference in likelihood of considering the agent's recommendation between the corrected-typo condition and the no-typo condition ($b = -.73$, $SE = .15$; 95% CI = [-1.04, -.42]). This mediation was significant both when humanness was measured first ($b = -.52$, $SE = .24$; 95% CI = [-.98, -.05]) and when it was

measured last ($b = -.89$, $SE = .21$; 95% CI = [-1.33, -.50]). In addition, perceived humanness also mediated the difference in participants' interest in future interaction with the agent between the corrected-typo condition and the no-typo condition ($b = -1.16$, $SE = .13$; 95% CI = [-1.44, -.91]), which was also significant both when humanness was measured first ($b = -1.47$, $SE = .21$; 95% CI = [-1.91, -1.08]) and when it was measured last ($b = -.88$, $SE = .17$; 95% CI = [-1.25, -.57]).

Discussion

Study 1C showed that across different service contexts, online service agents who made and corrected a typo in text messages were seen as more human than agents who made no typos. Furthermore, making and correcting a typo increased participants' interest in the agent's recommended product and future interactions with the same agent. Finally, Study 1C indicated that the effect of typo on these behavioral consequences was robust to whether participants explicitly considered the agent's perceived humanness before or after reporting their behavioral intentions.

Supplemental Method and Results for Study 2

Measures not reported in main text

In this study, we measured participants' information-sharing behavior during the conversation when they were asked a total of ten questions and were allowed to skip as many as they wanted. We also measured their information-sharing intention with the 18-item *information-sharing intention index*. In addition, participants rated their *overall satisfaction* after the chat ($r = .84$) and their *interpersonal closeness* to the agent ($\alpha = .91$).

Results on measures not reported in main text

Information disclosure. A one-way ANOVA on the number of questions participants actually answered revealed a marginally significant effect of the typo manipulation, $F(2, 515) =$

2.79, $p = .062$, $\eta^2 = .011$ ³. Pairwise comparisons further showed that the difference between the no-typo condition ($M = 7.40$, $SD = 2.42$) and the uncorrected-typo condition ($M = 7.94$, $SD = 2.11$) was marginally significant, $p = .055$, $d = 0.23$. However, inconsistent with our prediction, the difference between the corrected-typo condition ($M = 7.77$, $SD = 2.13$) and the no-typo condition was non-significant, $p = .27$, $d = .16$, as was that between the corrected-typo condition and the uncorrected-typo condition, $p = .76$, $d = .008$.

Given that only the corrected-typo condition showed an increase in perceived humanness, we next constructed two mediation models that contrasted the corrected-typo condition against either the no-typo or uncorrected-typo conditions. The first model showed that, consistent with previous studies, perceived humanness of the chat agent significantly mediated the relationship between the typo manipulation and participants' information-sharing behavior, $b = .14$, $SE = .06$, $95\% CI = [.03, .28]$. By contrast, the second model showed that perceived humanness did not significantly mediate the relationship between the corrected-typo vs. uncorrected-typo contrast and the information-disclosure behavior, $b = -.06$, $SE = .04$; $95\% CI = [-.15, .01]$, $p = .24$.

A one-way ANOVA on participants' information-disclosure intention showed that participants did not report significantly different intentions across conditions, $F(2, 515) = 0.34$, $p = .71$; on average, most participants were only willing to disclose two to three personal information items (corrected-typo: $M = 2.45$, $SD = 2.48$; no-typo: $M = 2.57$, $SD = 3.14$; uncorrected-typo: $M = 2.71$, $SD = 3.06$). However, mediation analyses showed that when comparing the corrected-typo and no-typo conditions, consistent with results in earlier studies, humanness perception again mediated the impact of typo manipulation on participants'

³ Because the number of questions participants answered was skewed towards the maximum value, we also conducted a Kruskal-Wallis Test—a non-parametric alternative to one-way ANOVA—and found a marginally significant omnibus effect across conditions: *Kruskal-Wallis* $\chi^2(2) = 4.48$, $p = .106$.

information-sharing behavior, $b = .27$, $SE = .10$, 95% CI = [.09, .49]. Furthermore, the second model contrasting the corrected- and uncorrected-typo conditions showed that perceived humanness also mediated the impact of typo manipulation on information-disclosure intention, $b = .26$, $SE = .11$, 95% CI = [.08, .48].

Other potential consequences. One-way ANOVA showed no effect of typo manipulation on participants' overall satisfaction with their chat experience, $F(2, 514) = .586$, $p = .56$ (corrected-typo: $M = 5.09$, $SD = 1.22$; no-typo: $M = 4.94$, $SD = 1.31$; uncorrected-typo: $M = 5.02$, $SD = 1.41$). However, consistent with previous studies, perceived humanness mediated the effect of the typo manipulation on satisfaction between the corrected-typo condition and the no-typo condition, $B = 0.19$, $SE = 0.06$; 95% CI = [0.08, 0.33].

We did not find a significant effect of typo manipulation on interpersonal closeness, $F(2, 515) = .062$, $p = .54$ (corrected-typo: $M = 3.85$, $SD = 1.58$; no-typo: $M = 3.65$, $SD = 1.53$; uncorrected-typo: $M = 3.74$, $SD = 1.68$). However, perceived humanness again mediated the effect of the typo manipulation on satisfaction between the corrected-typo condition and the no-typo condition, $B = 0.35$, $SE = 0.11$; 95% CI = [0.14, 0.56].

Supplemental Method and Results for Study 3

Measures not reported in main text

Participants responded to the same items as in Study 1A on the *perceived warmth* ($\alpha = .93$) and *competence* ($\alpha = .89$) of the agent, their *willingness to reward the agent*, and their *impressions of the company* (including a new item on business ethics, "I expect this cellular company to treat its customers fairly"; $\alpha = .94$). As secondary measures, participants indicated how error-prone and careless the agent seemed to be ($r = .76$), to assess an additional potential

downside of making typos, and they also reported their *information-disclosure intention* using the six-item index in Study 1A ($\alpha = .69$) and indicated how likely they would be to share personal information with the chat agent on a scale from 1 (*not at all likely*) to 7 (*very likely*). Finally, participants completed manipulation checks on the agent's identity ("a human employee," "an AI-powered chatbot," or "it wasn't mentioned") and what typo the agent made (if any), before reporting their gender, age, and comments about the study.

Results on measures not reported in main text

Perceived warmth and competence. A two-way ANOVA revealed a significant main effect of the typo manipulation on perceived warmth ($F(1, 380) = 8.96, p = .003, \eta^2_p = .023$), such that participants perceived the agent to be warmer when it made and corrected a typo ($M = 5.01, SD = 1.26$) than when it did not ($M = 4.61, SD = 1.35$). It also revealed a significant main effect of agent identity ($F(2, 380) = 3.27, p = .039, \eta^2_p = .017$), whereby participants perceived the agent to be warmer when the agent was a human ($M = 5.05, SD = 1.24$) compared to when its identity was ambiguous ($M = 4.73, SD = 1.35; t(257) = 2.98, p = .024, d = .24$) or a bot ($M = 4.64, SD = 1.34; t(254) = 2.55, p = .006, d = .31$). There was a non-significant interaction ($F(2, 380) = 1.48, p = .23, \eta^2_p = .008$). Examining the effect of the typo manipulation separately for each agent identity condition revealed a significant effect when the agent was ambiguous ($M_s = 5.01$ vs. $4.51, SD_s = 1.35, 1.31; t(128) = 2.14, p = .034, d = .38$) or was known to be human ($M_s = 5.35$ vs. $4.74, SD_s = 1.06, 1.35; t(127) = 2.86, p = .005, d = .50$), but not when it was known to be a chatbot ($M_s = 4.68$ vs. $4.60, SD_s = 1.30, 1.41; t(125) = .33, p = .37, d = .05$). Finally, perceived humanness of the agent mediated the difference between the no-typo condition and the corrected-typo condition ($b = .58, SE = .08; 95\% CI = [.43, .75]$).

By contrast, results on perceived competence showed no significant main effects of either the typo manipulation or agent identity, nor an interaction effect ($F_s < 1.08, p > .34$), again suggesting that correcting the typo did not harm perceptions of competence. The effect of the typo manipulation on perceived competence was not significant for any of the three agent identity conditions when tested in separate pairwise comparison tests ($p_s > .33$).

Willingness to reward the agent. A two-way ANOVA revealed a significant main effect of agent identity ($F(2, 380) = 38.10, p < .001, \eta^2_p = .17$), such that participants were more inclined to reward a human employee ($M = 4.44, SD = 1.58$) than an ambiguous agent ($M = 3.91, SD = 1.61; t(257) = 2.66, p = .008, d = .33$), and an ambiguous agent than a chatbot ($M = 2.73, SD = 1.64; t(255) = 5.83, p < .001, d = .73$). There was no main effect of the typo manipulation nor any interaction effect ($F_s < 1.84, p > .16$). Examining the effect of the typo manipulation separately for each agent identity condition, participants were directionally more inclined to reward the typo-correcting agent than the no-typo agent when the agent's identity was ambiguous ($M_s = 4.15$ vs. $3.72, SD = 1.52$ vs. $1.67; t(128) = 1.54, p = .13, d = .27$); by contrast, this pattern was completely absent when the agent was known to be a chatbot or a human ($p_s > .64$). Finally, perceived humanness of the agent significantly mediated the relationship between the typo manipulation and participants' willingness to reward the agent ($b = .73, SE = .11; 95\% CI = [.53, .95]$).

Impressions of the company. A two-way ANOVA showed a significant main effect of agent identity ($F(2, 380) = 8.19, p < .001, \eta^2_p = .041$), such that participants perceived a company using an ambiguous chat agent ($M = 4.60, SD = 1.16$) similarly positively to one using a human employee ($M = 4.69, SD = 1.31; t(257) = .56, p = .58, d = .07$) and more positively than one using a chatbot ($M = 4.09, SD = 1.35; t(211) = 3.27, p = .001, d = .41$). There was no main effect

of the typo manipulation nor was there an interaction effect ($F_s < .55, p_s > .49$). Examining the effect of the typo manipulation separately for each agent identity condition showed that the effect of the typo manipulation did not achieve statistical significance for any identity condition ($p_s > .35$), including the key condition in which the agent was ambiguous ($M_s = 4.69$ vs. 4.53 , $SD = 1.30$ vs. 1.04 , $t(128) = 0.75, p = .45, d = .13$). Nonetheless, perceived humanness showed a robust mediation effect on participants' impressions of the company, such that participants formed more favorable impressions of the company when they perceived the agent to be more human ($b = .64, SE = .08; 95\% CI = [.47, .81]$).

Perceived carelessness. Participants' responses on how careless and error-prone the agent seemed to be revealed a significant main effect of the typo manipulation ($F(1, 311) = 109.15, p < .001, \eta^2_p = .26$), yet no main effect of agent identity, and no interaction effect ($F_s < 1.51, p > .22$). As we have speculated, regardless of the agent's identity, participants who saw the agent make and then correct a typo perceived the agent to be more careless and error-prone than an agent who did not make any typos ($M_s = 3.84$ vs. $2.21, SD = 1.42$ & 1.35 , for the corrected-typo and no-typo conditions, respectively).

Information disclosure. The single-item disclosure intention question revealed a marginally significant main effect of the typo manipulation, $F(1, 311) = 3.03, p = .083, \eta^2_p = .009$, and a significant main effect of agent identity, $F(2, 311) = 3.35, p = .036, \eta^2_p = .22$, yet no interaction effect, $F(2, 311) = 1.28, p = .28, \eta^2_p = .008$. Overall, participants in the corrected-typo condition reported a slightly higher intention to disclose personal information than those in the no-typo condition ($M_s = 3.76$ vs. $3.46, SDs = 1.43$ & 1.64).

However, the six-item disclosure index revealed no main effect of the typo manipulation or of agent identity, nor did it reveal an interaction, $F_s < 0.87, p_s > .42$. Even though participants

in the corrected-typo condition indicated that they would be directionally more willing to share personal information ($M_s = 3.76$ vs. 3.46 , $SDs = 1.43$ & 1.64), the results were not significantly different from those in the no typo condition.

Mediation analyses on information disclosure. Across studies, humanness perception consistently mediates the effect of corrected typo on various consequences, yet for some consequences—information disclosure in particular—there was no direct effect of typo manipulation, suggesting that another mediation path might exist simultaneously and negatively mediate the relationship between the typo manipulation and information disclosure. Hence, we sought to test if perceived carelessness might negatively mediate the effect of corrected typo on information disclosure.

Mediation analyses on the single-item information-disclosure intention measure (PROCESS Macro in SPSS; Model 4) showed that perceived humanness of the agent positively mediated the effect of typo on disclosure intention, $b = .39$, $SE = .09$, $95\% CI = [.23, .58]$; simultaneously, perceived carelessness negatively mediated the effect of typo on disclosure intention, $b = -.32$, $SE = .11$, $95\% CI = [-.54, -.12]$ (see Figure S2).

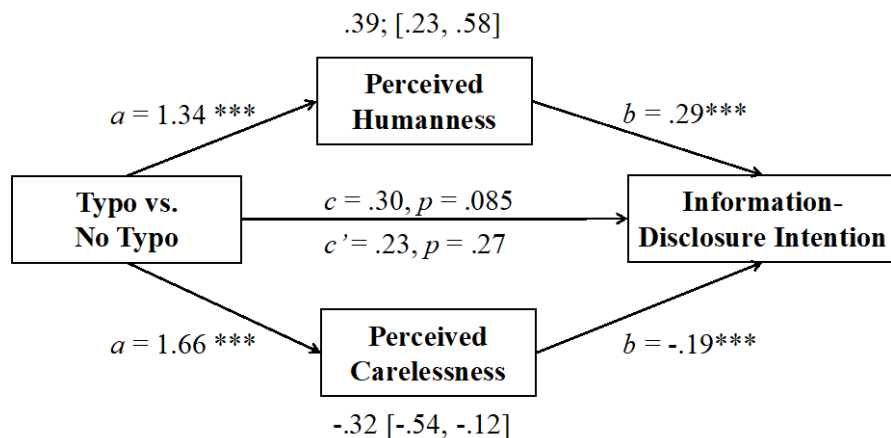


Figure S2. Results of mediation analysis on information-disclosure intention in Study 3, with perceived humanness and perceived carelessness as simultaneous mediators.

Taken together, the simultaneous presence of both a positive mediation path (i.e., perceived humanness) and a negative mediation path (i.e., carelessness) explained why the overall associations between the typo manipulation and various information disclosure measures have been weak or inconsistent. That is, while seeing corrected typos has always led participants to perceive greater humanness in the agent (thus increasing their willingness to provide private information to the agent), it could also lead people to perceive the agent as more careless, and hence decrease people's information-disclosure intention—potentially enough to even completely cancel out any increase from the increased humanness perception.

Discussion

Perceived carelessness negatively mediated the impact of corrected typos on information disclosure and canceled out the positive impact of increased humanness perception on information disclosure. In light of these findings, future research should investigate whether making an agent appear more human, yet no more careless, might lead consumers to become more inclined to disclose personal information.

Supplemental Study 1: Typo vs. Other Humanizing Cues

To compare the effect of typo to other potentially humanizing cues, this study examined not just typos but also two common and possibly humanizing features in online chats: whether an agent's profile photo portrayed a real human, and whether the agent introduced themselves by

name. In addition, we manipulated the agent's gender (male or female) to test whether agent gender might impact perceived humanness or interact with the effect of typo.

Method

Participants and design. We aimed to collect 50 participants in each of the sixteen experimental conditions, thus 800 participants across conditions. A total of 929 adults ($M_{\text{age}} = 32.62$, $SD_{\text{age}} = 12.02$; 53.1% female) from Prolific started our survey in exchange for \$0.64. Among them, 114 participants never completed the survey and were thus excluded, resulting in 815 participants in the final analysis ($M_{\text{age}} = 32.60$; 51.8% female). This study employed a 2 (error: typo, no-typo) \times 2 (profile photo: human, avatar) \times 2 (name: present vs. absent) \times 2 (agent gender: male, female) between-subjects design (16 separate conditions).

Procedure. Participants learned that their task was to read and evaluate online chat messages written by a service agent of a cellular company and then saw a screenshot displaying a customer service agent's greeting messages. Similarly to in Study 1C, the typo condition included a typo in one of the messages (misspelling "else" as "esle") that was corrected in a subsequent message by the same agent ("*Else**"; see Figure 2 and OSF for stimuli presentation). Messages in the no-typo condition included no typo or correction, but were otherwise identical. Along with the messages, participants saw either an avatar or a human photo (i.e., photo manipulation) depicting either a male or a female (i.e., gender manipulation). The agent also either introduced him/herself ("I am Angela [Michael], your customer service agent") or did not ("I am your customer service agent"; i.e., name manipulation). After viewing the agent's greeting messages, participants completed a survey evaluating the agent.

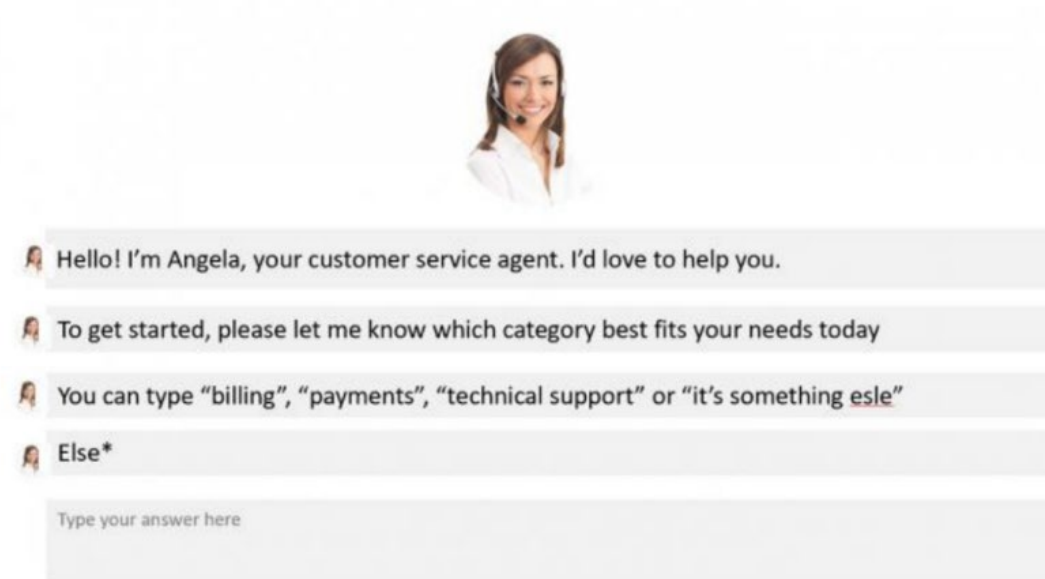


Figure S3. Example of experimental stimuli in the typo condition in Supplemental Study 1 (human photo / name present / female condition).

Measures. Participants reported their *perception of the agent's humanness* on a four-item scale as in previous studies ($\alpha = .89$), followed by how likely they would be to share personal information with the chat agent on a scale from 1 (*not at all likely*) to 7 (*very likely*). Participants then reported their *perceptions of the agent's warmth* and *competence*, their *willingness to reward the agent*, and their *expectations of the agent's helpfulness*. Next, participants reported their *impressions of the company* as in Study 1A. Finally, participants completed the study by responding to manipulation checks regarding whether the agent had a name, what their profile picture looked like, and what typo the agent made (if any), and then reported their demographic information.

Results

Manipulation checks. Among all participants, 89.1% passed the typo manipulation check (406 out of 411 in the no-typo condition, 328 out of 412 in the typo condition), 98.9% passed the

photo attention check (407 out of 411 in the no-typo condition, 407 out of 412 in the typo condition), and 77% passed the name attention check (328 out of 411 in the no-typo condition, 306 out of 412 in the typo condition). All findings remained the same when our analyses included or excluded those who failed the manipulation checks. Therefore, the results reported below are based on all participants.

Perceived humanness. A four-way ANOVA on the composite score of perceived humanness revealed a main effect of typo: Participants perceived the agent to be more human when the agent made and corrected a typo in the messages ($M = 4.08$, $SD = 1.81$) than when the agent did not make a typo ($M = 2.44$, $SD = 1.21$; $F(1, 798) = 227.33$, $p < .001$, $\eta^2 = .22$; see Figure S4). We also found a significant, although notably smaller, main effect of profile photo ($F(1, 798) = 4.19$, $p = .041$, $\eta^2 = .005$), such that participants perceived the agent with a human photo to be slightly more human ($M = 3.36$, $SD = 1.75$) than one with an avatar photo ($M = 3.18$, $SD = 1.74$). By contrast, mentioning the agent's name only led to a marginally significant increase in perceived humanness (name: $M = 3.34$, $SD = 1.74$; no name: $M = 3.20$, $SD = 1.74$; $F(1, 798) = 2.84$, $p = .092$, $\eta^2 = .004$), and the agent's gender showed no impact on perceived humanness ($F(1, 798) = .274$, $p = .601$, $\eta^2 < .001$).

Of all possible interaction effects, only two emerged as marginally statistically significant: The interaction between typo and profile photo ($F(1, 798) = 3.04$, $p = .081$, $\eta^2 = .004$) reflected a slightly larger humanizing effect of typo when the profile photo depicted a human (typo: $M = 4.29$, $SD = 1.72$; no typo: $M = 2.27$, $SD = 1.25$) than when it depicted an avatar (typo: $M = 3.88$, $SD = 1.88$; no typo: $M = 2.43$, $SD = 1.19$), and the interaction between gender and name ($F(1, 798) = 2.93$, $p = .087$, $\eta^2 = .004$) reflected a slightly larger humanizing effect of name when the agent was a female (typo: $M = 4.36$, $SD = 1.73$; no typo: $M = 2.45$, SD

= 1.04) than when it was male (typo: $M = 4.09$, $SD = 1.82$; no typo: $M = 2.48$, $SD = 1.28$). No other two-way, three-way, or four-way interactions reached statistical significance ($F_s < 2.13$, $p_s > .144$, $\eta^2_s < .001$).

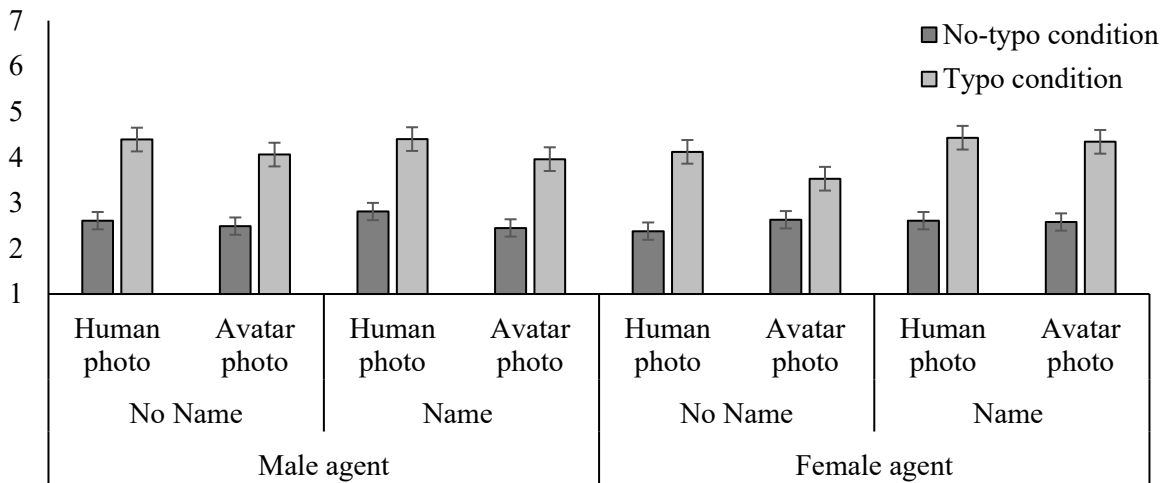


Figure S4. Participants' perceptions of an agent's humanness. Error bars represent ± 1 standard error around the means.

Perceived helpfulness. Supporting our predictions, participants in the typo condition believed that the agent would be able to better understand them and solve their problems ($M = 4.39$, $SD = 1.59$) than did those in the no-typo condition ($M = 3.43$, $SD = 1.65$; $F(1, 798) = 69.99$, $p < .001$, $\eta^2 = .081$). We found no main effects for other factors ($F_s < 2.15$, $p_s > .14$), and only a marginally significant two-way interaction between typo and name ($F(1, 798) = 3.73$, $p = .054$, $\eta^2 = .005$). Specifically, when an agent introduced themselves by name, their typo led to an even greater increase in perceived understanding and problem-solving ability (typo: $M = 4.51$, $SD = 1.57$; no typo: $M = 3.33$, $SD = 1.71$) than when they did not mention their name (typo: $M = 4.28$, $SD = 1.60$; no typo: $M = 3.53$, $SD = 1.62$). No other two-, three- or four- way interaction

effects between factors were significant ($F_s < 2.33$, $p_s > .127$, $\eta^2_s < .003$). Mediation analysis further showed that perceived humanness mediated the relationship between the typo manipulation and participants' expectations about the agent's helpfulness ($b = .59$, $SE = 0.04$, $95\% CI = [.50, .68]$)

Information disclosure. A four-way ANOVA showed no main effect of either typo, $F(1, 798) = .027$, $p = .870$, $\eta^2 < .001$, nor any other factors, $p_s > .26$. Two-way interaction between typo and agent's gender was marginally significant, $F(1, 798) = 3.62$, $p = .057$, $\eta^2 = .005$, suggesting that typo led to marginally more information-sharing intention when the agent was presented as a female (typo: $M = 2.83$, $SD = 1.50$; no typo: $M = 2.65$, $SD = 1.49$) rather than a male (typo: $M = 2.74$, $SD = 1.46$; no typo: $M = 2.92$, $SD = 1.54$). We also observed a marginally significant three-way interaction between typo, profile photo, and gender, $F(1, 798) = 2.79$, $p = .095$. No other two-, three-, or four-way interaction reached statistical significance, $p_s > .14$.

Although we observed no main effect of the typo manipulation, we did find that the chat agent's perceived humanness mediated the effect of typo on participants' intention to share information, $b = .39$, $SE = .06$, $95\% CI = [.27, .51]$.

Perceived warmth and competence. Participants perceived the agent who made and corrected a typo to be warmer ($M = 4.42$, $SD = 1.36$) than one who did not ($M = 3.85$, $SD = 1.54$; $F(1, 798) = 33.24$, $p < .001$, $\eta^2 = .040$). Participants also perceived an agent who introduced himself by name to be slightly warmer ($M = 4.24$, $SD = 1.48$) than one who did not ($M = 4.04$, $SD = 1.48$; $F(1, 798) = 4.01$, $p = .045$, $\eta^2 = .005$). We found no main effect on warmth of the agent's profile photo ($F(1, 798) = .481$, $p = .488$, $\eta^2 = .001$) or gender ($F(1, 798) = .419$, $p = .518$, $\eta^2 = .001$). Finally, we observed a small and unexpected three-way interaction among typo, name, and gender ($F(1, 798) = 4.51$, $p = .034$, $\eta^2 = .006$). No other interactions were significant ($F_s <$

2.14, $ps > .14$). Perceived humanness mediated the effect of typos on warmth perceptions ($b = 0.50$, $SE = 0.04$, $95\% CI = [.41, .59]$).

Regarding the agent's perceived competence, we found that participants actually perceived the agent in the typo condition to be even *more* competent ($M = 4.30$, $SD = 1.27$) than that in the no-typo condition ($M = 4.07$, $SD = 1.38$; $F(1, 798) = 6.96$, $p = .008$, $\eta^2 = .009$). No other main effects ($F_s < 2.16$, $ps > .14$) or interactions were significant ($F_s < 2.34$, $ps > .13$) besides a small three-way interaction among typo, name, and gender ($F(1, 798) = 6.07$, $p = .014$, $\eta^2 = .008$).

Willingness to reward the agent. Participants in the typo condition were more inclined to endorse a reward for the agent ($M = 3.76$, $SD = 1.74$) than those in the no-typo condition ($M = 3.43$, $SD = 1.79$; $F(1, 798) = 29.25$, $p < .001$, $\eta^2 = .035$). In addition, when the agent mentioned its name, participants were also marginally more likely to endorse a reward for the agent (name: $M = 3.54$, $SD = 1.80$; no name: $M = 3.31$, $SD = 1.77$; $F(1, 798) = 3.57$, $p = .059$, $\eta^2 = .004$). No other main effects or interaction effects reached statistical significance ($F_s < 2.68$, $ps > .10$). Perceived humanness mediated the relationship between the typo manipulation and endorsing a reward for the agent ($b = .50$, $SE = 0.04$, $95\% CI = [.42, .59]$).

Impressions of the company. As predicted, participants in the typo condition reported a more favorable opinion toward the company ($M = 4.32$, $SD = 1.31$) than participants in the no-typo condition ($M = 3.87$, $SD = 1.34$; $F(1, 798) = 24.61$, $p < .001$, $\eta^2 = .030$). We found no main effect for other factors ($F_s < 1.08$, $ps > .29$), and only a marginally significant two-way interaction between typo and name ($F(1, 798) = 3.00$, $p = .083$, $\eta^2 = .004$). Specifically, when the agent introduced themselves by name, their typo led to a greater increase in people's favorable impressions of the company (typo: $M = 4.44$, $SD = 1.31$; no typo: $M = 3.84$, $SD = 1.34$) than

when the agent did not mention its name (typo: $M = 4.20$, $SD = 1.30$; no typo: $M = 3.90$, $SD = 1.33$). In addition, we found a marginally significant three-way interaction between typo, name, and gender ($F(1, 798) = 3.67$, $p = .056$, $\eta^2 = .005$). No other interaction effects were significant ($F_s < 2.11$, $p_s > .14$). Finally, mediation analysis showed that perceived humanness mediated the relationship between the typo manipulation and participants' impressions of the company ($b = .42$, $SE = .04$, $95\% CI = [.34, .51]$).

Discussion

This study showed once again that observing a corrected typo in a static online chat, compared to observing no errors, leads people to perceive a customer service agent as more human. Further, the increase in humanness perception was associated with the positive downstream consequence of perceiving the agent to be more helpful. By contrast, seeing a human profile photo had a notably smaller effect on humanness perception and inconsistent effects on the downstream consequences, and neither the presence of the agent's name nor the agent's gender influenced people's perception of the agent's humanness at all.

Supplemental Study S2: Erring Without Correction Using a Button-Based Chatbot

As our first attempt to build a chatbot to investigate our research questions, we started with a button-based chatbot and examined whether making and correcting typos could humanize this chat agent. In addition to the corrected-typo and the no-typo conditions, this study also included a condition in which the agent made but did not correct a typo. Finally, this study featured typos that differed from those in other studies, and the corrections were not marked by asterisks (“*meat*” → “*meet*”; “*plajiarisim*” → “*plagiarism*”), thus introducing more variations to our research stimuli in this project.

Method

Participants. We aimed to recruit at least 100 participants in each of three experimental conditions, but ran more participants than expected because we scheduled data collection sessions ahead of time, and each session lasted for a week. In total, 391 participants ($M_{age} = 19.7$, $SD_{age} = 0.97$, 51.2% female) recruited from a behavioral research laboratory located on a university campus in the northeastern U.S. completed the study in exchange for course credit.

Design and procedure. Upon entering the study, participants were told that “[w]e are currently piloting a team of representatives (including tools that assist them, such as chatbots) who are waiting to interact with current students online.” Then they were told that they would be connected to an online agent named Angela. After a waiting screen that lasted for five seconds, participants were prompted to the next page in Qualtrics and were presented with a chat interface that displayed a female representative with a photo of a real person. As the agent was supposedly typing, a speech bubble blinked until a message showed up and she moved on to typing the next message. After she completed “typing” all messages and asked a question, participants were presented with a forward button, which allowed them to move to the next page and respond to the earlier question by choosing among a set of predetermined options (see Figure S5 for an example). We programmed this interface using customized CSS, HTML, and Java in Qualtrics.

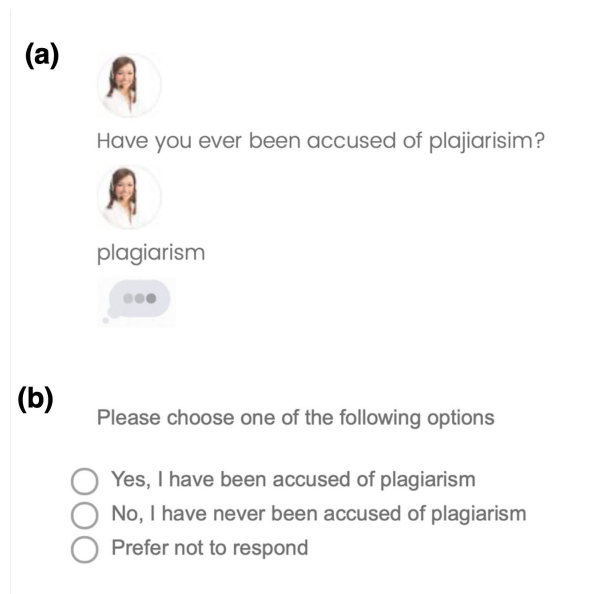


Figure S5. Example of the chat interface in Supplemental Study 2. (a) showcases the end state of a question-asking page, where the agent made and corrected a typo; and (b) showcases the subsequent page displaying multiple choices.

Participants were randomly assigned to either the no-typo condition, the corrected-typo condition, or the uncorrected-typo condition. In both the corrected-typo and the uncorrected-typo conditions, the agent made two typos over the conversation (i.e., “*it is nice to meat you*” at the beginning of the interaction, and “*Have you ever been accused of plajiarisim?*” as the fifth question of the list). In the corrected-typo condition, the agent corrected the typos (“*Meet you*” and “*Plagiarism*”; also note that different from other studies, the agent did not use an asterisk to indicate correction). In the uncorrected-typo condition, the agent continued the conversation without addressing the typos. In the no-typo condition, the agent did not make any typos.

During the conversation, the agent asked participants a total of eight questions—six multiple-choice questions (e.g., “*Who helps you pay for your tuition?*”, “*How often do you find yourself skipping classes?*”, “*Have you ever cheated on an exam?*”, and “*What is the lowest*

amount of payment you are willing to receive for writing someone's final exam?") and two open-ended questions (e.g., "What is the most sensitive issue that you ever encountered as a student at [school name]?") After the agent asked a multiple-choice question, a set of choices would appear on the next screen below the prompt, "Please choose one of the following options"; and after the agent asked an open-ended question, a text box would appear on the next page below a restatement of the question. For each question, participants could choose not to answer—all multiple-choice questions included an option of "Prefer not to answer," and the agent always wrote, "If you do not wish to answer, type S to skip!" in their message following each open-ended question.

Perceived humanness. After the chat conversation, participants were asked to indicate the agent's *perceived humanness* on the same set of three items from previous studies ($\alpha = .61$)⁴. As an additional, separate measure of humanness, we asked participants to indicate on a binary scale whether they thought that the agent was "a person from our customer service team" or "a chatbot." We added this binary *human-or-bot* measure to capture a simple, face-valid assessment of humanness that we expected would show similar results as the perceived humanness index. Participants then completed the *interpersonal impression* scale for warmth ($\alpha = .87$) and competence ($\alpha = .78$).

Other measures. We explored several other potential consequences, including agents' perceived *warmth* ($\alpha = .87$) and *competence* ($\alpha = .78$) on the same scales described in Study 1A in the main text; perceived *responsiveness of the agent* (i.e., "To what extent did you feel understood/heard/valued/supported?", $\alpha = .89$) from 1 (*not at all*) to 7 (*very much*); the extent to

⁴Although the humanness perception scale alpha was lower in Study 2 than in other studies, each individual item showed the same pattern of results. To maintain consistency with the other studies, we report the composite scores in the main text.

which participants felt *negative emotions* (i.e., angry/upset/frustrated, $\alpha = .83$) during the interaction from 1 (*not at all*) to 7 (*very much*); participants' overall *satisfaction* with this interaction on a scale from 1 (*sad face*) to 5 (*happy face*); participants' *intent to use the agent in the future* using two items ($\alpha = .87$): 1) "To what extent do you think Angela would be able to help you with any student challenges you have in the future?", and 2) "How likely are you to use Angela to help you solve a student issue in the future?", both from 1 (*not at all*) to 7 (*very much*); how long participants would be *willing to interact* with the chat agent in the future if they needed help; and how much participants would *trust the agent with their information* using two items ($\alpha = .84$): 1) "Did you feel comfortable sharing your personal information with Angela?", and 2) "To what extent do you trust Angela with your personal information?", both from 1 (*not at all*) to 7 (*very much*).

Finally, participants responded to a *manipulation check* item asking whether the agent made any typos in her messages (*Yes, No, or I do not remember*). Participants who chose "Yes" were then prompted to answer whether the agent corrected the typos (*Yes, No, or I do not remember*). They also reported how comfortable they were with technology, and how often they interacted with chat agents for customer service. Participants concluded the survey by reporting their gender, age, and any additional comments about the study.

Results

Manipulation check. Among all participants, 28.6% responded "I do not remember" to the first manipulation check (70 out of 132 in the no-typo condition, 4 out of 128 in the corrected-typo condition, and 38 out of 131 in the uncorrected-typo condition). Excluding those 112 participants who could not remember whether the chat agent made typos or not, 93.1% of the remaining participants correctly indicated whether the chat agent's messages contained typos

(56 out of 62 in the no-typo condition, 124 out of 124 in the corrected-typo condition, and 80 out of 93 in the uncorrected-typo condition). Among participants who correctly reported that the chat agent made a typo, 97.3% correctly identified whether or not the typo was corrected. All results remained the same when we analyzed the data including or excluding participants who failed the manipulation checks. Therefore, the results reported below are based on all participants.

Perceived humanness. A one-way ANOVA revealed a significant effect of the typo manipulation on perceived humanness of the chat agent, $F(2, 388) = 4.58, p = .011, \eta^2 = .023$. Pairwise comparisons further showed that when the agent corrected the typos, participants perceived the agent to be significantly more human ($M = 2.65, SD = 1.31$) than when there were no typos in the message ($M = 2.25, SD = 1.05$), $t(261) = 2.73, p = .007, d = 0.33$, or when the typos were not corrected ($M = 2.31, SD = 1.10$), $t(258) = 2.30, p = .022, d = 0.28$. Furthermore, when the agent did not make any typos or did not correct the typos, people perceived similar levels of humanness in these agents, $t(261) = -.04, p = .678, d = 0.05$. These results supported our hypothesis that it is the correction of a typo, rather than the mere presence of a typo, that leads people to perceive a chat agent as more human.

We separately analyzed the binary *human-or-bot* question by conducting a logistic regression on the proportions of participants who identified the agent as a human. We found that the proportions across three conditions were indeed significantly different from each other, $Wald(2) = 15.08, p = .001$. Furthermore, *a priori* contrasts showed that when the agent made and corrected typos, people were more likely to classify the agent as a human (20.3%) than when the agent did not make typos (4.7%), $Wald(1) = 6.14, p = .013, OR = .40$; however, the difference between the corrected-typo condition and the uncorrected-typo condition (10.0%) did not reach the .05 level of statistical significance, $Wald(1) = 2.11, p = .15, OR = .47$.

Perceived warmth and competence. A one-way ANOVA revealed no significant effect of the typo manipulation on perceived warmth, $F(2, 388) = .533, p = .587$, but a significant effect on perceived competence, $F(2, 388) = 3.30, p = .038$. Participants perceived the agent who did not make typos ($M = 3.28, SD = 1.00$) as more competent than the agent who made and corrected typos ($M = 2.95, SD = 1.02$), $t(258) = -2.65, p = .08$. There was no difference in perceived competence between the agent who corrected typos and the agent who did not correct typos ($M = 3.07, SD = 1.15$), $t(257) = -.920, p = .358$, or between the agent who did not make typos and the one who made but did not correct typos, $t(261) = 1.57, p = .116$.

Negative emotions. A one-way ANOVA revealed a marginally significant effect of typo manipulation on negative emotions, $F(2, 388) = 2.37, p = .095$, with a significant difference between the no-typo condition ($M = 1.86, SD = 1.30$) and the uncorrected-typo condition ($M = 2.23, SD = 1.42$), $t(261) = -2.15, p = .032$. No other differences were significant.

Finally, we did not find any significant effects of the typo manipulation on perceived *responsiveness of the agent* ($F(2, 388) = .557, p = .573$), *satisfaction with the interaction* ($F(2, 388) = .911, p = .403$), *intent to use the agent in the future* ($F(2, 384) = .328, p = .721$), *how long participants were willing to interact with the agent* ($F(2, 384) = .428, p = .652$), or *trust toward the agent* ($F(2, 388) = 1.19, p = .304$).

Supplemental References

- Appel, Jana, Astrid von der Pütten, Nicole C. Krämer, and Jonathan Gratch (2012), “Does Humanity Matter? Analyzing the Importance of Social Cues and Perceived Agency of a Computer System for the Emergence of Social Reactions during Human-Computer Interaction,” *Advances in Human-Computer Interaction* (2012), 1–10.
- Kang, Sin-Hwa and Jonathan Gratch (2014), “Exploring users’ social responses to computer counseling interviewers’ behavior,” *Computers in Human Behavior*, 34, 120–130.
- Larivière, B., Bowen, D., Andreassen, T.W., Kunz, W., Sirianni, N.J., Voss, C., Wunderlich, N.V. and De Keyser, A. (2017), “Service Encounter 2.0”: An investigation into the roles of technology, employees and customers. *Journal of Business Research*, 79, 238-246.
- LivePerson (2017), “How consumer view bots in customer care,”
<https://www.liveperson.com/resources-reports/bots-in-customer-care>

End Notes

¹ We developed the chat program using JavaScript, CSS, and HTML codes and embedded the interactive chat interface in a Qualtrics survey. Our OSF folder includes a demonstration video of this chatbot and a Qualtrics survey file (.qsf), to allow interested readers to download the code and create basic chatbots for their own research. Interested readers are invited to try the chatbot at: <https://tinyurl.com/humantypo>. Those who wish to deploy similar chatbots in their research may contact the correspondence author. As reported in the main text, this program detected keywords in participants' responses and provided minimal, pre-determined responses to convey listening and understanding. For example, when a participant's response contained a question mark, the program would interpret it as a question and respond, "*Please try your best to answer this question.*" This response worked in most circumstances, yet occasionally participants used question marks for rhetorical or clarification purposes (e.g., "What is your next question?"), which triggered the automatic response and therefore exposed that the agent was an algorithm, warranting data exclusion based on our pre-registered exclusion criteria. Besides scripting the conversation to make the chatbot appear at least somewhat "intelligent," we also implemented a few design features in the chat interface—modeled after a private chat room—that simulated a typical chat experience with a customer service agent. First, before each utterance, the chatbot would pause for a brief moment proportional to the number of words in a participant's previous message, which was meant to convey the impression that the agent spent time reading the message before responding. Second, when the agent was in typing mode, the screen would display "*Angela is typing! Please wait.*" Above the chat box to indicate ongoing activity from the agent and thereby mimic the experience of a chat conversation with service agents. Third, the chatbot was set to be "typing" at a reasonably fast human speed (approximately 7 characters per second, spaces included), creating a realistic waiting time before each message appeared on the screen. As we developed these design features, we also conducted multiple pilot studies to ensure that participants on average perceived the agent to be moderately human.

² Note that for the sake of brevity, we are only able to report participants' perceptions of humanness and helpfulness in the main text. We also measured several other downstream consequences of perceived humanness across our studies, including perceptions of the agent's a) warmth and b) competence (Study 1A, Study 1B, Study 3, Supplemental Study 2), c) perceived interpersonal closeness to the agent and d) satisfaction with the conversation (Study 1A, Study 1B, Study 2), e) willingness to reward the agent and f) impression of the company using the agent (Study 1A, Study 3, Supplemental Study 2), g) consideration of the agent's recommendation and h) interest in future interaction with the agent (Study 1C), and i) whether participants disclose personal information to the agent (Study 1A, Study 1B, Study 2, Study 3, Supplemental Study 2). Many of these secondary measures also have positive effects from corrected typos (e.g., perceived warmth, willingness to reward the agent, consideration of the agent's recommendations), though other measures have less consistent effects (e.g., disclosure of personal information). We describe and fully report all of these measures in the Web Appendix.

³ This study also manipulated whether perceived humanness was measured before or after perceived helpfulness; regardless of the order, participants always perceived the agent in the typo condition to be more helpful, $t_s > 3.93$, $p_s < .001$, ruling out the possibility that any effect of condition on perceived helpfulness resulted only from participants considering the humanness first.

⁴ 44 participants used a question mark that triggered an out-of-context automatic response from the agent and 7 participants attempted to ask the agent multiple questions during the greeting phrase that the agent could not answer.

⁵ Similar to Study 1C, measuring perceived humanness before or after perceived helpfulness did not affect the results. An ANOVA revealed no interaction effects between measurement order and the other factors on perceived humanness ($p_s > .11$) or perceived helpfulness ($p_s > .30$).