Untouchable them: the effect of chatbot gender on angry customers

Shichang Liang
School of Business, Guangxi University, Nanning, China and China-Asean Institute of Financial Cooperation, Guangxi University, Nanning, China, and
Rulan Li and Bin Lan
School of Business, Guangxi University, Nanning, China and Key Laboratory of Interdisciplinary Science of Statistics and Management (Guangxi University), Education Department of Guangxi, Nanning, China
Yuxuan Chu
School of Business, Sun Yat-Sen University, Guangzhou, China
Min Zhang
School of Management, Guang Dong Peizheng College, Guangzhou, China, and
Li Li
School of Business, Nanning University, Nanning, China

Abstract
Purpose – This study explores how chatbot gender and symbolic service recovery may improve the satisfaction of angry customers in the context of service failures. It provides a strategy for companies to deploy chatbots effectively in customer anger.

Design/methodology/approach – This research relies upon a systematic literature review to propose three hypotheses, and we recruit 826 participants to examine the effect of chatbot gender on angry customers through one lab study and one field study.

Findings – This research shows that female chatbots are more likely to increase the satisfaction of angry customers than male chatbots in service failure scenarios. In addition, symbolic recovery (apology vs. appreciation) moderates the effect of chatbot gender on angry customers. Specifically, male (vs. female) chatbots are more effective in increasing the satisfaction of angry customers when using the apology method, whereas female (vs. male) chatbots are more effective when using the appreciation method.

Originality/value – The rapid advancements in artificial intelligence technology have significantly enhanced the effectiveness of chatbots as virtual agents in the field of interactive marketing. Previous research has concluded that chatbots can reduce negative customer feedback following a service failure. However, these studies have primarily focused on the level of chatbot anthropomorphism and the design of conversational texts, rather than the gender of chatbots. Therefore, this study aims to bridge that gap by examining the effect of chatbot gender on customer feedback, specifically focusing on angry customers following service failures.

The authors thank Prof. Yuanyuan Li for her comments and advice on China Marketing International Conference 2022, and thank Linlin Ren, Qin Jiang for their help in experimental design and data collection.

Funding: This research was funded by National Natural Science Foundation of China under Grant (72162002 and 72232003); Guangxi Higher education undergraduate teaching reform project under Grant (2021JGA120); Interdisciplinary Scientific Research Foundation of Applied Economics of Guangxi University under Grant (2023 JJXJ14); National Social Science Fund Projects under Grant (21XZD01); Key Research Base of Humanities and Social Sciences in Guangxi Universities and Guangxi Development Strategy Institute under Grant (2023GDSIZD01); Henan Province Philosophy and Social Science Planning Project under Grant (2021BJJ110); and Guangdong Province Philosophy and Social Science Planning Project (GD20CGL47).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, upon reasonable request.
1. Introduction
Driven by the rapid development of artificial intelligence (AI) technology, the interactive marketing landscape has evolved from direct marketing between individuals to two-way interactions with platform-based virtual conversational agents (Lim et al., 2022; Wang, 2021, 2023). The utility of chatbots as virtual conversational agents has improved significantly in the field of interactive marketing (Araujo, 2018; Shumanov and Johnson, 2021). Duplex, which was designed by Google in 2018, can help users make salon and restaurant reservations in a natural and fluent tone (Sykes and Gonzalez-Lloret, 2020). Chinese shopping platforms such as Taobao’s “Ali Xiaomi” and Jingdong’s “Xiaodong” are trying to improve customer chat experience through the use of social language (Yin and Qiu, 2021). However, do chatbots provide satisfactory service to customers? Imagine that you bought a camera on a shopping platform for an upcoming trip. After receiving the goods, you find that the camera lens is damaged; thus, you will not be able to use the camera on your trip. You are angry and immediately contact the platform’s customer service. At this moment, which gender of customer service chatbot (male vs. female) will be able to pacify your anger? This is the question that this study hopes to address.

Numerous scholars and companies have increasingly focused on how to alleviate customer anger stemming from service failures (Chen et al., 2021; McColl-Kennedy and Sparks, 2003). Customer anger is a negative emotion that can affect the evaluation of services, escalating from initial dissatisfaction to a complaint, negative word of mouth and even resistance toward the company or product (Bougie et al., 2003; Grégoire and Fisher, 2008). Scholars have noted that AI can mitigate the influence of service failures by reducing customers' negative evaluations of companies (Li et al., 2023; Lv et al., 2021). The intelligent design of chatbots allows them to deliver high-level responses, alleviating the negative emotions and results arising from service failures (Fredrickson et al., 2000; Sands et al., 2022). Furthermore, the anthropomorphism of chatbots can influence customer anger after service failures (Go and Sundar, 2019; Shin, 2022), and the high expectations associated with anthropomorphism are the main reason for such negative effects (Nowak and Biocca, 2003). However, these researches mainly focused on the degree of chatbot anthropomorphism and the design of conversational texts rather than the gender of chatbots. Therefore, this study examines the effect of chatbot gender on angry customers following service failures.

People have a solidified pattern of gender perception, believing that males and females behave in specific ways (Eagly, 1995; Eagly and Wood, 1999). Based on gender stereotypes, females are generally considered superior to males in emotional aspects and possess a stronger ability to understand and feel others’ emotions, that is empathy (Bora et al., 2008). Based on these researches, this study suggests that customers perceive female chatbots to be more empathetic toward and understanding of angry customers than male chatbots, leading to higher customer satisfaction. In addition, companies urgently need to implement effective service recovery strategies to alleviate customer anger resulting from service failures (Chen et al., 2021). Symbolic recovery is a timely and effective way to calm customer anger and increase consumer satisfaction (Zhang et al., 2023). To sum up, we propose that symbolic recovery (apology vs. appreciation) moderates the effect of chatbot gender on angry customers following service failures. Specifically, male chatbots are more effective in increasing the satisfaction of angry customers when using the apology method, whereas female chatbots are more effective when using the appreciation method.
2. Theoretical background

2.1 Service failures and customer anger

Service failure refers to the inconsistency between a product or service and consumers’ expectations during the service process, where customers receive something different from what they anticipated (Khamitov et al., 2020). Scholars have classified service failures into various types based on different dimensions, but there is a consensus regarding the classification of result-based and process-based failures (Bagherzadeh et al., 2020; Roschk and Gelbrich, 2014). Result-based failures arise when the products provided by a company fail to meet consumers’ needs, while process-based failures pertain to deficiencies in the service itself during the service delivery process (Jean, 2012; Obeidat et al., 2020). Relatively high probability of service failure is due to the unique nature and complexity of the service (Wu et al., 2023). Customers evaluate companies based on their experiences with products and services, and when a service failure occurs, the customer experience is compromised, leading to dissatisfaction and negative word-of-mouth (Choi and Choi, 2014; Yim, 2023). Meanwhile, negative customer engagement behaviors can also impact potential consumers’ attitudes and behaviors toward service providers (Azer and Alexander, 2020; Azer et al., 2023; Lee, 2023). Moreover, service failures are a root cause of negative emotions, including anger and frustration, and can prompt customers to switch behaviors (Harrison-Walker, 2019).

Alleviating customer anger triggered by service failures has become a focus of attention (Chen et al., 2021; McColl-Kennedy and Sparks, 2003). Customer anger is a negative emotion that can have implications for the evaluation of services, escalating from initial dissatisfaction to complaint, negative word of mouth and even resistance toward the company or product (Bougie et al., 2003; Grégoire and Fisher, 2008). AI can effectively alleviate negative customer reviews after a service failure (Li et al., 2023). Specifically, the intelligent design of chatbots empowers them to provide empathetic responses that demonstrate a high level of understanding, thereby alleviating the adverse effects (Decety and Svetlova, 2012). These empathetic responses align with customers’ objectives when seeking a resolution to problems after experiencing service failures, which make them develop positive evaluations (Ma et al., 2013), ultimately, help to mitigate negative emotions arising from service failures (Fredrickson et al., 2000; Sands et al., 2022). In conclusion, these researches indicate that chatbots can reduce negative feedback from customers following service failures. Thus, our study places particular emphasis on how chatbot gender can influence customer responses in the context of service failures.

2.2 Chatbot and service failure

Chatbots are computer programs developed using AI technology, such programs allow the exchange of text-based messages and images between chatbots and humans conversationally and are widely used by service providers (Brünker et al., 2022). For example, Jill Watson is a teaching assistant designed by Georgia Tech, and the Let’s Talk assistant launched by the American education technology company K12 Insight (Fernoagă et al., 2018; Kim et al., 2019) play the role of a teacher well and provide question-and-answer services. Chinese shopping platform AI is trying to improve customer chat experience using social language (Yin and Qiu, 2021). From the perspective of interactive marketing, Kull et al. (2021) argued that managers can strategically adjust the initial messages of chatbots to promote the connection between users and brands. Other scholars provided a model of user acceptance of chatbots (Wirtz et al., 2018) by endowing chatbots with human-like qualities to increase customer satisfaction with service (Han, 2021; Holzwarth et al., 2006). Chatbot anthropomorphism can improve interactive marketing outcomes. Specifically, chatbots’ human-like interface increases customers’ trust in the technology (Go and Sundar,
Chatbot avatars can also make online shopping experience enjoyable, thereby increasing customers’ purchase intention (Han, 2021). Furthermore, chatbots can influence customer anger after service failures (Crolic et al., 2022; Roschk and Gelbrich, 2014). These researches have predominantly emphasized elements such as the design of conversational texts and the level of anthropomorphism in chatbots, while giving comparatively less consideration to the role of chatbot gender. Our study aims to examine the influence of chatbot gender on customer emotions following service failures, particularly among angry customers.

2.3 Angry customers and chatbot gender

Customers’ reactions to a company are influenced by their emotional state, especially feelings of anger (Ask and Landström, 2010; Bougie et al., 2003). Anger as an individual maladaptive emotion leads to anti-humanizing mental activity (Anderson and Bushman, 2002; DeWall et al., 2011). Therefore, angry customers will likely blame others and produce other negative behaviors (Lench et al., 2016). Unlike other emotions (e.g. sadness, frustration and so on), anger tends to increase action or aggressive behavior and is a key factor influencing customer consumption, psychology, satisfaction and loyalty (Battigalli et al., 2019; Rychalski and Hudson, 2017). Effectively managing the influence of negative emotions resulting from service failures is a crucial concern for companies seeking to improve customer satisfaction (Chen et al., 2021; McColl-Kennedy and Sparks, 2003).

Research has indicated that the intelligent design of chatbots enables them to deliver high-level empathetic responses, which can alleviate the negative emotions and associated effects resulting from service failures (Sands et al., 2022; Yim, 2023). Furthermore, studies on human–computer interaction showed that people tend to endow physical robots with female (vs. male) appearance more communal qualities, including warmth, friendliness and empathy (Eyssel and Hegel, 2012; Stroessner and Benitez, 2019). Such endowed communal qualities are considered to be able to pacify others’ negative emotions, ultimately influencing their behaviors (Borau et al., 2021; Judd et al., 2005). Zogaj et al. (2023) demonstrated the positive influence of matching the gender of chatbots with the gender of consumers on consumer behavior. Moreover, Brahmam and De Angeli (2012) suggest that chatbots’ gender can influence the style of interaction and communication between consumers and chatbots. Additionally, the research investigated the effectiveness of female chatbots in reducing user frustration, enhancing satisfaction and fostering trust in the agent (Hone, 2006).

In summary, we assume that chatbot gender can influence customer satisfaction when customers are angry. Specifically, when chatbots with gender characteristics provide a service, compared with male chatbots, customers perceive female chatbots to possess more empathy and understand the reasons behind their anger, which can directly influence customer satisfaction. Accordingly, we propose the following hypotheses:

\[H1\]. Female chatbots will more likely increase the satisfaction of angry customers than male chatbots.

\[H2\]. Perceived empathy mediates the effect of chatbot gender on the satisfaction of angry customers.

2.4 Service recovery and chatbot gender

Service recovery refers to the actions undertaken by service providers in response to service failures, aiming to address dissatisfaction that arises from such failures, while also serving as a crucial means for companies to restore customer loyalty (Jung and Seock, 2017; You et al., 2020). Extensive discussions have been conducted on the influence of human
employees in the service recovery process (Jeong and Lee, 2017). For example, within the context of service recovery following a failure, male individuals are often regarded as more suitable for managerial roles compared to their female counterparts (Eagly and Karau, 2002; Koenig and Eagly, 2014). Meanwhile, customers frequently possess specific gender-related expectations toward service providers during the service recovery, leading to potential biases that can influence customer satisfaction (Hekman et al., 2010). Furthermore, using AI can ameliorate negative evaluations of companies after service failures (Li et al., 2023; Lv et al., 2021).

To address the adverse effects of service failures, companies express verbal acknowledgment of their mistakes first. Unlike utilitarian recovery (e.g. refunds, coupons and so on), which incurs high costs for a company, symbolic recovery is costless and a timely and efficient way to restore consumer satisfaction (Du et al., 2010; Fang et al., 2019). Symbolic recovery, which refers to a company’s verbal acknowledgment of its failures and provision of psychological compensation to consumers (Jung and Seock, 2017), can be divided into two aspects, namely, apology and appreciation (You et al., 2020). Specifically, an apology means that an individual realized his/her mistake and is conveying his/her submission to others (Simpson et al., 1996), whereas appreciation involves conveying approval and admiration to others by highlighting their advantages and contributions in the service encounter, which can lead customers to shift their attention from their anger to their advantages and contributions (González et al., 2019). Furthermore, the findings of studies on self-consistency suggested that people feel good about being admired and accepted by others (Nia and Zaichkowsky, 2000).

Gender stereotypes state that males possess traits of being upright and unyielding (Kjeldgaard-Christiansen, 2020). Apologizing conveys humility as a way of maintain the receiver’s image and show respect (Ludwig et al., 2022). This humility makes the receiver feel submissive and has a positive impact on increasing customer satisfaction (Farrell et al., 2015; Min et al., 2021). By contrast, females are believed to be tender and vulnerable (Pun et al., 2016; Sultana, 2010). When saying “thank you”, the speaker shows appreciation for the recipient’s behavior and confirms the recipient’s merits by acknowledging his/her contribution (Tesser et al., 1968), which will induce him/her to feel the admiration of the speaker, who is in a vulnerable position, thereby improving the recipient’s self-esteem and generating a positive influence (Grant and Gino, 2010; Kramer et al., 1993). In addition, the introduction of gender stereotypes in nonhuman subjects produced results consistent with those of human subjects (Adam et al., 2021). Accordingly, we propose the following hypotheses:

**H3**. Symbolic recovery (apology vs. appreciation) moderates the effect of chatbot gender on the satisfaction of angry customers.

**H3a**. In the service failures scenario, male (vs. female) chatbots are more effective in increasing the satisfaction of angry customers when using the apology method.

**H3b**. In the service failures scenario, female (vs. male) chatbots are more effective in increasing the satisfaction of angry customers when using the appreciation method.

### 3. Study 1

Study 1 validated H1 by showing that female chatbots are more likely to increase angry customer satisfaction than male chatbots. It was mediated by perceived empathy H2.
3.1 Methods

3.1.1 Participants and design. We recruited 324 residents (222 women; $M_{\text{age}} = 21.29$ years, SD = 6.54) from the community to participate in the study in exchange for a small amount of money (RMB 5). The study involved a $2 \times 2$ (chatbot gender: male vs. female) × 2 (after-sale result: ambiguous vs. resolved) between-subjects design.

3.1.2 Stimulus materials. Previous studies pointed out that males and females represent different colors (Jalil et al., 2012; Wong and Hines, 2015). Furthermore, the auditory and visual systems may induce different behavioral outcomes in individuals (Heimler et al., 2014). Therefore, we showed name cards of different colors to activate the gender chatbot. Meanwhile, we drew on the manipulation of anger by Crolic et al. (2022) and subdivided the post-sale outcome of the anger scenario (ambiguous, i.e. anger group vs. resolved, i.e. control group).

In the ambiguous scenario, the participants read the following:

You buy sunscreen clothes online (RMB 300) for a trip you will take in 15 days. When it arrives, you find that it is not the product you ordered. After contacting the website’s after-sales service (the website uses a chatbot after-sales system), the customer service chatbot says that the sunscreen clothes are currently out of stock, and the replenishment time has yet to be determined, which means that the product you selected may arrive before you depart on your trip or after you leave.

In the resolved scenario, the specific delivery date for the correct product is communicated, ensuring the customer that the product will arrive before the scheduled travel departure date. In the female chatbot condition, we showed the participants a pink-colored female name card, whereas, in the male chatbot condition, we showed the participants a blue-colored male name card (see Figure 1; Appendix 1).

3.1.3 Procedure and measurements. We instructed all the participants to read the post-sale result scenario and then chat with the chatbot based on the post-sale result. After conversing with the chatbots, the participants reported their preference for the name card color and perceived empathy (Czaplewski et al., 2002; Yoon and Lee, 2021) and evaluated

![Figure 1](image_url)

Figure 1.
Stimuli with different color name cards of chatbot gender (Study 1)

Source: Authors’ own work
their satisfaction with the chatbot’s handling of their query. Then, we asked the participants to report their feelings about the gender of the chatbot and their level of anger.

3.2 Results

3.2.1 Manipulation checks. The results of independent samples t-tests showed that the participants in the ambiguous group were significantly angrier than those in the resolved group ($M_{\text{ambiguous}} = 4.80$, SD = 1.35; $M_{\text{resolved}} = 3.32$, SD = 1.59; $t[322] = 8.973$, $p < 0.001$). In addition, we observed significant differences in the participants’ perception of chatbot gender ($M_{\text{male}} = 3.44$, SD = 1.19; $M_{\text{female}} = 4.92$, SD = 1.30; $t[322] = 10.698$, $p < 0.001$).

3.2.2 Customer satisfaction. The interaction between chatbot gender and after-sales results had a significant effect on customer satisfaction ($F[1, 320] = 19.753$, $p < 0.001$, $\eta^2 = 0.058$). The results of the simple-effects analysis revealed that in the ambiguous condition, the female chatbot was more likely to increase customer satisfaction than the male chatbot ($M_{\text{male}} = 3.45$, SD = 1.59; $M_{\text{female}} = 4.74$, SD = 1.32; $F[1, 320] = 39.619$, $p < 0.001$, $\eta^2 = 0.110$). Meanwhile, in the resolved condition, satisfaction with the female or male chatbot did not differ significantly ($M_{\text{male}} = 4.57$, SD = 1.07; $M_{\text{female}} = 4.58$, SD = 1.09; $F[1, 320] = 0.003$, $p > 0.05$, $\eta^2 = 0.000$).

3.2.3 Perceived empathy. We observed the significant effect of chatbot gender interacting with after-sales outcomes on perceived empathy ($F[1, 320] = 10.787$, $p = 0.001$, $\eta^2 = 0.032$). The simple-effects analysis revealed that the female chatbot gained higher perceived empathy than the male chatbot in the ambiguous condition ($M_{\text{female}} = 4.40$, SD = 1.22; $M_{\text{male}} = 3.56$, SD = 1.26; $F[1, 320] = 20.476$, $p < 0.001$, $\eta^2 = 0.060$). This effect was not significant in the resolved condition ($M_{\text{female}} = 4.48$, SD = 1.07; $M_{\text{male}} = 4.50$, SD = 1.06; $F[1, 320] = 0.007$, $p > 0.05$, $\eta^2 = 0.000$; see Figure 2).

3.2.4 Mediating role of perceived empathy. In the moderated mediation analysis (Hayes, 2015; Model 7, 5,000 bootstrap resamples), we used perceived empathy as a mediator and after-sales result as a moderator (1 = "resolved" and –1 = "ambiguous"). We found

![Figure 2](source:Authors' own work)
a significant moderating mediator index ($\beta = 0.3557, \text{SE} = 0.1143, 95\% \text{ CI} = [0.1370, 0.5922]$, see Figure 3). Specifically, the indirect effect of the chatbots on customer satisfaction through perceived empathy was significant in the ambiguous condition ($\beta = -0.3493, \text{SE} = 0.0916, 95\% \text{ CI} = [-0.5368, -0.1762]$), whereas the indirect effect of the chatbots on customer satisfaction through perceived empathy was not significant in the resolved condition ($\beta = 0.0064, \text{SE} = 0.0695, 95\% \text{ CI} = [-0.1280, 0.1437]$).

3.2.5 Alternative explanations. We used name card color preference as a mediating variable (Hayes, 2015; Model 7, 5,000 bootstrap resamples) and did not find that name card color preference mediated the effect of chatbot gender on customer satisfaction in the ambiguous group ($\beta = -0.0650, \text{SE} = 0.0464, 95\% \text{ CI} = [-0.1628, 0.0200]$). In addition, the ANOVA with the participants’ gender as a covariate showed that the participants’ gender had no significant effect on their satisfaction ($F[1, 319] = 0.468, p > 0.05, \eta^2 = 0.001$). We repeated the above analysis with the participants’ frequency of online shopping as the covariate, and the results showed that the frequency of online shopping had no significant effect on customer satisfaction ($F[1, 319] = 0.214, p > 0.05, \eta^2 = 0.000$; see Table I).

4. Study 2
In Study 2, we conducted experiments in realistic conditions. Through field experiments, we provided a replicable strategy for the effect of chatbot gender on customer satisfaction, and verified that in a consumer anger scenario following a service failure chatbot gender on customer satisfaction.

4.1 Methods
4.1.1 Participants and design. We randomly assigned the participants to conditions in a 3 (chatbot: female vs. genderless vs. male) × 2 (symbolic recovery: appreciation vs. apology) × 2 (waiting time: 2 min. vs. 30 min.) between-subjects design. The correlation between waiting time and anger can be found in the Pretest (see Appendix 2). We recruited 502 customers (327 women; $M_{\text{age}} = 28.83$ years, SD = 8.06) at the cafe to participate in this study in exchange for a small amount of money.

4.1.2 Stimulus materials. Cartoon photos can help minimize unrealistic chatbot expectations (De Cicco et al., 2020). Therefore, we used cartoon photos to manipulate the

Indirect effect: .3557 (95%CI = .1370 ~ .5922)
<table>
<thead>
<tr>
<th>Study 1</th>
<th>M (SD)</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer satisfaction</td>
<td>FC</td>
<td>MC</td>
<td></td>
</tr>
<tr>
<td>Chatbot Gender × After-sales Result</td>
<td>Ambiguous (N = 160)</td>
<td>4.74 (1.32)</td>
<td>3.45 (1.59)</td>
</tr>
<tr>
<td></td>
<td>Resolved (N = 164)</td>
<td>4.58 (1.09)</td>
<td>4.57 (1.07)</td>
</tr>
<tr>
<td>Perceived empathy</td>
<td>Chatbot gender</td>
<td>4.02 (1.25)</td>
<td>4.45 (1.14)</td>
</tr>
<tr>
<td>Chatbot Gender × After-sales Result</td>
<td>Ambiguous</td>
<td>4.40 (1.22)</td>
<td>3.56 (1.26)</td>
</tr>
<tr>
<td></td>
<td>Resolved</td>
<td>4.48 (1.07)</td>
<td>4.50 (1.06)</td>
</tr>
</tbody>
</table>

β = -0.3493, SE = 0.0916, 95% CI = [-0.5368, -0.1762]

β = 0.0064, SE = 0.0695, 95% CI = [-0.1280, 0.1437]

### Notes:
- FC = female chatbot group; GC = genderless chatbot group; MC = male chatbot group
- “p > 0.05” = nonsignificant; Statistics in parentheses are standard deviations
- Source: Authors’ own work

---

### Alternative explanation
- Participants’ gender: F [1, 319] = 0.468, p > 0.05, η² = 0.001
- Frequency of online shopping: F [1, 319] = 0.214, p > 0.05, η² = 0.000
- Name card color preference: η = 0.0650, SE = 0.0464, 95% CI = [-0.1628, 0.0200]
gender of the chatbot. Meanwhile, drawing on the classification and manipulation of symbolic recovery of You et al. (2020), in the appreciation condition, the message read, “Thank you for your patience, I really appreciate it!” In the apology condition, the message read, “I’m sorry for keeping you waiting, I sincerely apologize!” For the manipulation of the chatbot’s gender, in the male (female) chatbot group, we presented the participants with a cartoon figure of a male (female) chatbot, Nimeld (Nimilia). In the genderless chatbot group, we presented the participants with a cartoon figure (see Figure 4; Appendix 3).

4.1.3 Procedure and measurements. We conducted this study in a café (which used a chatbot for food orders) in Southern China for one month, starting in June 2022. The research assistant’s clothing and hairstyle were not changed during the entire experiment. In addition, the research assistant wore a face mask, which prevented interference from the research assistant’s appearance in the experiment.

We conducted the study during lunch and dinner time during the day. On the day of the study, a research assistant acted as a waiter in the café (which uses a chatbot ordering system) and controlled the meal delivery time after the customer finished ordering using the ordering system. We conducted the anger manipulation by making the customers wait 2 min or 30 min for their meal. The research assistant handed them an iPad and informed them that it was a survey on the café’s service quality while serving the meal. The iPad showed a picture of a chatbot with a message (see Appendix 4).

After viewing the picture, the participants answered the questions about their preference for the cartoon figure and satisfaction with how the chatbot addressed the problem (“I really like the cartoon photo” and “I’m very satisfied with how the chatbot handled the problem”; 1 = “strongly disagree” and 7 = “strongly agree”; α = 0.89). Finally, the participants reported their feelings about the chatbot’s gender and level of anger.

4.2 Results
4.2.1 Manipulation checks. The results of the independent samples t-tests for waiting time and symbolic recovery showed significant differences in waiting time. Specifically, the
participants in the 30-min group were significantly angrier than those in the 2-min group \((M_{2\text{min}} = 2.23, \text{SD} = 1.21; M_{30\text{min}} = 5.79, \text{SD} = 0.91; t [500] = 36.988, p < 0.001). Moreover, the participants in the appreciation group significantly perceived the chatbot’s appreciation compared with those in the apology group \((M_{\text{apology}} = 2.04, \text{SD} = 1.03; M_{\text{appreciation}} = 4.76, \text{SD} = 0.91; t [500] = 22.758, p < 0.001). In addition, the results of the one-way ANOVA on perceived chatbot gender showed significant differences in the participants’ perception of the chatbot’s gender \((M_{\text{male}} = 1.78, \text{SD} = 0.74; M_{\text{genderless}} = 3.63, \text{SD} = 0.63; M_{\text{female}} = 6.30, \text{SD} = 0.65; F [2, 499] = 1879.611, p < 0.001, \eta^2 = 0.883)."

4.2.2 Moderating role of symbolic recovery. The results of MANOVA showed the significant effect of the interaction between chatbot gender and waiting time on customer satisfaction \((F [2, 490] = 9.875, p < 0.001, \eta^2 = 0.038), and the significant effect of the interaction of chatbot gender, waiting time and symbolic recovery on customer satisfaction \((F [2, 490] = 56.121, p < 0.001, \eta^2 = 0.186). Specifically, the interaction of symbolic recovery and chatbot gender had a significant effect on customer satisfaction when the waiting time was 30 min \((F [2, 490] = 113.665, p < 0.001, \eta^2 = 0.316). The simple pairwise comparison showed that in the 30-min waiting time condition, the female chatbot with an appreciation increased customer satisfaction \((M_{\text{female}} = 5.58, \text{SD} = 0.64; F [2, 490] = 84.985, p < 0.001, \eta^2 = 0.257) compared with the male chatbot \((M_{\text{male}} = 3.01, \text{SD} = 0.72) and genderless chatbot \((M_{\text{genderless}} = 3.54, \text{SD} = 0.91). Similarly, compared with the female chatbot \((M_{\text{female}} = 4.14, \text{SD} = 1.40) and genderless chatbot \((M_{\text{genderless}} = 4.36, \text{SD} = 1.06), the male chatbot induced higher satisfaction with an apology \((M_{\text{male}} = 5.95, \text{SD} = 0.53; F [2, 490] = 47.970, p < 0.001, \eta^2 = 0.163). When the waiting time was 2 min, the interaction of symbolic recovery and chatbot gender did not have a significant effect on customer satisfaction \((F [2, 490] = 0.027, p > 0.05, \eta^2 = 0.001; \text{see Figure 5})."

4.2.3 Alternative explanations. We used preference for the chatbot cartoon figure as a mediator variable (Hayes, 2015; Model 7, 5,000 bootstrap resamples) and did not find that cartoon figure preference mediated the effect of chatbot gender on customer satisfaction in the 30-min waiting time group \((\beta = -0.0133, \text{SE} = 0.0312, 95\% \text{ CI} = [-0.0770, 0.0469]). In addition, the result of ANOVA showed that the participants’ gender had no significant effect on customer satisfaction \((F [1, 489] = 3.016, p > 0.05, \eta^2 = 0.006). We repeated the above analysis with the participants’ online shopping frequency as the covariate, and the results revealed that online shopping frequency had no significant effect on customer satisfaction \((F [1, 489] = 0.727, p > 0.05, \eta^2 = 0.001; \text{see Table II})."

5. General discussion
This research examines how chatbot gender influences the satisfaction of angry customers in the context of service failures. The two studies show that female chatbots are more likely to increase the satisfaction of angry customers in the context of service failures than male chatbots (H1). This result occurs because female chatbots are perceived to be more empathetic than male chatbots (H2). In addition, the symbolic recovery (apology vs. appreciation) moderates the effect of chatbot gender on the satisfaction of angry customers following service failures (H3).

Results from one lab study and one field study provide robust support for our theorizing. In the two studies, we manipulate chatbot gender with different stimulus materials to enhance the robustness of the results. Specifically, different-colored name cards with visual system cues (Study 1), and cartoon photos (Study 2). In addition, to improve the external validity of the experiment, we conduct the field study in Study 2 in a realistic setting (a café) to provide firms with recovery strategies that are easy to copy.
5.1 Theoretical contributions

First, current research contributes to the literature on the influence of chatbot gender within the context of service failures. We provide valuable insights into the role of chatbot gender in shaping customer experiences. Some research demonstrated the positive influence of matching the gender of chatbots with the gender of consumers on consumer behavior (Hone, 2006; Zogaj et al., 2023). However, these researches have predominantly emphasized elements such as the design of conversational texts and the level of anthropomorphism in
### Table II: Statistics descriptive results of Study 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>M (SD)</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Customer satisfaction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chatbot Gender × Waiting Time × Symbolic recovery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 min</td>
<td>Appreciation (N = 122)</td>
<td>FC 5.58 (0.64)</td>
<td>MC 3.01 (0.72)</td>
</tr>
<tr>
<td></td>
<td>Apology (N = 127)</td>
<td>FC 4.14 (1.40)</td>
<td>MC 5.95 (0.53)</td>
</tr>
<tr>
<td>2 min</td>
<td>Appreciation (N = 126)</td>
<td>FC 5.42 (0.63)</td>
<td>MC 5.44 (1.13)</td>
</tr>
<tr>
<td></td>
<td>Apology (N = 127)</td>
<td>FC 5.38 (0.59)</td>
<td>MC 5.45 (1.14)</td>
</tr>
<tr>
<td><strong>Alternative explanation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants' gender</td>
<td>F[1, 489] = 3.016, p &gt; 0.05, ( \eta^2 = 0.006 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency of online shopping</td>
<td>F[1, 489] = 0.727, p &gt; 0.05, ( \eta^2 = 0.001 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cartoon figure preference</td>
<td>( \beta = -0.0133, SE = 0.0312, 95% CI = [-0.0770, 0.0469] )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes**: FC = female chatbot group; GC = genderless chatbot group; MC = male chatbot group; “\( p > 0.05 \)” = nonsignificant; Statistics in parentheses are standard deviations.

**Source**: Authors' own work
chatbots. Therefore, this paper focuses on chatbot gender and service failure interactions affect customer satisfaction, extends the literature on chatbot gender in service failure contexts.

Second, our research adds valuable insights to the literature on service recovery concerning chatbots. Previous research has indicated that consumers tend to hold specific gender-related expectations toward service providers during the service recovery process (Hekman et al., 2010). However, the paper mainly focuses on the influence of the gender of human identity on service recovery strategies. In addition, Chamani (2014) has explicated the influence of gender and social status on the choice of English apology strategies in various communication contexts. Therefore, we analyze how the symbolic recovery strategies can be matched correctly with the gender of a chatbot to reduce downstream negative consequences and increase customer satisfaction.

Finally, our paper enriches research on perceived empathy in the field of AI. Studies on human–computer interaction demonstrated that gender stereotypes apply to nonhuman subjects (Adam et al., 2021), and people perceive female bots to have more communal traits, such as empathy, than male bots (Eyssel and Hegel, 2012; Stroessner and Benitez, 2019). Our study shows that customers perceive female chatbots to be more empathetic toward and understanding of their anger than male chatbots, and this perceived empathy directly affects customer satisfaction under the influence of gender stereotypes.

5.2 Managerial implications
The results of this paper suggest that female chatbots are more likely to increase the satisfaction of angry customers than male chatbots in the context of service failures. The rapid advancements in AI technology have significantly enhanced the effectiveness of chatbots as virtual agents in the field of interactive marketing (Tsai et al., 2021). Customers are increasingly interacting with companies through chatbots, and this market demand has led to the emergence of other bot types. However, female chatbots seem to have a natural market advantage, especially for angry customers following service failures. Therefore, companies should pay attention to chatbot avatars and names and take advantage of female chatbots by giving them avatars and names with female characteristics to improve customer satisfaction.

Our study also provides strategies on how companies can effectively retain angry customers following service failures. While it is not possible to prevent customer anger resulting from service failures, companies can realize timely service recovery and the selection of the appropriate recovery method to reduce negative downstream effects. We subdivide two types of symbolic recovery and examine how chatbot gender can be matched with the recovery method to effectively increase customer satisfaction, and leverage the positive influence of AI in the interactive marketing field. For example, female chatbots can say thank you to angry customers, while male chatbots can say sorry to offer sincere apologies.

5.3 Limitations and directions for future research
A limitation of this paper is that it focuses on Chinese-speaking participants. However, specific cultural values based on socialization and interpersonal relationships may also influence an individual’s perception of gender differences (Steinmetz et al., 2014). For example, in most countries, the skirt is seen as a symbol of femininity, but in the UK the kilt was originally designed exclusively for men. Furthermore, from a linguistic perspective, there may be a distinguish in the understanding of apology and appreciation in different languages (Chamani, 2014). Therefore, future research could focus on a more extensive range of cultural contexts. Another limitation of this study is
that it selects chatbots that are all serving the e-commerce industry and the restaurant industry. Although some chatbots are also used in medical, legal and financial applications, these are not AI medical, AI legal and AI financial (Xu et al., 2021), but still tend to be more of an information inquiry and reception service. Therefore, the influence of chatbot gender interactions with symbolic service recovery in these scenarios needs to be further investigated.

ORCID iDs
Shichang Liang https://orcid.org/0000-0001-6990-8816
Bin Lan https://orcid.org/0000-0001-9249-9803
Li Li https://orcid.org/0009-0000-9731-9875

References


Chatbot gender on angry customers


Yin, J. and Qiu, X. (2021), “Influence mechanism research on the AI marketing technology of online shopping platform on consumer/s' Purchase Intention. 6th Annual International Conference on Social Science and Contemporary Humanity Development (SSCHD 2020).


## Appendix 1. Chat Conversation with a Chatbot (Study 1)

### Female Chatbot - Ambiguous

<table>
<thead>
<tr>
<th>Nimilia</th>
<th>Hello! I’m your customer service assistant, Nimilia. I’m happy to serve you! Please tell me why you are contacting us today.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nimilia</td>
<td>It's unfortunate to hear about the shipping error. This sunscreen clothe is now out of stock, once the stocks are replenished, we will be the first to ship it to you.</td>
</tr>
<tr>
<td>Nimilia</td>
<td>Once the exact time is set, we will notify you in advance. Is that all we can help you with today?</td>
</tr>
</tbody>
</table>

Your reply to Nimilia:

**SHIPPING ERROR**

Next

**DELIVERY TIME**

Next

END
<table>
<thead>
<tr>
<th>Male Chatbot - Resolved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hello! I’m your customer service assistant, Nimeld. I’m happy to serve you! Please tell me why you are contacting us today.</td>
</tr>
<tr>
<td>Your reply to Nimeld:</td>
</tr>
<tr>
<td>SHIPPING ERROR</td>
</tr>
<tr>
<td>Next</td>
</tr>
<tr>
<td>Nimeld</td>
</tr>
<tr>
<td>It's unfortunate to hear about the shipping error. This sunscreen clothe is now out of stock, once the stocks are replenished, we will be the first to ship it to you.</td>
</tr>
<tr>
<td>Your reply to Nimeld:</td>
</tr>
<tr>
<td>DELIVERY TIME</td>
</tr>
<tr>
<td>Next</td>
</tr>
<tr>
<td>Nimeld</td>
</tr>
<tr>
<td>The factory can ship the products after 3 days, and the items will reach you on time after 1-2 days to ensure that you can use them in time. Is that all we can help you with today?</td>
</tr>
<tr>
<td>Your reply to Nimeld:</td>
</tr>
<tr>
<td>END</td>
</tr>
</tbody>
</table>
Appendix 2. Study 2 Pretest

Participants and design. Before conducting the formal study, we recruited 261 (159 women; $M_{\text{age}} = 21.21$ years, SD = 6.23) participants for our pretest in exchange for a small amount of money. The length of waiting time can influence an individual’s patience and mood, and the longer the waiting time, the more likely the negative emotions (Shaddy and Lee, 2019). Drawing on the classification of waiting time of Shaddy and Lee (2019), we used a one-factor, three-level (waiting time: 2 min. vs. 10 min. vs. 30 min.) between-participants design to verify the differences in the participants’ anger emotion in response to the length of waiting time.

Procedure. During the study, we arranged for the participants to sit by themselves, without interference from one another. Before giving out the reading material, we played soothing music to calm the participants before conducting the study. Next, we showed the participants a picture of a milk tea store and instructed them to imagine that they are waiting in line to buy a product. We asked the participants to read the text description (“You ordered from the milk tea store in the picture and are now waiting in line for your order. You have been waiting for 2 min/10 min/30 min because the staff made a mistake in your order”). After reading the text, we instructed the participants to answer the anger-related questions (“I am now very angry at this milk tea store,” “I am now very angry at the staff of this store,” and “I will never come to this store again”; 1 = “strongly disagree” and 7 = “strongly agree”; $a = 0.91$). To exclude interference from the milk tea brand on the study results, we also asked the participants whether they were familiar with the milk tea brand and whether they liked the milk tea displayed in the picture. Finally, we recorded the participants’ age, gender and frequency of online shopping.

Customer anger. The one-way ANOVA with the wait time as the independent variable and customer anger as the dependent variable showed a significant difference in anger by waiting time ($F[2, 258] = 31.481, p < 0.001, \eta^2 = 0.196$). Specifically, anger was lowest at 2 min of waiting time and highest at 30 min of waiting time ($M_{2\text{min}} = 3.35$, SD = 1.49; $M_{10\text{min}} = 4.06$, SD = 1.23; $M_{30\text{min}} = 5.01$, SD = 1.42). Therefore, we used the waiting time of 2 min and 30 min as manipulation of emotions in our formal study.

Alternative explanations. Using the participants’ familiarity with the milk tea brand as a mediator (Hayes, 2015; Model 4, 5,000 bootstrap resamples), we did not find that familiarity mediated the effect of waiting time on feelings of anger ($\beta = 0.0196$, SE = 0.0336, 95% CI = [−0.0416, 0.0919]). Repeating the analysis with the degree of milk tea preference, we found that it had no significant mediating effect ($\beta = 0.0424$, SE = 0.0365, 95% CI = [−0.0251, 0.1187]). Therefore, the results excluded alternative explanations for the participants’ familiarity with the milk tea brand and degree of milk tea preference.
Appendix 3. Chat Conversation with a Chatbot (Study 2)

Female Chatbot - Appreciation

Hello! I’m your order assistant, Nimilia, and I’m happy to serve you! Please tell me why you are contacting us today.

Nimilia

Your reply to Nimilia:

DELIVER FOOD

Thank you for your patience, I really appreciate it! I will immediately urge the kitchen to deliver your food. Is that all we can do for you today?

Nimilia

Your reply to Nimilia:

END
Female Chatbot - Apology

Hello! I’m your order assistant, Nimilia, and I’m happy to serve you! Please tell me why you are contacting us today.

Nimilia

Your reply to Nimilia:

DELIVER FOOD

Next

I’m sorry for keeping you waiting, I sincerely apologize! I will immediately urge the kitchen to deliver your food. Is that all we can do for you today?

Nimilia

Your reply to Nimilia:
Appendix 4. Field Experiments in a Café (Study 2)

The ordering interface of the café

Customers viewing experimental materials
Appendix 5. Study β

Study β verified H1, stating that female chatbots are more likely to increase the satisfaction of angry customers than male chatbots. Study β manipulated the chatbot gender by playing an audio clip for the participants. Moreover, technical willingness is a concept that reflects personality traits, which refers mainly to an individual’s tendency to use new technologies to accomplish a certain task (Sadik-Rozsványai and Bertrandias, 2019). To eliminate any potential bias toward new technologies from the participants themselves, we also measured their technical willingness (Appendix 5).

Methods: Participants and design. We recruited 321 students (219 women; $M_{\text{age}} = 20.82$ years, $SD = 4.61$) at a university in Southern China to participate in this study for course credit. For Study 1, we adopted a 3 (chatbot gender: male vs. genderless vs. female) × 2 (emotional scenario: control vs. anger) between-subjects design.

Stimulus materials. Drawing on previous descriptions of control and anger scenarios (Crolic et al., 2022) and the classification of chatbot gender (Chen et al., 2022; Pogacar et al., 2021), we instructed the participants in the control scenario to read the following situation:

You buy a luggage online for RMB 3,000 (from an online shopping platform that uses a chatbot to provide customer service). When it arrives, you check the luggage and see that it does not have any problems. Finally, you ask customer service about precautions in using the luggage then confirm the delivery.

The participants in the anger scenario read the following situation: “You buy a luggage online. After it arrives, you find that the luggage is badly damaged and cannot be used. Finally, you contact customer service to deal with the after-sales issue.” (see Appendix 6).

In the male chatbot group, the participants heard a distinctly male voice that said, “Hello! I’m your customer service assistant, Nimile. I’m happy to serve you!” In the female chatbot group, the participants heard a distinctly female voice that said, “Hello! I’m your customer service assistant, Nimila. I’m happy to serve you!” In the genderless chatbot group, the participants heard a distinctly robotic voice that said, “Hello! I’m your customer service assistant, and I’m happy to serve you!” (see Appendix 7).

Procedure and Measurements. We randomly assigned the participants to read one of the control or anger scenarios. Then, we instructed the participants in the control or anger scenario to engage in a chat conversation with a chatbot.

After browsing the stimuli, the participants reported their feelings about the gender of the chatbot (“I think the service assistant bot has female characteristics”; 1 = “strongly disagree” and 7 = “strongly agree”; $\alpha = 0.89$) and level of their anger (“Now, I feel very angry/bad/frustrated”; 1 = “strongly disagree” and 7 = “strongly agree”; $\alpha = 0.95$). We then asked the participants to report their satisfaction with the chatbot’s handling of the query (Agarwal and Prasad, 1998; Spreng et al., 1996; Venkatesh and Davis, 2000). Finally, we also asked the participants to answer questions related to their technical willingness (Courtney, 2008), and recorded the participants’ age, gender, and frequency of online shopping.

Results: Manipulation checks. We conducted one-way ANOVA on perceived chatbot gender and extent of anger, and the results showed that in the emotional scenario, the anger group was significantly angrier than the control group ($M_{\text{anger}} = 4.44$, $SD = 1.43$; $M_{\text{control}} = 2.55$, $SD = 1.69$; $F[1, 319] = 116.897$, $p < 0.001$, $\eta^2 = 0.268$). We observed significant differences in the participants’ perception of chatbot gender ($M_{\text{male}} = 3.55$, $SD = 1.14$; $M_{\text{genderless}} = 4.25$, $SD = 0.74$; $M_{\text{female}} = 4.58$, $SD = 1.08$; $F[2, 318] = 28.791$, $p < 0.001$, $\eta^2 = 0.153$), verified the validity of the gender manipulation (see Figure A1).

Customer satisfaction. We conducted ANOVA, with customer satisfaction as the dependent variable. The results showed that emotion had a significant effect on customer satisfaction ($F[1, 315] = 29.874$, $p < 0.001$, $\eta^2 = 0.086$), the anger group was significantly less satisfied than the control group ($M_{\text{anger}} = 3.95$, $SD = 1.35$; $M_{\text{control}} = 4.63$, $SD = 1.00$), and the interaction of chatbot gender and emotion had a significant effect on customer satisfaction ($F[2, 315] = 14.691$, $p < 0.001$, $\eta^2 = 0.089$). The results of the simple-effects analysis showed that for the participants in the anger group, the customer satisfaction of the female chatbot group ($M_{\text{female}} = 4.68$, $SD = 1.27$; $F[2, 315] = 19.668$, $p < 0.001$, $\eta^2 = 0.111$) was significantly higher than that of the genderless chatbot group ($M_{\text{genderless}} = 3.82$, $SD = 1.23$) and that of the male chatbot group ($M_{\text{male}} = 3.33$, $SD = 1.20$). In addition, for those in the control group,
we observed no significant effect of chatbot gender on customer satisfaction ($M_{\text{male}} = 4.85, \text{SD} = 0.83; M_{\text{genderless}} = 4.52, \text{SD} = 0.83; M_{\text{female}} = 4.52, \text{SD} = 1.26; F[2, 315] = 1.543, p > 0.05, \eta^2 = 0.009$); thus, H1 was verified (see Figure A2).

Alternative explanations. To exclude alternative explanations for technical willingness, first, we found that chatbot gender did not affect individual technical willingness ($M_{\text{male}} = 4.54, \text{SD} = 0.85; M_{\text{genderless}} = 4.52, \text{SD} = 0.87; M_{\text{female}} = 4.65, \text{SD} = 1.04; F[2, 318] = 0.631, p > 0.05, \eta^2 = 0.004$; see Table A1). Second, we used technical willingness as a mediating variable (Hayes, 2015; Model 7, 5,000 bootstrap resamples). We did not observe the effect of technical willingness mediating the effect of chatbot gender on customer satisfaction in the anger group ($\beta = -0.1114, \text{SE} = 0.0712, 95\% \text{CI} = [-0.2580, 0.0199]$). Therefore, the above results excluded alternative explanations for technical
willingness. In addition, the ANCOVA with the participants' gender as a covariate showed that the participants' gender had no significant effect on customer satisfaction ($F[1, 314] = 1.915, p > 0.05, \eta^2 = 0.006$). We repeated the above analysis with the participants' frequency of online shopping as the covariate, and the results revealed that the frequency of online shopping had no significant effect on customer satisfaction ($F[1, 314] = 0.628, p > 0.05, \eta^2 = 0.002$) (see Table A2).

### Table A1. The effect of chatbot gender on technical willingness

<table>
<thead>
<tr>
<th>Variable</th>
<th>FC</th>
<th>M (SD)</th>
<th>GC</th>
<th>M (SD)</th>
<th>MC</th>
<th>M (SD)</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical willingness</td>
<td>Optimism</td>
<td>4.88 (1.19)</td>
<td>4.86 (1.09)</td>
<td>4.75 (1.04)</td>
<td>$F[2, 318] = 0.436$</td>
<td>0.647</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Creativity</td>
<td>4.65 (1.27)</td>
<td>4.62 (1.11)</td>
<td>4.58 (1.02)</td>
<td>$F[2, 318] = 0.095$</td>
<td>0.910</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discomfort</td>
<td>4.36 (1.32)</td>
<td>4.02 (1.19)</td>
<td>4.27 (1.19)</td>
<td>$F[2, 318] = 2.211$</td>
<td>0.111</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lack of security</td>
<td>4.74 (1.24)</td>
<td>4.61 (1.07)</td>
<td>4.57 (1.05)</td>
<td>$F[2, 318] = 0.665$</td>
<td>0.547</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>4.65 (1.04)</td>
<td>4.52 (0.87)</td>
<td>4.54 (0.85)</td>
<td>$F[2, 318] = 0.631$</td>
<td>0.533</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** FC = female chatbot group; GC = genderless chatbot group; MC = male chatbot group; "p > 0.05" = nonsignificant

### Table A2. Statistics descriptive results of Study β

<table>
<thead>
<tr>
<th>Study β</th>
<th>M (SD)</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer satisfaction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chatbot Gender × Emotion</td>
<td>Anger (N = 161)</td>
<td>4.68 (1.27)</td>
<td>3.33 (1.20)</td>
</tr>
<tr>
<td>Control (N = 160)</td>
<td>4.52 (1.26)</td>
<td>4.85 (0.83)</td>
<td>4.52 (0.83)</td>
</tr>
</tbody>
</table>

**Alternative explanation**

<table>
<thead>
<tr>
<th>Participants' gender</th>
<th>$F[1, 314] = 1.915, p &gt; 0.05, \eta^2 = 0.006$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of online shopping</td>
<td>$F[1, 314] = 0.628, p &gt; 0.05, \eta^2 = 0.002$</td>
</tr>
<tr>
<td>Technical willingness</td>
<td>$\beta = -0.1114, \text{SE} = 0.0712, 95% \text{ CI} = [-0.2580, 0.0199]$</td>
</tr>
</tbody>
</table>

**Notes:** FC = female chatbot group; GC = genderless chatbot group; MC = male chatbot group; "p > 0.05" = non-significant; Statistics in parentheses are standard deviations

**Source:** Authors' own work
<table>
<thead>
<tr>
<th>Female Chatbot - Anger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your Reply:</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

It sounds like you have a question. Please select the product you need help with.

<table>
<thead>
<tr>
<th>Your Reply:</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUITCASE</td>
</tr>
</tbody>
</table>

Unfortunate to hear that there is a problem with your suitcase! Is there anything we can do for you?

<table>
<thead>
<tr>
<th>Your Reply:</th>
</tr>
</thead>
</table>
Female Chatbot - Control

Your Reply:

It sounds like you have a question. Please select the product you need help with.

Your Reply:

SUITCASE

Next

Nice to hear that there is no problem with your suitcase! Is there anything we can do for you?

Your Reply:

END
Appendix 7. Manipulation of Gender (Study β)

Female chatbot group

The above Chinese audio is expressed in English as “Hello! I'm your customer service assistant, Nimilia. I'm happy to serve you!”

Male chatbot group

The above Chinese audio is expressed in English as “Hello! I'm your customer service assistant, Nimeld. I'm happy to serve you!”

Genderless chatbot group

The above Chinese audio is expressed in English as “Hello! I'm your customer service assistant, and I'm happy to serve you!”
Appendix 8. Study 3

In Study 3, we further verified that symbolic recovery (apology vs. appreciation) moderated the effect of chatbot gender on the satisfaction of angry customers in the context of service failure. As previous studies have shown, service failures in booking door-to-door services can also trigger customer anger, which ultimately influences satisfaction (Avermann and Schlüter, 2019). Therefore, in this study, we manipulated customer anger by simulating service failures within the context of door-to-door services. Drawing on You et al. (2020) the experimental manipulation of symbolic service recovery, we verify our conjecture by using the stimulus of the chatbot gender of Study 1 (Appendix 8).

Methods: Participants and design. Study 3 employed a 2 (symbolic recovery: apology vs. appreciation) × 2 (chatbot: female vs. male) between-subjects design. 482 MTurk workers (287 women; \(M_{\text{age}} = 31.91 \text{ years}, SD = 9.40\)) completed the study for monetary compensation and were randomly assigned to one of the four conditions.

Procedure. At the beginning of the experiment, we played a piece of soothing music to calm the participants down before proceeding with the study. To induce feelings of anger, participants were asked to read a text description describing a service failure scenario.

Due to a broken water pipe at their door-to-door, the participant had booked a door-to-door service at 10 a.m. through an online platform, with the chatbot customer service assuring that the staff will arrive on time. However, by 10 p.m., no staff have arrived.

Then, participants were asked to indicate their emotional response by selecting either the “angry” or “not angry” option from a provided radio button.

Next, participants were presented with a picture of a chatbot (see Figure 1) and a chat message (see Appendix 9) in which a customer contacts the door-to-door service platform to deal with a problem. In the appreciation condition, the participant read the following chat message, “Thank you for your patience, I really appreciate it!” In the apology condition, the participant read the following chat message, “Sorry to keep you waiting, I sincerely apologize!” As in Study 3, after reading the material, we assessed participants’ perception of the chatbot’s gender and their satisfaction with the chatbot’s problem-solving approach (α = 0.96). In addition, to examine the effectiveness of manipulation, we also measured symbolic recovery (“I think the chat message is an apology” and “I feel that the chatbot apologized to me”; 1 = “Strongly disagree” and 7 = “Strongly agree”; α = 0.94).

Finally, to verify that gender stereotypes do exist in the chatbot, participants were asked to indicate their perception of gender stereotypes by selecting a radio button labeled “upright and unyielding” or “tender and vulnerable”. Meanwhile, we also collected additional demographic information from participants, such as gender, age, and education level. At the end of the study, we expressed our gratitude to the participants for their participation.

Results: Manipulation checks. Regarding the priming item of anger, out of the 482 participants, 438 participants chose “anger” while 44 participants chose “not angry”. To prime the participants’ anger, the 44 participants who chose “not angry” were excluded from the data analysis. The results of an independent samples t-test revealed that participants in the apology condition reported significantly higher perceptions of the chatbot expressing apology than participants in the appreciation condition (\(M_{\text{apology}} = 1.86, SD = 1.05; M_{\text{appraisal}} = 5.27, SD = 1.73, t(436) = 24.913, p < 0.001\)). Additionally, a one-way ANOVA was conducted to examine the perceived chatbot gender. The results revealed significant differences in the perception of chatbot gender between participants in the male and female chatbot conditions (\(M_{\text{male}} = 2.94, SD = 1.70; M_{\text{female}} = 6.12, SD = 0.69, F(1, 436) = 651.998, p < 0.001, \eta^2 = 0.599\)).

Customer satisfaction. A one-way ANOVA on customer satisfaction showed that chatbot gender has a significant effect on customer satisfaction (\(M_{\text{male}} = 4.79, SD = 1.68; M_{\text{female}} = 5.15, SD = 1.44, F(1,436) = 5.585, p < 0.05, \eta^2 = 0.013\)). H1 was confirmed again. A two-way ANOVA was conducted with customer satisfaction as the dependent variable, revealing a significant interaction effect between chatbot gender and symbolic recovery on customer satisfaction (\(F(1,434) = 28.039, p < 0.001, \eta^2 = 0.061\)). Specifically, participants were more satisfied with female chatbots than male chatbots when the symbolic recovery was appreciation (\(M_{\text{male}} = 4.33, SD = 1.81; M_{\text{female}} = 5.46, SD = 1.22, F(1,434) = 29.938, p < 0.001, \eta^2 = 0.065\)). Participants were more satisfied with male chatbots than female chatbots when the symbolic recovery was an apology (\(M_{\text{male}} = 5.26, SD = 1.41; M_{\text{female}} = 4.86, SD = 1.59, F(1,434) = 4.036, p < 0.05, \eta^2 = 0.009\)). Therefore, H3 was validated again (see Table A3).
Perception of gender stereotypes. We used Chi-square analysis to test whether gender stereotypes existed in chatbots, and the participant’s perception of gender stereotypes were used as the dependent variable (−1 = “tender and vulnerable”, 1 = “upright and unyielding”). The results showed that participants did have gender stereotypes about the chatbots’ gender ($\chi^2 [1] = 220.836, p < 0.001$). Specifically, participants in the female chatbot condition (98.2 percent) were more likely to choose “tender and vulnerable” than those in the male chatbot condition (29.9 percent). Participants in the male chatbot condition (70.1 percent) were more likely to choose “upright and unyielding” than participants in the female chatbot condition (1.8 percent). The results verified that gender stereotypes exist in chatbots.

Discussion: In Study 3, we conducted additional scenarios to ensure the consistency of our findings. The results revealed that the effect of chatbot gender on customer satisfaction, in the context of customer anger caused by various service failures, was moderated by symbolic recovery (apology vs. appreciation). We further extended the application of the experimental findings to provide a practical strategy for companies in handling service failures. Moreover, unlike Studies 1, 2 which used specific products to induce result-based service failure conditions, in Study 3 we used online booking services, i.e. process-based service failure to evoke anger. This approach aimed to enhance the internal validity of the study.

<table>
<thead>
<tr>
<th>Study 3</th>
<th></th>
<th></th>
<th>df</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer satisfaction</td>
<td></td>
<td></td>
<td>df</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>FC</td>
<td>MC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chatbot Gender × Angry</td>
<td>5.15 (1.44)</td>
<td>4.79 (1.68)</td>
<td>$F[1,436] = 5.585$</td>
<td>0.019</td>
</tr>
<tr>
<td>Chatbot Gender × Angry × Symbolic recovery</td>
<td>5.46 (1.22)</td>
<td>4.33 (1.81)</td>
<td>$F[1,434] = 29.938$</td>
<td>0.000</td>
</tr>
<tr>
<td>Appreciation ($N = 218$)</td>
<td>4.85 (1.59)</td>
<td>5.26 (1.41)</td>
<td>$F[1,434] = 4.036$</td>
<td>0.045</td>
</tr>
<tr>
<td>Apology ($N = 220$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A3.
Statistics descriptive results of Study 3

Notes: FC = female chatbot group; GC = genderless chatbot group; MC = male chatbot group; “p > 0.05” = non-significant; Statistics in parentheses are standard deviations

Source: Authors’ own work
Appendix 9. Chat Conversation with a Chatbot (Study 3)

Female Chatbot - Appreciation

Hello! I’m your service assistant, Nimilia, and I’m happy to serve you! Please tell me why you are contacting us today.

Your reply to Nimilia:

Thank you for your patience, I really appreciate it! I will immediately urge the staff to provide you with home service. Is that all we can do for you today?

Your reply to Nimilia:
Hello! I’m your service assistant, Nimeld, and I’m happy to serve you! Please tell me why you are contacting us today.

I’m sorry for keeping you waiting, I sincerely apologize! I will immediately urge the staff to provide you with home service. Is that all we can do for you today?

Your reply to Nimeld:
About the authors
Shichang Liang Conceptualization, methodology, resources, writing-review and editing.
Rulan Li Visualization, supervision and data curation.
Bin Lan Software, Validation, Writing-Original draft preparation, Investigation. Bin Lan can be contacted at: 2102302014@st.gxu.edu.cn
Yuxuan Chu Project administration, funding acquisition.
Min Zhang Resources and data curation.
Li Li Project administration, funding acquisition.

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com