

Unraveling the Curvilinear Impact of AI Agent Anthropomorphism on Service Failure Tolerance

Ruoran Liu ¹
Jin-Xing Hao ^{2,*}
Yan Yu ³
Wei Shan ⁴

*Corresponding author

¹ PhD student, Beihang University, Beijing, China, liuruoran@buaa.edu.cn

² Professor, Beihang University, Beijing, China, hao@buaa.edu.cn

³ Associate Professor, Renmin University of China, Beijing, China, yanyu@ruc.edu.cn

⁴ Professor, Beihang University, Beijing, China, shanwei@buaa.edu.cn

ABSTRACT

Anthropomorphic AI agents are increasingly prevalent, yet their impact on customer responses to service failures remains contentious. Drawing on cognitive appraisal theory, this study explores the curvilinear effects of AI agent anthropomorphism on customer service failure tolerance. Through two video-based experiments and one text-based experiment involving a total of 526 subjects, we aim to investigate whether an inverted U-shaped relationship exists between anthropomorphism and failure tolerance and if likability serves as a mediating factor in this relationship. We further explore the potential moderating role of failure severity, hypothesizing that high severity may attenuate the effect of anthropomorphism on likability while low severity could strengthen the indirect relationship between anthropomorphism and failure tolerance via likability. These insights are expected to enrich cognitive appraisal theory and the uncanny valley effect by uncovering potential internal emotional mechanisms, offering valuable implications for designing and implementing AI agents.

Keywords: AI agent, anthropomorphism, cognitive appraisal, likability, failure tolerance, emotional mechanism.

INTRODUCTION

AI agents are increasingly influential across sectors for their autonomous capabilities, with 55% of businesses incorporating AI technology in at least one unit by 2023, up significantly from prior years (Maslej *et al.*, 2024). However, AI adoption in the service industry is cautious due to the complexity of human interactions and the risk of service failures (Choi *et al.*, 2021), which can damage both customer experiences and company reputations (Cameron *et al.*, 2021). Addressing service failure in this context is, therefore, urgently needed to reduce the negative impact by enhancing customer tolerance.

Prior studies have demonstrated that anthropomorphic strategies have the potential to mitigate AI agent service failure, but their role in customer tolerance is still controversial (Xie & Lei, 2022; Yang *et al.*, 2024). Some studies have found that customers show a higher tolerance for AI agents with high anthropomorphism (e.g., Arikan *et al.*, 2023); others have suggested that AI agents with low anthropomorphism are more likely to gain customer tolerance (e.g., Cheng, 2023; Cui & Zhong, 2023; Zhang *et al.*, 2023). Therefore, this study aims to delve further into such intricate relationships and attempts to investigate how and why AI agent anthropomorphism affects customer service failure tolerance.

Particularly, this study will examine the curvilinear effect of AI agent anthropomorphism on service failure tolerance through the internal emotion mechanism of likability under different failure severity conditions. Based on cognitive appraisal theory, when customers face service failure, they assess stress and generate emotional responses that directly affect tolerance decisions (Folkman *et al.*, 1986; Smith & Ellsworth, 1985). In this process, personal and situational factors jointly influence stress assessment (Lazarus & Folkman, 1984). As a personal factor, likability, defined as the ease with which one is liked and possesses pleasant or appealing qualities (Sandoval *et al.*, 2021), is a crucial positive emotional representation in customer-AI collaboration (Salem *et al.*, 2013). As a situational factor, the anthropomorphism of the AI agent, defined as the process of attributing human-like traits, emotions, or intentions to nonhuman entities (Duffy, 2003), influences the tolerance decision through emotion changes (Smith & Ellsworth, 1985). Failure severity, as another indispensable situational factor reflecting the customer's perception of the intensity of the service problem (Hess, 2008; Wang *et al.*, 2011), affects both the customer's emotional state and tolerance decision (Smith *et al.*, 1999; Tsarenko & Roosrani Tojib, 2011).

To explore these dynamics, this study uses three experimental designs to simulate AI agent interactions in service scenarios. Findings aim to contribute to AI agent service literature by clarifying the inverted U-shaped relationship between

anthropomorphism and tolerance, highlighting likability's role as a mediator, and identifying failure severity as a boundary condition. These insights offer a foundation for optimizing AI design and implementation in service contexts.

LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

AI Agents

AI agents utilize advanced technologies such as natural language processing, machine learning, and robotics to achieve autonomy, intelligence, adaptability, and social interaction. In the service sector, they manifest as service robots, chatbots, and virtual assistants, performing vital functions like reception, concierge services, and item delivery. These agents not only reduce labor costs but also facilitate value co-creation through customer-AI interactions (Han et al., 2023). Through experiential and emotional capabilities, AI agents can enhance customer service experiences, ultimately fostering stronger connections between service providers and customers (Huang & Rust, 2018).

The growing adoption of AI agents has drawn significant interest from both scholars and industry professionals. Their appeal lies in several factors, including attracting customers and increasing traffic (Zhang et al., 2022); enhancing service experiences and customer satisfaction (Lajante et al., 2023); effectively addressing diverse customer needs through personalized services (Xu et al., 2023); and contributing to positive word-of-mouth and brand reputation, thus promising a brighter future for the service industry (Lu et al., 2021).

AI Agent Service Failure and Failure Tolerance in Cognitive Appraisal

Service failures involving AI agents are common due to their limited adaptability. When failures occur, AI agents often struggle to recover through pre-programmed procedures, which can lead to customer dissatisfaction. For instance, Connie at the Hilton Hotel faced navigation issues (Koo et al., 2021), and Botlr at the Aloft Hotel encountered delivery problems (Çakar & Aykol, 2021). Therefore, in situations where service failures are inevitable, it is crucial to understand customer perceptions of these failures to mitigate negative impacts.

Cognitive appraisal theory provides a framework for analyzing emotional responses to service failures. It posits that customers perceive service failures as threats to their well-being, eliciting negative emotions (Lazarus & Folkman, 1984). To cope with these emotions, customers may use various strategies, with failure tolerance being an effective strategy (Lv et al., 2021). Failure tolerance indicates the extent to which customers can endure substandard service before considering switching providers (Cui & Zhong, 2023; Zhang et al., 2023). This concept reflects customer evaluations of service quality and overall satisfaction, influencing their behavioral intentions (Zeithaml et al., 1993). By enhancing failure tolerance, customers can manage service failures more effectively, reducing emotional distress.

The Curvilinear Relationship Between Anthropomorphism and Failure Tolerance

Research shows that anthropomorphism can enhance customer acceptance of AI agents by fostering cognitive, emotional, and associative connections. Studies suggest that customers empathize more with human-like AI agents, which elicits positive emotions and fulfills the need for social connection (Choi et al., 2021; Mende et al., 2019; Yang et al., 2022). Cognitive appraisal theory suggests that such empathy leads to higher failure tolerance when the agent's limitations are perceived as understandable. However, excessive anthropomorphism can trigger the uncanny valley effect, leading to discomfort and reduced tolerance when AI agents seem nearly human yet flawed (Mori et al., 2012; Yang et al., 2024). This study posits that anthropomorphism's effect on failure tolerance follows an inverted U-shape, with modest anthropomorphism yielding the highest tolerance.

H1: AI agent anthropomorphism has an inverted U-shaped relationship with failure tolerance.

The Curvilinear Relationship Between Anthropomorphism and Likability

According to cognitive appraisal theory, customers assess failure events based on perceived impacts on personal well-being, which drives emotional responses (Lazarus & Folkman, 1984; Smith & Ellsworth, 1985). Likability reflects the emotional connection customers feel toward AI agents (McGuire, 1985). This study posits that the anthropomorphism-likability relationship is also curvilinear. Modest anthropomorphism satisfies customers' identity needs, fostering personal well-being and higher likability. However, when anthropomorphism becomes excessive, it may threaten customers' sense of identity, reducing personal well-being and lowering likability (Mende et al., 2019; Stein & Ohler, 2017). Thus, we propose:

H2: AI agent anthropomorphism has an inverted U-shaped relationship with likability.

The Relationship Between Likability and Failure Tolerance

Likability is a key aspect of human psychology that generates positive emotional and behavioral responses (Tellefsen & Thomas, 2005; Zeithaml et al., 1993). In the context of service failure, likability can act as a cognitive buffer, helping customers manage negative emotions and adopt constructive coping strategies, which include failure tolerance. Customers with high failure tolerance are more likely to tolerate service failures, thereby preserving satisfaction and brand image (Lv et al., 2021). Therefore, we argue that likability enhances service failure tolerance.

H3: Likability has a positive influence on failure tolerance.

The Mediating Role of Likability

According to cognitive appraisal theory, anthropomorphism affects failure tolerance through a customer's emotional response. Low levels of anthropomorphism yield minimal likability, failing to mitigate negative emotions and resulting in low failure tolerance. As anthropomorphism increases, perceived likability also rises, enhancing failure tolerance. However, when anthropomorphism surpasses a threshold, customers may experience the uncanny valley effect (Mori et al., 2012), where overly human-like AI agents induce discomfort and reduce likability. This diminished likability can lead to lower failure tolerance because negative emotions are not effectively managed. Combining insights from H1, H2, and H3, we propose:

H4: Likability mediates the inverted U-shaped relationship between AI agent anthropomorphism and failure tolerance.

The Moderating Role of Failure Severity

Failure severity, defined as the perceived intensity of a service problem, is crucial in understanding service failure (Wang et al., 2011; Webster & Sundaram, 1998). Higher failure severity is associated with greater perceived loss (Weun et al., 2004). Research shows that failure severity moderates customer perceptions of positive emotions. For example, Yang et al. (2022) find that the positive effects of an agent's humor on service failure evaluations are more pronounced in low-severity situations, while they become negative under high-severity conditions. Similarly, Lv et al. (2021) suggest that customers view AI assistance as more cuteness in low-severity failures, leading to greater tolerance, whereas the positive influence of AI cuteness diminishes in high-severity contexts.

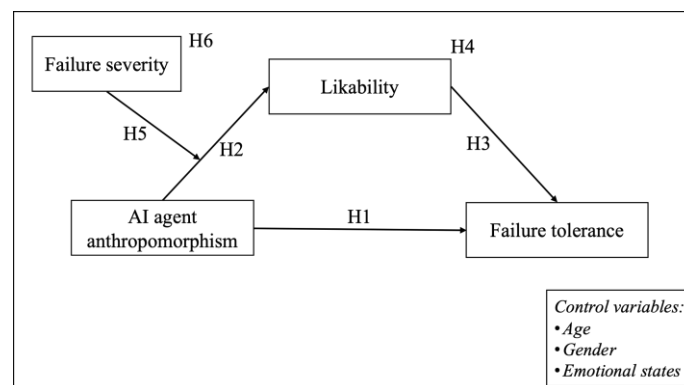
Cognitive appraisal theory posits that varying failure severity influences customers' evaluations of potential losses, inducing different emotional responses (Wang et al., 2011). In low-severity contexts, customers perceive failures as minor inconveniences, enhancing their perception of the likability of anthropomorphic AI agents. Conversely, high-severity failures evoke frustration and negative emotions, overshadowing the positive effects of anthropomorphism. Therefore, we propose:

H5: Failure severity moderates the curvilinear relationship between AI agent anthropomorphism and likability, such that the curvilinear effect is attenuated (vs. amplified) when customers encounter high (vs. low) failure severity.

Furthermore, failure severity is closely associated with failure tolerance. When a customer experiences a low-severity service failure, the likability elicited by AI agent anthropomorphism positively influences failure tolerance by reducing the perceived potential loss, thereby enhancing the customer's failure tolerance. In contrast, during a high-severity service failure, the negative emotions stemming from the customer's assessment of potential loss significantly diminish failure tolerance, leading the customer to overlook the likability induced by the anthropomorphism associated with the service failure. Thus, we argue that failure severity is the moderator in the indirect relationship between anthropomorphism and failure tolerance via likability:

H6: Failure severity moderates the indirect relationship between AI agent anthropomorphism and failure tolerance through likability, such that the relationship becomes stronger (vs. weaker) when failure severity is low (vs. high).

The conceptual framework of the research is illustrated in Figure 1.



Source: This study.

Figure 1: Conceptual framework.

METHODOLOGY

To test these hypotheses, we designed three laboratory experiments. In each experiment, subjects were asked to complete tasks under the experimental guidance. In the first study, we adopted a video-based scenario experiment to manipulate and test three different levels of anthropomorphism of AI agents. In the second study, we used a text-based scenario experiment to manipulate and verify two levels of service failure. These two studies provided the basis for subsequent main experiments with stimulus materials. In the third study, we conducted a video-vignettes-based experiment to test Hypotheses 1-6.

Study 1

This study investigates the effectiveness of anthropomorphism in AI agents within customer service scenarios. Participants were recruited from the Credamo platform and randomly assigned to one of three conditions with varying levels of anthropomorphism. Participants accessed the experimental system via their computers and watched a fullscreen video depicting an AI agent interacting with a customer, focusing on anthropomorphic traits. The video ended after the AI agent acknowledged the customer's needs, minimizing the impact of service outcomes. Afterward, participants rated the AI agent's anthropomorphism on a slider scale from 1 to 100 and provided demographic information, including age and gender. Participants received a cash reward upon completion.

Study 2

This study aims to validate the manipulation of two levels of failure severity in service scenarios based on a hotel check-in scenario adapted from Smith et al. (1999). Textual scripts for the scenarios were created and then reviewed by two experts for fluency and authenticity, ensuring all factors except severity were identical. Participants were recruited and randomly assigned to either a low-severity or high-severity scenario and read the text before proceeding to a questionnaire. To measure perceptions of service failure severity, we used a 5-point Likert scale adapted from Hess (2008), consisting of three items. Participants also provided demographic information, including age and gender, and received a cash reward upon completion.

Study 3

This study investigates the curvilinear relationship between anthropomorphism and failure tolerance through likability, examining the boundary role of failure severity using a 3 (anthropomorphism: low, modest, high) \times 2 (failure severity: low, high) factorial design with video vignettes. We produced 9 video vignettes of a check-in service including three introductions from anthropomorphic AI agents and six scenarios depicting service failures at varying severity levels.

Participants with recent travel experience were recruited from the Credamo platform and were randomly assigned to one of six conditions. They reported their emotional states before viewing a video vignette where the AI agent introduced itself. After the introduction, subjects evaluated the agent's anthropomorphism using a slider scale, followed by another video vignette in which the AI agent encountered a service failure. After viewing, participants completed questionnaires measuring likability, failure severity, failure tolerance, and scenario realism, alongside attention checks and demographic questions. A cash bonus was awarded for participation.

EXPECTED CONCLUSION AND CONTRIBUTION

This study enhances the understanding of AI agent anthropomorphism in service failures by demonstrating an inverted U-shaped relationship between anthropomorphism and failure tolerance, mediated by likability. It shows that modest anthropomorphism significantly increases both likability and failure tolerance, while high-severity failures weaken the direct effect of anthropomorphism on likability and the indirect effect of likability on failure tolerance.

Theoretical implications include a refinement of cognitive appraisal theory by offering new personal and situational factors and enriching the uncanny valley theory by expanding behavioral response from emotional response. Additionally, the research theoretically contributes to the internal mechanism (i.e., likability) and conditional boundary (i.e., failure severity) of the anthropomorphism-tolerance relationship. Practically, the findings suggest organizations employ AI agents with modest anthropomorphism and avoid AI agents with high anthropomorphism. Companies are also advised on service assignment strategies.

ACKNOWLEDGMENT

The work described in this paper was partially supported by grants from the National Natural Science Foundation of China (Project Nos: 72472008, 71974010).

REFERENCES

- Arikan, E., Altinigne, N., Kuzgun, E., & Okan, M. (2023). May robots be held responsible for service failure and recovery? The role of robot service provider agents' human-likeness. *Journal of Retailing and Consumer Services*, 70, 103175. <https://doi.org/10.1016/j.jretconser.2022.103175>
- Çakar, K., & Aykol, Ş. (2021). Understanding travellers' reactions to robotic services: a multiple case study approach of robotic hotels. *Journal of Hospitality and Tourism Technology*, 12(1), 155–174. <https://doi.org/10.1108/JHTT-01-2020-0015>
- Cameron, D., de Saille, S., Collins, E. C., Aitken, J. M., Cheung, H., Chua, A., Loh, E. J., & Law, J. (2021). The effect of social-cognitive recovery strategies on likability, capability and trust in social robots. *Computers in Human Behavior*, 114, 106561. <https://doi.org/10.1016/j.chb.2020.106561>
- Cheng, L. K. (2023). Effects of service robots' anthropomorphism on consumers' attribution toward and forgiveness of service failure. *Journal of Consumer Behaviour*, 22(1), 67–81. <https://doi.org/10.1002/cb.2112>
- Choi, S., Mattila, A. S., & Bolton, L. E. (2021). To Err Is Human(-oid): How do consumers react to robot service failure and recovery? *Journal of Service Research*, 24(3), 354–371. <https://doi.org/10.1177/1094670520978798>

- Cui, J., & Zhong, J. (2023). The effect of robot anthropomorphism on revisit intentions after service failure: A moderated serial mediation model. *Asia Pacific Journal of Marketing and Logistics*, 35(11), 2621–2644. <https://doi.org/10.1108/APJML-10-2022-0862>
- Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3–4), 177–190. [https://doi.org/10.1016/S0921-8890\(02\)00374-3](https://doi.org/10.1016/S0921-8890(02)00374-3)
- Folkman, S., Lazarus, R. S., Dunkel-Schetter, C., DeLongis, A., & Gruen, R. J. (1986). Dynamics of a stressful encounter: cognitive appraisal, coping, and encounter outcomes. *Journal of Personality and Social Psychology*, 50(5), 992. <https://doi.org/10.1037/0022-3514.50.5.992>
- Han, E., Yin, D., & Zhang, H. (2023). Bots with feelings: Should AI agents express positive emotion in customer service? *Information Systems Research*, 34(3), 1296–1311. <https://doi.org/10.1287/isre.2022.1179>
- Constituent Paths Are Nonlinear. *Multivariate Behavioral Research*, 45(4), 627–660. <https://doi.org/10.1080/00273171.2010.498290>
- Hess, R. L. (2008). The impact of firm reputation and failure severity on customers' responses to service failures. *Journal of Services Marketing*, 22(5), 385–398. <https://doi.org/10.1108/08876040810889157>
- Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155–172. <https://doi.org/10.1177/1094670517752459>
- Koo, B., Curtis, C., & Ryan, B. (2021). Examining the impact of artificial intelligence on hotel employees through job insecurity perspectives. *International Journal of Hospitality Management*, 95, 102763. <https://doi.org/10.1016/j.ijhm.2020.102763>
- Lajante, M., Tojib, D., & Ho, T. I. (2023). When interacting with a service robot is (not) satisfying: The role of customers' need for social sharing of emotion. *Computers in Human Behavior*, 146, 107792. <https://doi.org/10.1016/j.chb.2023.107792>
- Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer.
- Lu, L., Zhang, P., & Zhang, T. C. (2021). Leveraging “human-likeness” of robotic service at restaurants. *International Journal of Hospitality Management*, 94, 102823. <https://doi.org/10.1016/j.ijhm.2020.102823>
- Lv, X., Liu, Y., Luo, J., Liu, Y., & Li, C. (2021). Does a cute artificial intelligence assistant soften the blow? The impact of cuteness on customer tolerance of assistant service failure. *Annals of Tourism Research*, 87, 103114. <https://doi.org/10.1016/j.annals.2020.103114>
- Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., & Clark, J. (2024). The AI index 2024 annual report. *AI Index Steering Committee, Institute for Human-Centered AI, Stanford University*. <https://doi.org/10.48550/arXiv.2405.19522>
- McGuire, W. J. (1985). Attitudes and attitude change. In *Handbook of Social Psychology* (pp. 233–346). Random House.
- Mende, M., Scott, M. L., Van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, 56(4), 535–556. <https://doi.org/10.1177/0022243718822827>
- Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>
- Sandoval, E. B., Brandstatter, J., Yalcin, U., & Bartneck, C. (2021). Robot likeability and reciprocity in human robot interaction: Using ultimatum game to determinate reciprocal likeable robot strategies. *International Journal of Social Robotics*, 13(4), 851–862. <https://doi.org/10.1007/s12369-020-00658-5>
- Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joubin, F. (2013). To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, 5, 313–323. <https://doi.org/10.1007/s12369-013-0196-9>
- Smith, A. K., Bolton, R. N., & Wagner, J. (1999). A model of customer satisfaction with service encounters involving failure and recovery. *Journal of Marketing Research*, 36(3), 356. <https://doi.org/10.2307/3152082>
- Smith, C., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48(4), 813. <https://doi.org/10.1037/0022-3514.48.4.813>
- Stein, J. P., & Ohler, P. (2017). Venturing into the uncanny valley of mind — The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, 160, 43–50. <https://doi.org/10.1016/j.cognition.2016.12.010>
- Tellefsen, T., & Thomas, G. P. (2005). The antecedents and consequences of organizational and personal commitment in business service relationships. *Industrial Marketing Management*, 34(1), 23–37. <https://doi.org/10.1016/j.indmarman.2004.07.001>
- Tsarenko, Y., & Rooslan Tojib, D. (2011). A transactional model of forgiveness in the service failure context: A customer-driven approach. *Journal of Services Marketing*, 25(5), 381–392. <https://doi.org/10.1108/08876041111149739>
- Wang, Y. S., Wu, S. C., Lin, H. H., & Wang, Y. Y. (2011). The relationship of service failure severity, service recovery justice and perceived switching costs with customer loyalty in the context of e-tailing. *International Journal of Information Management*, 31(4), 350–359. <https://doi.org/10.1016/j.ijinfomgt.2010.09.001>
- Webster, C., & Sundaram, D. S. (1998). Service consumption criticality in failure recovery. *Journal of Business Research*, 41(2), 153–159. [https://doi.org/10.1016/S0148-2963\(97\)00004-0](https://doi.org/10.1016/S0148-2963(97)00004-0)
- Weun, S., Beatty, S. E., & Jones, M. A. (2004). The impact of service failure severity on service recovery evaluations and post-recovery relationships. *Journal of Services Marketing*, 18(2), 133–146. <https://doi.org/10.1108/08876040410528737>

- Xie, L., & Lei, S. (2022). The nonlinear effect of service robot anthropomorphism on customers' usage intention: A privacy calculus perspective. *International Journal of Hospitality Management*, 107, 103312. <https://doi.org/10.1016/j.ijhm.2022.103312>
- Xu, J., Hsiao, A., Reid, S., & Ma, E. (2023). Working with service robots? A systematic literature review of hospitality employees' perspectives. *International Journal of Hospitality Management*, 113, 103523. <https://doi.org/10.1016/j.ijhm.2023.103523>
- Yang, M., Li, J., Feng, L., Chen, S. C., & Tseng, M. L. (2024). Service robot anthropomorphism on consumer usage intention: Curvilinear and linear effect. *Industrial Management & Data Systems*, 124(2), 612–639. <https://doi.org/10.1108/IMDS-08-2023-0538>
- Yang, Y., Liu, Y., Lv, X., Ai, J., & Li, Y. (2022). Anthropomorphism and customers' willingness to use artificial intelligence service agents. *Journal of Hospitality Marketing & Management*, 31(1), 1–23. <https://doi.org/10.1080/19368623.2021.1926037>
- Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1993). The Nature and Determinants of Customer Expectations of Service. *Journal of the Academy of Marketing Science*, 21, 1–12. <https://doi.org/10.1177/0092070393211001>
- Zhang, M., Cui, J., & Zhong, J. (2023). How consumers react differently toward humanoid vs. nonhumanoid robots after service failures: A moderated chain mediation model. *International Journal of Emerging Markets*, ahead-of-print(ahead-of-print). <https://doi.org/10.1108/IJOEM-06-2022-1023>
- Zhang, X., Balaji, M. S., & Jiang, Y. (2022). Robots at your service: Value facilitation and value co-creation in restaurants. *International Journal of Contemporary Hospitality Management*, 34(5), 2004–2025. <https://doi.org/10.1108/IJCHM-10-2021-1262>