No Hard Feelings: The Protective Power of AI Empathy During Interaction Failures

Fotis Efthymiou[1], Alex Mari[2], Ertugrul Uysal[3] and Jeffrey Brooks[4]

[1]AI Empathy Research Initiative, Zurich, Switzerland
[2]American University Of Bahrain, Riffa, Bahrain
[3]Chair of Technology Marketing, ETH Zurich, Zurich, Switzerland
[4]Hume AI Inc., New York, NY, USA

Abstract

This study investigates the protective role of AI-delivered empathy during interaction failures in voice-based human–AI encounters. In a 2 (AI Empathy: empathic vs. non-empathic) × 2 (Interaction Failures: present vs. absent) between-subjects experiment ($N = 394$), participants interacted with a banking voice assistant powered by Hume AI's empathic voice interface. Results show that when interaction failures occurred, empathic AI significantly increased user satisfaction and reduced verbal aggression, with effects mediated by mind perception and social presence. Beyond self-report measures, a sentiment analysis and real-time prosodic emotion analysis revealed that empathic AI shifted users' language toward more positive sentiment and prevented anger escalation during failure episodes. These findings position AI empathy as an essential protective mechanism in conversational AI, and introduce utterance-level emotion measurement as a methodological innovation for studying dynamic affective processes in human–AI interaction.

Keywords

AI empathy, conversational AI, voice assistants, interaction failures, emotion recognition, human-AI interaction

1. Introduction

Conversational AI is rapidly transforming human-AI interactions in everyday contexts, such as frontline services. Voice assistants (VAs), powered by large language models and affective computing, are increasingly employed in high-stakes industries such as banking to handle tasks traditionally performed by humans [1, 2]. Yet, despite improvements in accuracy and efficiency, AI-led service encounters remain vulnerable to process breakdowns, which frustrate users, undermine satisfaction, and often elicit hostility [3, 4]. Past research highlights empathy as a cornerstone of effective service delivery [5], but whether *AI-delivered empathy* can meaningfully improve user experiences—particularly when failures occur—remains an open question [6, 7]. This study investigates the role of AI empathy in shaping user satisfaction and verbal aggression during interaction failures, while also introducing new methodological opportunities to study user emotions dynamically and in real time.

2. Method and Design

We conducted a 2 (AI Empathy: empathic vs. non-empathic) × 2 (Interaction Failures: present vs. absent) between-subjects experiment in a simulated banking scenario. 394 participants ($M_{age} = 39.6$, 42% male) were randomly assigned to one of the four experimental conditions. Participants acted as customers contacting a bank's VA regarding a missing transfer of $500. Interactions were powered by Hume AI's empathic voice interface (EVI), a real-time, voice-to-voice conversational model that detects users' emotional expressions through their voice and responds with adaptive speech [8, 9].

EmpathiCH'26 Workshop Co-located with CHI'26 Conference on Human Factors in Computing Systems, April 13–17, 2026, Barcelona, Spain
fotis.efthymiou1@gmail.com (F. Efthymiou); Alex.Mari@aubh.edu.bh (A. Mari); euysal@ethz.ch (E. Uysal); jeff@hume.ai (J. Brooks)

In the empathic condition, the VA actively recognized emotional signals in the participant's voice and responded with supportive and validating language, dynamically adjusting tone and prosody. In the non-empathic condition, the VA remained polite and professional but withheld explicit emotional acknowledgement. We designed these configurations through an iterative prompting process informed by empathy literature, systematically operationalizing theoretical constructs into specific conversational behaviors. This approach enabled precise experimental control over empathic expression while preserving the dynamic, naturalistic qualities of voice interactions.

Interaction failures were designed by prompting the VA to repeatedly misunderstand user responses at predetermined points in the conversation, before ultimately resolving the issue. All participants completed post-interaction measures of mind perception [10], social presence [11], satisfaction [12], and verbal aggression [13], while EVI simultaneously captured real-time utterance-level emotional expression data throughout the exchange and generated the interaction transcripts.

3. Main Results

Our findings show that when no interaction failures occurred, empathic and non-empathic VAs performed comparably. However, when interaction failures occurred, users interacting with an empathic VA reported higher satisfaction ($M_{Empathic} = 4.34$ vs. $M_{Non\text{-}Empathic} = 3.36$) and lower verbal aggression ($M_{Empathic} = 3.79$ vs. $M_{Non\text{-}Empathic} = 4.22$) than those in the non-empathic condition. That is, empathic design becomes critical when things go wrong—shielding users from the frustration and hostility that interaction failures typically provoke.

These effects unfolded through a sequential mediation mechanism: empathic VAs were attributed greater mind perception, which in turn enhanced their social presence. Higher social presence then predicted greater satisfaction and reduced aggression. For satisfaction, the indirect effect of AI empathy through mind perception and social presence was significant ($b = .21$, $SE = .06$, 95% CIs [.10, .34]). For verbal aggression, the indirect effect was likewise significant, and negative ($b = -.15$, $SE = .05$, 95% CIs [-.26, -.07]). In both cases, direct effects were nonsignificant, indicating full mediation.

4. Beyond Self-Report: Dynamic Emotional Expression Analyses

While self-reported satisfaction and aggression capture participants' reflections after the interaction, our design also enabled analysis of emotional expressions in real time.

4.1. Sentiment Analysis of User Speech

Transcripts of participants' utterances were analyzed using the AFINN lexicon [14]. Results revealed that participants interacting with the empathic VA expressed significantly more positive sentiment ($M = 0.20$) than those in the non-empathic condition ($M = -0.16$; $p < .001$) (Figure 1). This suggests that empathic AI not only improves retrospective evaluations but also shifts the semantic tone of users' language during the interaction itself.

4.2. Prosodic Emotion Analysis of Vocal Expressions

Leveraging EVI's emotion recognition [8], we examined the trajectory of anger during interaction failures. Critically, these emotional expression scores are not self-reports collected after the fact—they are extracted from participants' vocal expressions during the interaction itself, providing a behavioral trace of emotional experience as it unfolds in real time. This utterance-level measurement allows us to observe the exact moment an interaction failure triggers anger in the user, and whether empathic AI design prevents that emotional spike from occurring. As expected, anger levels spiked at the moment of failure for participants in the non-empathic VA condition, with a significant positive increase over message sequence ($b = .00077$, $SE = .00027$, $p = .004$, 95% CI[.00025,.00130]). Strikingly, in the empathic VA condition, this anger spike did not occur; the regression slope was flat and nonsignificant

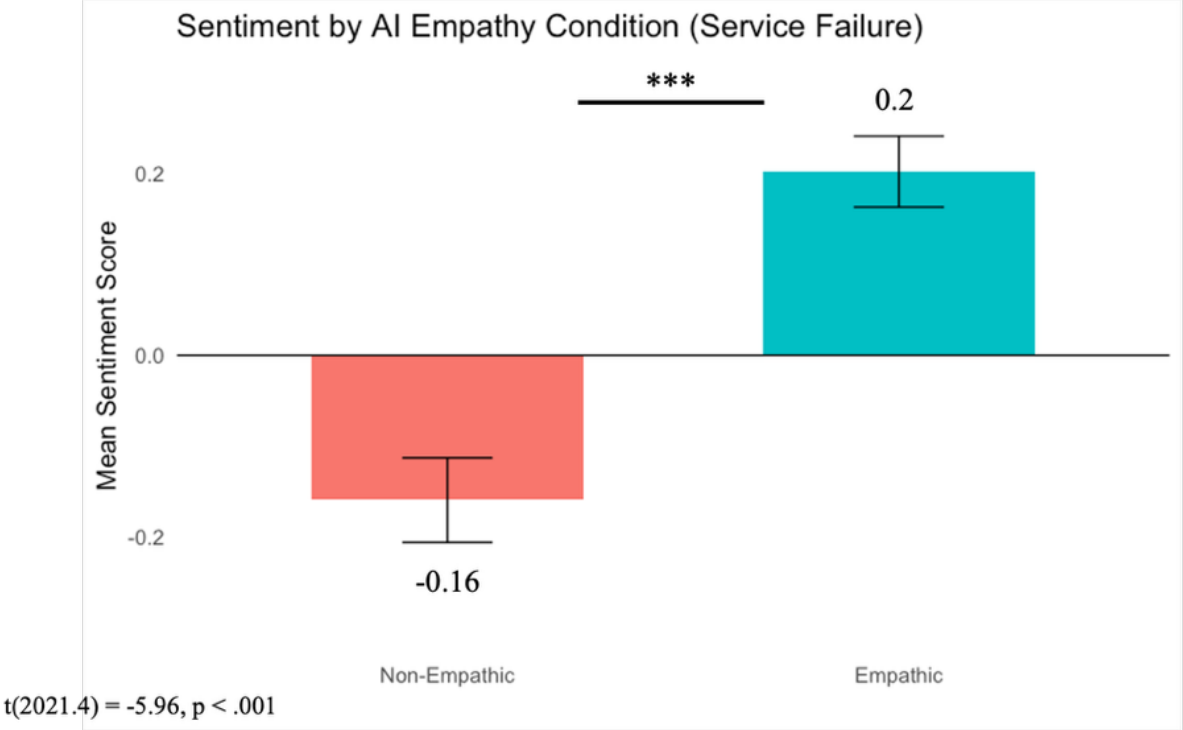

Figure 1: The effect of AI empathy on users' sentiment in the presence of a service failure.

($b = -.00019$, $SE = .00027$, $p = .474$, 95% CI[-.00071, .00033]). Participants' vocal expressions remained stable, showing no surge of anger during the failure episode (Figure 2). This novel finding demonstrates that AI-delivered empathy can prevent the escalation of negative emotions in real time, effectively neutralizing the emotional cost of interaction breakdowns.

Together, these analyses extend our conclusions beyond self-report to show how AI empathy tangibly alters the emotional dynamics of human-AI interactions. They also highlight the methodological innovation of combining survey data with utterance-level emotion measurement, setting a new benchmark for ecological validity in human–AI interaction research.

5. Discussion and Contributions

This study makes several contributions to the emerging literature on AI empathy, voice-based human-AI interactions, as well as customer service. First, contrary to prior research suggesting that humanlike features can sometimes backfire and lower satisfaction in high-frustration contexts [15, 16], we provide causal evidence that AI empathy consistently improves satisfaction and reduces misbehavior, especially under failure conditions. Rather than being a "nice-to-have," AI empathy emerges as an essential protective feature in high-stakes interaction contexts.

Second, we introduce real-time emotion analysis as a methodological innovation. Whereas prior research has largely relied on self-reports or scripted chatbot interactions [7], our work shows how emotion recognition at the utterance level provides richer insight into users' dynamic affective states. This enables a new class of behavioral evidence for how empathic AI regulates emotions during encounters.

Finally, we contribute to the literature on service recovery [17, 18] by showing that AI-delivered empathy can act as a buffer against double deviations, mitigating both dissatisfaction and hostility when the AI itself causes the failure. In doing so, we position AI empathy not as an optional enhancement but as a strategic safeguard for organizations adopting conversational AI at scale.

By combining survey evidence, sentiment analysis, and vocal emotion trajectories, this work shows

that empathy is both a psychological mechanism and a real-time regulator of user emotions in AI-led encounters. As conversational AI becomes ubiquitous, embedding empathic capacities may be critical to ensuring emotionally intelligent and constructive digital interactions.

This study offers a practical example of how empathic AI design can be operationalized in voice-based human-AI collaboration. By examining how emotion recognition and adaptive responses shape user experiences, we demonstrate that empathy in conversational AI technologies is not simply about mimicking human warmth—it functions as a protective mechanism whose benefits emerge particularly in moments of system failure and user frustration. This situational quality suggests that as we continue developing empathic capabilities in AI, careful attention must be paid to the contexts in which empathy is deployed. Effective implementation requires understanding not just whether empathic responses improve interactions, but when, how, and under what conditions they do so. Such contextual awareness is essential for creating AI systems that engage users in both effective and ethically sound ways.

Declaration on Generative AI

During the preparation of this work, the author(s) have used Claude to improve writing clarity.

References

[1] J. Wirtz, P. G. Patterson, W. H. Kunz, T. Gruber, V. N. Lu, S. Paluch, A. Martins, Brave new world: Service robots in the frontline, Journal of Service Management 29 (2018) 907–931. doi:10.1108/josm-04-2018-0119.

[2] Y. Liu-Thompkins, S. Okazaki, H. Li, Artificial empathy in marketing interactions: Bridging the human-ai gap in affective and social customer experience, Journal of the Academy of Marketing Science 50 (2022) 1198–1218. doi:10.1007/s11747-022-00892-5.

[3] Y. Sun, S. Li, L. Yu, The dark sides of ai personal assistant: Effects of service failure on user continuance intention, Electronic Markets 32 (2022) 17–39. doi:10.1007/s12525-021-00483-2.

[4] J. Riquel, A. B. Brendel, F. Hildebrandt, M. Greve, A. Dennis, "f*** you!" – an investigation of humanness, frustration, and aggression in conversational agent communication, in: Proceedings of the 42nd International Conference on Information Systems, 2021.

[5] J. Wieseke, A. Geigenmüller, F. Kraus, On the role of empathy in customer-employee interactions, Journal of Service Research 15 (2012) 316–331. doi:10.1177/1094670512439743.

[6] N. Heirati, V. Pitardi, J. Wirtz, C. Jayawardhena, W. Kunz, S. Paluch, Unintended consequences of service robots – recent progress and future research directions, Journal of Business Research 194 (2025) 115366. doi:10.1016/j.jbusres.2025.115366.

[7] X. Lv, Y. Yang, D. Qin, X. Cao, H. Xu, Artificial intelligence service recovery: The role of empathic response in hospitality customers' continuous usage intention, Computers in Human Behavior 126 (2022). doi:10.1016/j.chb.2021.106993.

[8] J. A. Brooks, V. Tiruvadi, A. Baird, P. Tzirakis, H. Li, C. Gagne, M. Oh, A. Cowen, Emotion expression estimates to measure and improve multimodal social-affective interactions, in: International Conference on Multimodal Interaction, 2023, pp. 353–358. doi:10.1145/3610661.3616129.

[9] A. S. Cowen, H. A. Elfenbein, P. Laukka, D. Keltner, Mapping 24 emotions conveyed by brief human vocalization, American Psychologist 74 (2019) 698–712. doi:10.1037/amp0000399.

[10] H. M. Gray, K. Gray, D. M. Wegner, Dimensions of mind perception, Science 315 (2007) 619–619.

[11] J. van Doorn, M. Mende, S. M. Noble, J. Hulland, A. L. Ostrom, D. Grewal, J. A. Petersen, Domo arigato mr. roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences, Journal of Service Research 20 (2017) 43–58. doi:10.1177/1094670516679272.

[12] E. W. Anderson, C. Fornell, S. K. Mazvancheryl, Customer satisfaction and shareholder value, Journal of Marketing 68 (2004) 172–185. doi:10.1509/jmkg.68.4.172.42723.

[13] A. H. Buss, M. Perry, The aggression questionnaire, Journal of personality and social psychology 63 (1992) 452.
[14] J. Silge, D. Robinson, Text mining with R: A tidy approach, O'Reilly Media, 2016.
[15] C. Crolic, F. Thomaz, R. Hadi, A. T. Stephen, Blame the bot: Anthropomorphism and anger in customer–chatbot interactions, Journal of Marketing 86 (2022) 132–148. doi:10.1177/00222429211045687.
[16] W.-H. S. Tsai, D. Lun, N. Carcioppolo, C.-H. Chuan, Human versus chatbot: Understanding the role of emotion in health marketing communication for vaccines, Psychology & marketing 38 (2021) 2377–2392.
[17] M.-H. Huang, R. T. Rust, The caring machine: Feeling ai for customer care, Journal of Marketing 88 (2024) 1–23. doi:10.1177/00222429231224748.
[18] L. L. Bove, Empathy for service: Benefits, unintended consequences, and future research agenda, Journal of Services Marketing 33 (2019) 31–43. doi:10.1108/jsm-10-2018-0289.

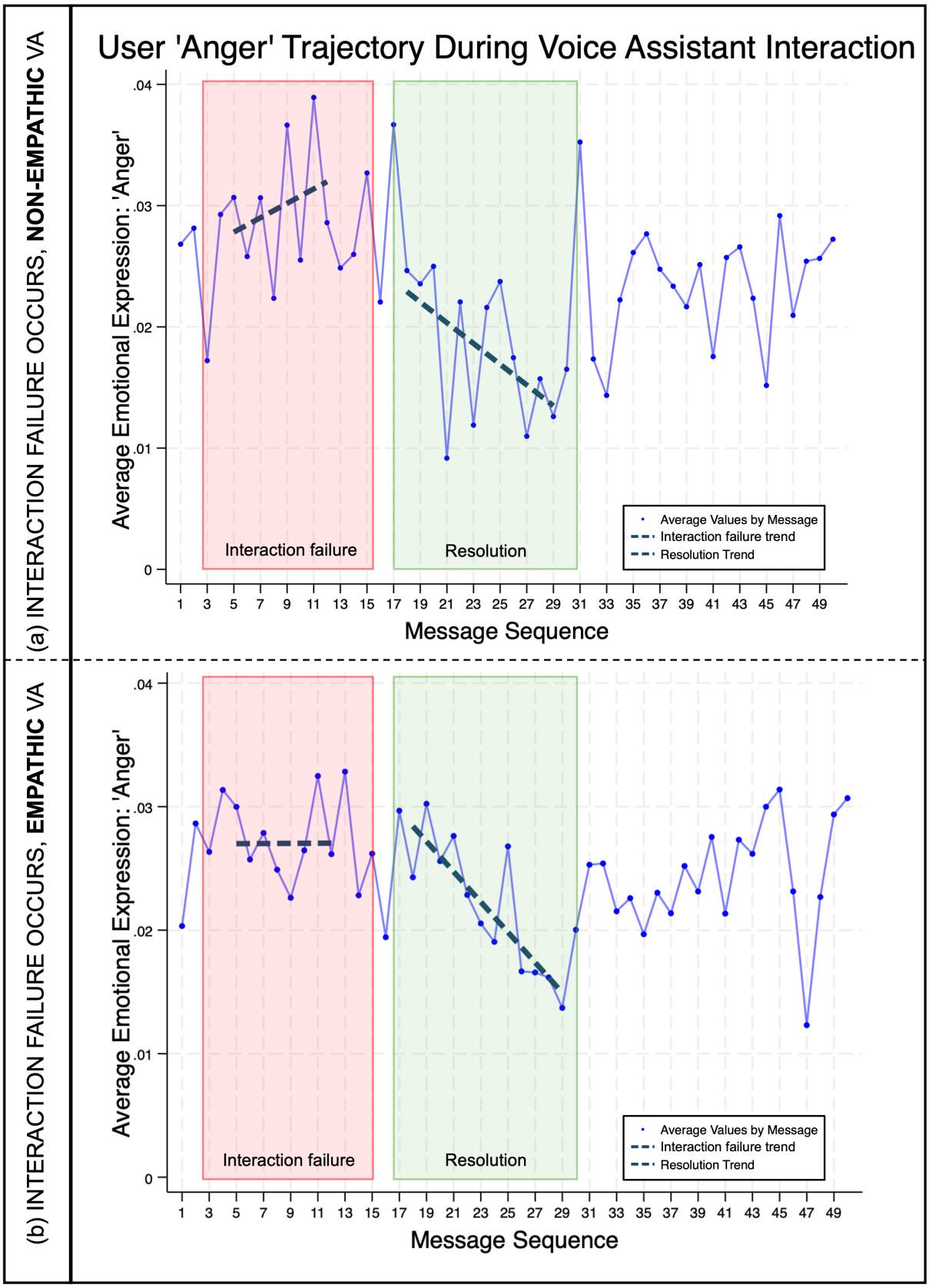

Figure 2: AI-delivered empathy can prevent the escalation of users' anger (as measured through user's voice in real time) during service failures.