Please quote as: Benner, D., Janson, A. & Tan, C.-W. (2025). Less is More or More is Better? Comparing Static and Generative AI-Based Recovery for Chatbot Service Breakdowns. International Conference on Information Systems (ICIS), Nashville, USA.

Less is More or More is Better? Comparing Static and Generative AI-Based Recovery for Chatbot Service Breakdowns

Short Paper

Dennis Benner
University of Kassel
Information Systems
benner@uni-kassel.de

Andreas Janson
University of St.Gallen
Institute of Information Systems and Digital Business
andreas.janson@unisg.ch

Chee-wee Tan
The Hong Kong Polytechnic University
Management & Marketing, Management Information Systems
chee-wee.tan@polyu.edu.hk

Abstract

Conversational breakdowns in chatbot-based customer service interactions can impair service quality and brand perception. Our study investigates how static versus generative AI (GAI)-based recovery affects user experience in such scenarios. Using a dual-study design, Study 1 examines two static strategies: Inform (explanation-focused) and Repair (action-focused). We find that Repair significantly enhances recovery quality and service perception, while Inform offers limited benefits when used alone. Study 2 explores the potential of GAI-driven recovery, employing a chatbot capable of nuanced, interactive support. Preliminary findings suggest that GAI-based recovery may increase resolution success, particularly in complex interactions, but could demand higher cognitive effort from users. By comparing both approaches and introducing a sequencing approach, our study provides novel insights into how recovery design and GAI influence service outcomes. Our results inform the design of future chatbot service systems, suggesting that adaptive, generative recovery may complement traditional static recovery strategies under specific conditions.

Keywords: Conversational Breakdown, Breakdown Recovery, Generative AI

Introduction

Whether assisting with inquiries or resolving service issues, chatbots increasingly mediate digital service encounters (Han et al., 2024; Huang & Rust, 2020). Chatbots offer natural, human-like interactions and aim to deliver personalized, efficient service at scale (Blümel & Jha, 2023; Han et al., 2024). As a result, their adoption continues to accelerate, especially in the light of generative artificial intelligence (GAI), which contributes to this trend (Schöbel et al., 2023). Despite their growing importance, recent advances, and research efforts, chatbot interactions still often fall short of expectations (Chhabra et al., 2024; Janssen et al., 2021). A key point of failure occurs when the chatbot misinterprets or simply does not understand user input or when users are unsure how to engage effectively with the chatbot. This can then lead to an abrupt termination of the dialogue and service process (Chhabra et al., 2024; Choi et al., 2021). These conversational breakdowns, defined as a failure to understand and respond appropriately (Ashktorab et al., 2019; Zeng et al., 2024), can severely harm service outcomes (Choi et al., 2021; Klein et al., 2024). Not only do they degrade perceived service and process quality (Chen et al., 2022; Klein et al., 2024), but they can also damage brand reputation through reduced satisfaction, diminished repurchase intentions, and increased negative word of mouth (Klein et al., 2024; Schuetzler et al., 2021; van Vaerenbergh et al., 2012). To mitigate these consequences, researchers and practitioners have explored various recovery strategies from a theoretical and practical perspective (e.g., Haupt et al., 2023; Benner et al., 2021; Le, Hoang T. P. M. et al., 2023; Zeng et al., 2024). In general, recovery strategies can be defined as design features or conversational responses aimed and restoring a functional dialogue (Ashktorab et al., 2019; Benner et al., 2021; Zeng et al., 2024). Existing research often examines isolated mechanisms or short-term outcomes, leaving a gap in our understanding of how recovery strategies influence broader user experiences and brand-related perceptions (Choi et al., 2021; Gelbrich, 2010; Janssen et al., 2021; Mariani et al., 2023). An emerging but underexplored opportunity lies in the application of GAI to chatbot recovery. Unlike statically programmed or rule-based systems, GAI-powered chatbots can generate nuanced, context-sensitive responses that may better support users during breakdowns (Feuerriegel et al., 2023; Sun et al., 2024). Thus, GAI-driven chatbots hold the potential to go beyond scripted responses or button-based repair strategies by engaging users in more fluid, cooperative recovery dialogues. However, research into the effectiveness of GAI for conversational breakdown recovery remains limited (Villarroel Ordenes et al., 2025). Key questions about the impact of GAI on factors like interaction fluency, cognitive load, and service quality remain unanswered and merit empirical investigation. Moreover, the impact of pitfalls of GAI like hallucination on breakdown recovery is unclear as well. Against this backdrop, we aim to contribute a more holistic understanding of recovery strategies by integrating both static and GAI-based approaches to breakdown recovery. Therefore, we raise the following research question (RQ): ***How do static versus generative recovery strategies impact conversational breakdowns in customer service?***

To answer our RQ, this research in progress article encompasses two studies on this topic. Our first study investigates the impact of traditional i.e., static recovery while our second study aims to explore the potential of GAI for breakdown recovery. Overall, we aim to investigate the potential of GAI for breakdown recovery and to analyze how static vs. generative recovery impact conversational recovery from a service perspective. To this end, we present our approach and planned investigation in this regard. First insights from our dual study approach suggest that when it comes to static recovery, less may be more as elaborate responses including information and explanation do not outperform the control group while simple instructions and dialog options significantly and positively impact the recovery. Regarding our second study and investigation of the potential of GAI, we can describe a trend where more may be better as longer conversations but more comprehensive interactions with the chatbot may lead to improved recovery.

Conversational Breakdown and Recovery in Digital Service

Despite advancements in conversational technology, chatbots continue to experience conversational breakdowns (Weiler et al., 2022), i.e., situations where they fail to understand or appropriately respond to user input. These failures stem from the inherent ambiguity of natural language, lack of contextual awareness, limitations in dialogue management, or simply the inability of users who may have a different understanding of the chatbots capabilities (Ashktorab et al., 2019; Do et al., 2023). While such failures are inherent to the medium, chatbots can engage in recovery strategies to overcome these breakdowns (Ashktorab et al., 2019). Recovery strategies may help to clarify the situation and reestablish a common ground or mutual understanding between chatbot and human user or actively guide the user toward a resolution (Weiler et al., 2022). These strategies vary in form: from simple information provisioning that signal a problem occurred and may try to explain it, to more proactive mechanisms like instructing the user to take certain action or even provide response or dialogue options (Ashktorab et al., 2019; Benner et al., 2021; Weiler et al., 2022). Emerging studies on recovery strategies for conversational breakdowns differentiate between informational strategies (e.g., “I didn’t understand that because am not programmed to do this.”) and actionable strategies (e.g., suggesting rephrasing the last request to the user). The effectiveness of such strategies often depends on how they shape the user’s cognitive load, expectations, and perceived agency during the recovery process (Cheng et al., 2022; Hong et al., 2004). For example, providing an explanation without a clear path forward may increase user frustration or lead to task abandonment, highlighting that how a recovery is initiated may matter more than the acknowledgment itself (Choi et al., 2021; van Vaerenbergh et al., 2012). Moreover, empirical work that directly investigates how specific recovery strategies mitigate the negative consequences of conversational breakdowns in the context of digital service, particularly in terms of process quality, service quality, and brand reputation, remains limited (Benner et al., 2021; Mariani et al., 2023). Furthermore, with the recent advance in GAI, novel opportunities but also challenges have emerged. GAI introduces more advanced capabilities for breakdown recovery, such as conversational adaptability, but also risks, including hallucinations and unpredictable responses (Bang et al., 2023; Villarroel Ordenes et al., 2025). A particular concern is hallucination, where the system fabricates information or misrepresents facts (Bang et al., 2023). Hallucinations during breakdown recovery could therefore be detrimental and have an even worse outcome than static implementations that are standard today. Thus, we argue to investigate the impact of state-of-the-art static breakdown recovery versus emerging GAI-driven breakdown recovery. In sum, these dual potentials necessitate a systematic investigation into how GAI influences breakdown recovery. In this regard, current research, especially empirical studies on GAI-enabled breakdown recovery, is sparse (Mariani et al., 2023; Villarroel Ordenes et al., 2025). We theorize that both static and generative approaches have their merits and drawbacks. While static recovery suits small problem-solution spaces very well, problems that may require discussion (i.e., similar to human agent support) may benefit or even require a generative approach with a wider problem-solution space. However, a generative approach similar to a chat implementation like ChatGPT may not include dialogue options which in return may require a higher degree of cognitive effort from the user to truly leverage the, in theory, superior capabilities of GAI (Hong et al., 2004; Villarroel Ordenes et al., 2025). Therefore, the cognitive aspect may play a difficult dual role in our study. One the one hand, lowering the cognitive effort required is desirable, but on the other hand it may not be avoidable due to the nature of the dialogue interaction and even act as mediator. Regarding the recovery strategies, we follow related studies that have proposed conceptualizations and frameworks based on literature reviews and theory (i.e., Benner et al., 2021; Ashktorab et al., 2019). While both studies present multiple strategies, we find that not all are actually viable for recovery. For example, the “ignore” strategy (Ashktorab et al., 2019) simply ignores any potential breakdown and continues disregarding its own failure. Likewise, the “handover” or “defer” strategy (Ashktorab et al., 2019; Benner et al., 2021) doesn’t solve the conversational breakdown but shifts it to a human agent. While this may resolve the service breakdown (Poser et al., 2021), it does not offer potential to recover the conversational breakdown but rather cements it. Thus, we only consider strategies that have potential to recover the d human-chatbot interaction and focus on two types of strategies: 1) ones that provide information (i.e., “Inform”), and 2) ones that provide instruction or action for repair (i.e., “Repair”). Regarding the Inform strategy, it starts with acknowledging the potential breakdown and follows with explaining it to the user. The idea is to create common ground between the user and the chatbot but leaves the responsibility of recovery on the side of the user. In contrast, the Repair strategy, provides actionable instructions or options on which the user can act. However, this strategy omits any other potentially useful information. Thus, we focus on these two contrasting but potentially complementary strategies (see Table 1).

	Inform	**Repair**
Role	Passive, user needs to recover, chatbot doesn’t attempt recovery	Active, chatbot takes proactive role in recovery, user more passive
Examples	• Acknowledge issue, clarify to user • Explain reason for issue	• Instructions, advice, command • Buttons, dialog options, choices
Potential Benefits	Creates understanding, adjusts expectation of the user	Provides actionable pathways to the user, can be efficient if issue is in scope
Potential Issues	Potentially high cognitive effort required, no solution or instruction	Instructions and options are insufficient without information or explanation
GAI Opportunities	More in-detail information & explanation, can elaborate in detail	Can reason and discuss with the user, ask for more data and understand it also
GAI Challenges	Hallucination, provide false/fake information or explanation	Misinterpretation, failure to comprehend or reason repair dialog

Table 1: Inform vs Repair Recovery Strategies

Research Approach and Methodology

We employ an overlapping study design (Dennis & Valacich, 2001). Thus, we employ a dual study approach (see Table 2) where we first test two strategies for conversational recovery in study 1 before we investigate in study 2 how GAI and its orchestration perform against the best combined strategies of study 1 (i.e., study 1, T3 = study 2, T0). We hypothesize that the order of which static and GAI recovery are administered will differ. Both studies follow an in-between subject design with four experimental and are executed online using the Prolific platform, balanced sample, and randomization. All participants participate on voluntary basis and receive fair treatment and payment.

Study 1	-	Inform		**Study 2**			
-	T0	T1		Static	Static→GAI	GAI→Static	GAI
Repair	T2	**T3**	→	**T0**	T1	T2	T3

Table 2. Dual Study Setup

Study 1: Assessing Static Recovery

Study 1 was conducted with a total of 180 participants. Participants were randomly assigned to one of four experimental conditions: a control group (no recovery strategy, T0), strategy 1 group ("Inform", T1), strategy 2 group ("Repair", T2), and a group combining both strategies (T3). Regarding the design of the strategies (see Figure 1), we focus on acknowledging the breakdown and explaining the issue to the user for the Inform strategy, whereas for the Repair strategy we ask the user to rephrase the request and/or pick a potential option. For the generative chatbot (study 2) we implemented these strategies by prompting them according to the implementation in study one. To test the strategies, we constructed a relatable self-service scenario that participants had to follow and playthrough with the chatbot. Participants engaged in a simulated customer service interaction involving a subscription modification task. The scenario was intentionally designed to be simple and relatable across diverse demographic backgrounds. The task requires participants to suspend their account and service for a business trip and make it available to family and friends. A scenario common in subscription-based service that usually required customer service agent intervention as each case is individual.

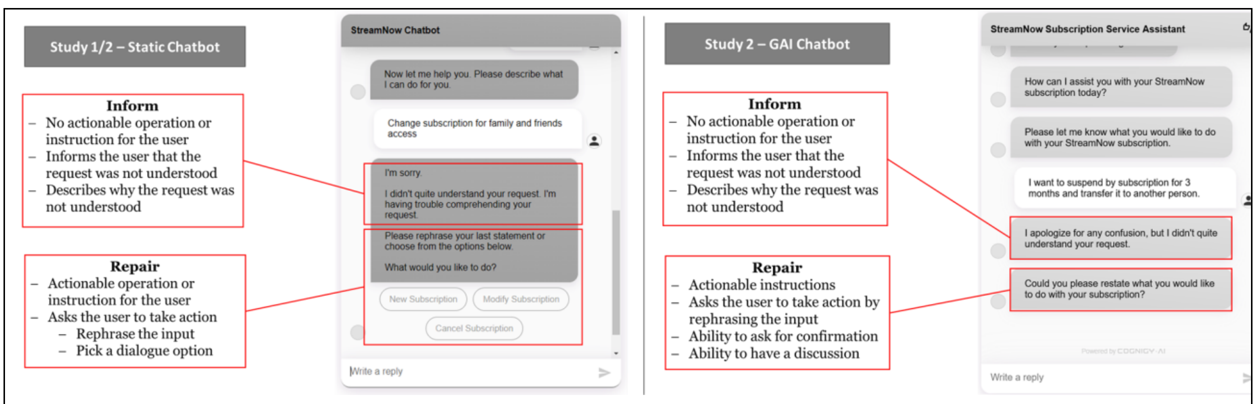

Figure 1. Chatbot Design & Implemented Recovery Strategies

Before conducting our main study, we prepared a pre-test in order to assess our constructed scenario regarding its complexity, relatability, and perceived stake (i.e., urgency or value of service). To this end, we conducted a survey (n = 20) study where we asked participants to carefully read our constructed scenario and evaluate it regarding complexity, relatability, and perceived stake on a 7-point Likert scale. As results, we can report that our constructed scenario rates 3.00 for complexity (*T=-4.47, p=0.00****), 4.49 for reliability (*T=2.19, p=0.04**) and 4.36 for stake (*T=1.61, p=0.12*). Thus, we conclude that our task indeed depicts a scenario that participants perceive as significantly relatable. While complexity is on the lower side, potentially limiting application to complex real-world settings, it also implies that any observed cognitive load or effort may be attributed to the interaction between participants and the chatbot. To complete their task, the participants then engage in an interaction with the customer service chatbot. The interaction follows a predetermined path with controlled breakdowns (see Figure 2) but allows for the interaction to play out naturally. Additionally, the chatbot interface was kept neutral in design and branding, and excluded anthropomorphic features, to minimize bias and isolate the effects of the recovery strategies. During the interaction, two predefined conversational breakdowns were systematically embedded into the dialogue flow to ensure consistent experimental conditions. Depending on the assigned treatment group, the chatbot either provided no recovery support (control), only information about the failure (Inform strategy), actionable options to resolve the issue (Repair strategy), or both. This setup allowed for a controlled examination of how different breakdown recovery strategies impact user experience, service perception, and cognitive effort. Additionally, if participants could not resolve the problem after three attempts per constructed breakdown, the interaction would end with the chatbot admitting to not be able to help. After the interaction, participants were redirected to a questionnaire.

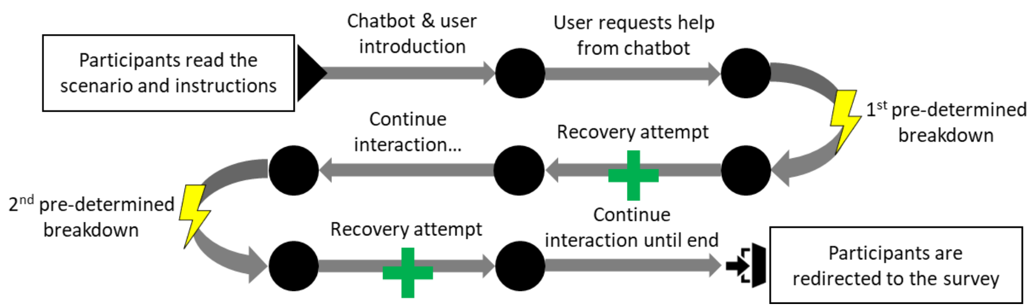

Figure 2. Interaction Flow

To assess the impact of chatbot recovery strategies in our customer service setting, we measure the following key constructs: recovery quality (PQ), service quality (SQ), brand reputation (BR), and cognitive effort (CE). RQ was assessed using the co-creation service recovery scale (Dong et al., 2008), while SQ was measured with digital service quality instruments (Eisingerich & Bell, 2008; Rafaeli et al., 2008). BR was evaluated through a coping- and frustration-oriented lens (Gelbrich, 2010), focusing on emotional reactions and potential word-of-mouth behavior. CE was included as a control and potential moderating variable (Hong et al., 2004). All constructs demonstrated satisfactory internal consistency, with Cronbach's alpha exceeding 0.7. To ensure data quality, attention checks were used in the survey (Oppenheimer et al., 2009).

Study 2: Investigating the Potential of GAI

For study 2 we expand our scope from standard static recovery to GAI-driven recovery. To this end, we developed a GAI version of our chatbot referring to prompt engineering guides and best practices introduced by Mollick and Mollick (2023) and Braun et al. (2024). Our approach to prompt engineering the GAI version of our chatbot was to have the chatbot mimic the conversational flow of the non-GAI i.e., static chatbot version but allow for free conversation at any given time. However, we did place guardrails to ensure that the chatbot sticks to the topic and task without veering off. To embed conversational breakdowns alike to the static chatbot, we instructed the GAI chatbot to pretend to not or misunderstand the conversation at the exact same predefined steps as the static chatbot. For the recovery process itself, we did not implement any other limitations or restrictions other than those relevant to the task (e.g., if speaking about suspending the account, do not allow users to create a new one). Unlike in study 1, where a scripted and rule-based approach to recovery is applied, in study 2 we focus on GAI. The chatbot is instructed i.e., prompted to inform, and explain breakdowns to the user and present conversational help (e.g., asking for info, confirmation, making suggestions to the customer). To empower the chatbot to recover from the breakdown we have prompt-engineered the strategies following study 1 and the theory behind them (see Figure 2 above). Furthermore, we also slightly adapted the task theme to make it even more relatable and also expanded the task to increase complexity compared to study 2. The general customer service scheme remains for study 2, however, now participants have to suspend and transfer their subscription to a specific person (i.e., persona from the instructions) and also have the option to negotiate a better deal with the chatbot. Our reasoning is that the former should increase relatability while the latter should better reflect a similar real-world setting. Regarding the breakdown structure and general conversational flow beyond these adaptations we did not change anything in order to ensure both our studies can be compared to each other for evaluation. To conduct study 2, we plan to follow the same approach as for study 1. We will conduct the study on Prolific with approximately 200 participants in total (i.e., target of 50 per group). For now, we have already conducted a small-scale pretest with 20 participants in total to assess the viability of our approach and potentially draw first insights or trends from data. Moreover, for study 2 we will collect all chat logs including timestamps as well as free text comments from participants. With this approach, we will be able to get deeper qualitative insights into the interactions between participants and our chatbot per treatment group. We will also be able to calculate the exact time taken for the interaction and determine if changes in the pace within an interaction occur. For instance, in T1 and T2 of study 2 where static and GAI chatbots switch and thus where we expect to observe different timings in the interaction. The free text component will also potentially enable us to run sentiment and emotion analyses on the participants feedback in order to determine if and how the different chatbot designs affect them on a deeper, more personal level instead just relying on a quantitative survey.

Preliminary Results

In this section we present our results from study 1 and preliminary insights and expected results from study 2. Prior to analyzing group differences, we tested homogeneity of variances using Levene's test. This test is essential in determining the appropriate type of ANOVA, as the standard ANOVA assumes equal variances across groups. Our results indicate that this assumption is violated for the constructs under investigation, suggesting significant variance heterogeneity ($p < 0.05$ for all). Given the lack of homogeneity, we applied the Welch-ANOVA, a robust alternative to the standard ANOVA that does not require equal variances and is particularly suitable for designs with unequal group sizes, as is the case in our sample. The results of the Welch-ANOVA indicate statistically significant differences between treatment groups across all measured constructs ($p < 0.05$ for all). This finding implies that the experimental manipulations had a meaningful impact on participants' responses. To estimate effect sizes, we additionally conducted a standard ANOVA with a Bonferroni correction (Dunn, 1961). Although the standard ANOVA is less suitable when variance homogeneity is not met, it allows us to assess the effect size (ε^2). The Bonferroni adjustment compensates for the increased risk of Type I error due to multiple comparisons and is considered one of the most conservative correction methods (Dunn, 1961). Across all constructs, we observe large effect sizes ($\varepsilon^2_{PQ} = 0.445$; $\varepsilon^2_{SQ} = 0.466$; $\varepsilon^2_{BR} = 0.395$; $\varepsilon^2_{CE} = 0.314$), indicating that a substantial portion of the variance in outcomes can be attributed to the experimental conditions. To further explore the group differences revealed by the ANOVA, we conducted Games-Howell post-hoc tests, which are appropriate given the lack of variance homogeneity and unequal group sizes (Games & Howell, 1976). These comparisons reveal clear patterns in the effects of recovery strategies (see Figure 3).

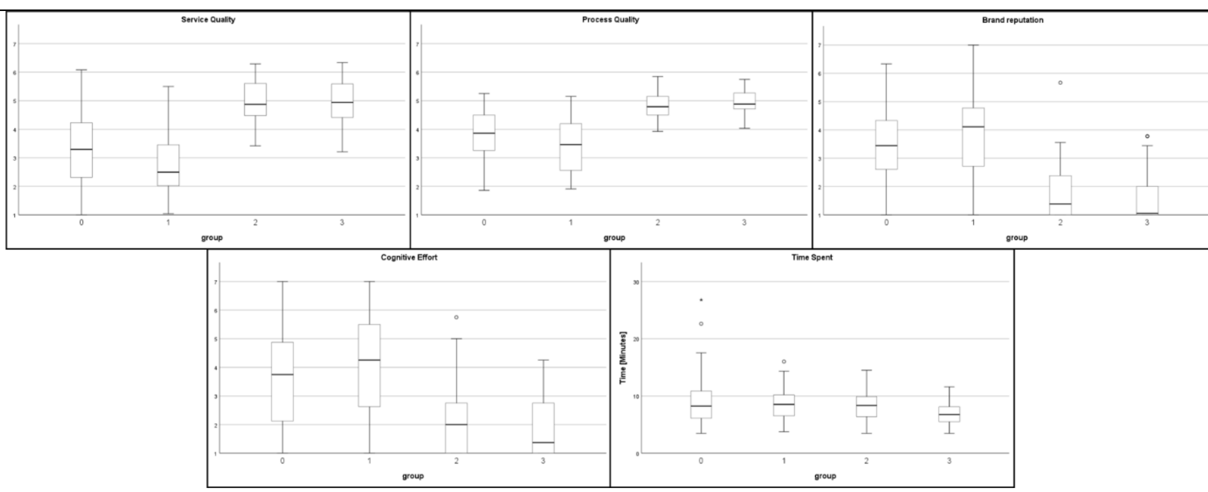

Figure 3. Study 1 Results

From our tests, we find no significant differences between groups T0 (control) and T1 (Inform) as well as between groups T2 (Repair) and T3 (Inform + Repair) of study 1 (all *n.s.* $p > 0.05$). However, we report significant differences between groups T0/T1 and T2/T3, respectively (all $p < 0.05$). These results underscore the strong impact of the Repair strategy across the board, in contrast to the limited effectiveness of Inform strategy alone. Moreover, the results also suggest that the Inform strategy (T1) has little added effect compared to the control (T0), and that combining Inform and Repair (T3) does not outperform the Repair strategy alone (T2). Moreover, while not statistically significant, we can observe a slight trend towards less time required for T2/3 that rely on the Repair strategy. Regarding our planned study 2, we can report first tendencies from our small-scale pre-test. From the pre-test we can observe that the control treatment T0 (i.e., T3 from study 1) concludes the interaction the fastest. However, the new GAI treatment (T3) concludes the interaction with the highest success rate. Regardless, because of the small sample size of our pre-test we are not confident to report statistical measurements at this point in time. Nevertheless, we expect to observe this trend to strengthen in our full-scale study. Regardless, we are confident enough to state that GAI-driven recovery is a viable alternative to the industry standard static approach and may augment or even replace these in the future.

Discussion, Next Steps & Expected Outcome

Our findings from study 1 indicate that the Inform strategy, when applied in isolation, does not significantly improve breakdown recovery whatsoever as none of our measured variables present significant results. This result may contrast with prior research, which has reported positive effects of this mechanisms on user satisfaction and behavioral intentions (e.g., Cheng et al., 2022; Choi et al., 2021). However, a closer look at studies like these reveals that information or explanations were often combined with other elements in attempt to recover a breakdown e.g., emotional cues or empathetic response. Thus, we argue that in these cases the isolated Inform strategy is not affective unlike what theory may suggest but rather complements other approaches such as social recovery by including emotional cues or trying to appeal to empathy. Furthermore, in our experiment, the interaction was strictly between a human and a chatbot, devoid of anthropomorphic features or social framing. While this approach may limit the effectiveness of the Inform strategy alone it also underscores the ineffectiveness of this strategy when used on its own further supporting our results from study 1. Moreover, while not statistically significant, we could observe a slight increase in cognitive effort and task failure rates for T1 (i.e., Inform) in study 1 that further underscores the ineffectiveness of this approach. We theorize that users may not necessarily be equipped to act on the information provided or interpret it in a way that resolves a breakdown. Thus, we argue that further research on truly effective design for breakdown recovery is required as current approaches to inform users are ineffective in practice while theory suggests otherwise. By contrast, presence of the Repair strategy showed clear benefits in both cases (T2, T3). Providing users with actionable recovery options, such as rephrasing prompts or selectable inputs, was associated with higher scores across all outcome measures, consistent with theoretical expectations and prior research (e.g., Cheng et al., 2022; Do et al., 2023). In practical terms, this strategy appears to shift the cognitive burden away from the user and toward the chatbot, which may be especially important in failure contexts where user frustration or uncertainty is already elevated (Gelbrich, 2010). From a theoretical perspective this may hint at the importance of delegation and agency in breakdown situations (Baird & Maruping, 2021; Fügener et al., 2022). We argue that such delegation and shift of agency may prove as particularly important or effective in breakdown recovery if chatbots can leverage GAI. Therefore, we expect results from study 2 that reflect our theorized outcome. However, at the same time GAI may also lead to higher cognitive effort similar to the Inform strategy which again may negatively impact the interaction. Therefore, we aim to investigate the role of cognitive load and effects of our strategies in this regard with study 2 in more detail. Furthermore, we expect that we may also see effects from switching static and GAI recovery sequences. We hypothesize that sequencing may lead to similar recovery effects as observed with the service recovery paradox.

Acknowledgements

This research is partially funded by Connex Programme of the Dr. Hans Riegel-Stiftung.

References

Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient Chatbots. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19* (pp. 1–12). ACM Press. https://doi.org/10.1145/3290605.3300484

Baird, A., & Maruping, L. M. (2021). The Next Generation of Research of IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS Quarterly*, *45*(1), 315–341.

Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., Do, Q. V., Xu, Y., & Fung, P. (2023, February 8). *A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity*. https://arxiv.org/pdf/2302.04023

Benner, D., Elshan, E., Schöbel, S., & Janson, A. (2021). What do you mean? Recovery Strategies to Overcome Breakdowns of Conversational Agents. *International Conference on Information Systems*.

Blümel, J. H., & Jha, G. (2023). Designing a conversational AI agent: Framework combining customer experience management, personalization, and AI in service techniques. *Hawaii International Conference on System Sciences*, 1397–1406. https://doi.org/102456

Braun, M., Greve, M., Kegel, F., Kolbe, L. M., & Beyer, P. E. (2024). Can (A)I Have a Word with You? A Taxonomy on the Design Dimensions of AI Prompts. *Hawaii International Conference on System Sciences*, 559–568. https://hdl.handle.net/10125/107396

Chen, Q., Gong, Y., Lu, Y., & Tang, J [Jing] (2022). Classifying and measuring the service quality of AI chatbot in frontline service. *Journal of Business Research*, *145*, 552–568. https://doi.org/10.1016/j.jbusres.2022.02.088

Cheng, Y., Qiu, L., & Tan, C.-W. (2022). Restoring Justice: The Moderating Role of AI Agent in Consumers Reactions to Service Recovery. *International Conference on Information Systems*.

Chhabra, S., Kaushal, V., & Girija, S. (2024). Determining the causes of user frustration in the case of conversational chatbots. *Behaviour & Information Technology*, 1–19. https://doi.org/10.1080/0144929X.2024.2362956

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

Dennis, A. R., & Valacich, J. S. (2001). Conducting Experimental Research in Information Systems. *Communications of the Association for Information Systems (CAIS)*, *7*, Article 5.

Do, H. J., Kong, H.-K., Tetali, P., Lee, J., & Bailey, B. P. (2023). To Err is AI: Imperfect Interventions and Repair in a Conversational Agent Facilitating Group Chat Discussions. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW1), 1–23. https://doi.org/10.1145/3579532

Dong, B., Evans, K. R., & Zou, S. (2008). The effects of customer participation in co-created service recovery. *Journal of the Academy of Marketing Science*, *36*(1), 123–137. https://doi.org/10.1007/s11747-007-0059-8

Dunn, O. J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, *56*(293), 52. https://doi.org/10.2307/2282330

Eisingerich, A. B., & Bell, S. J. (2008). Perceived Service Quality and Customer Trust. *Journal of Service Research*, *10*(3), 256–268. https://doi.org/10.1177/1094670507310769

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2023). Generative AI. *Business & Information Systems Engineering*.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems Research*, *33*(2), 678–696. https://doi.org/10.1287/isre.2021.1079

Games, P. A., & Howell, J. F. (1976). Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study. *Journal of Educational Statistics*, *1*(2), 113–125. https://doi.org/10.3102/10769986001002113

Gelbrich, K. (2010). Anger, frustration, and helplessness after service failure: Coping strategies and effective informational support. *Journal of the Academy of Marketing Science*, *38*(5), 567–585. https://doi.org/10.1007/s11747-009-0169-6

Han, E., Yin, D., & Zhang, H. (2024). When and How to Implement Choices on Customer Service Chatbots: Hilton Hawaiian Village Waikiki Beach Resort, January 3-6, 2024. *Hawaii International Conference on System Sciences*. Advance online publication. https://doi.org/107396

Haupt, M., Rozumowski, A., Freidank, J., & Haas, A. (2023). Seeking empathy or suggesting a solution? Effects of chatbot messages on service failure recovery. *Electronic Markets*, *33*(1), Article 56. https://doi.org/10.1007/s12525-023-00673-0

Le, Hoang T. P. M., Park, J., & Lee, S. (2023). Emotion and trust in virtual service assistant design for effective service recovery. *Journal of Retailing and Consumer Services*, *74*, 103368. https://doi.org/10.1016/j.jretconser.2023.103368

Hong, W., Thong, J. Y. L., Tam, K., & Yana, R. (2004). The Effects of Information Format and Shopping Task on Consumers' Online Shopping Behavior: A Cognitive Fit Perspective. *Journal of Management Information Systems*, *21*(3), 149–184. https://doi.org/10.1080/07421222.2004.11045812

Huang, M.-H., & Rust, R. T. (2020). Engaged to a Robot? The Role of AI in Service. *Journal of Service Research*. Advance online publication. https://doi.org/10.1177/1094670520902266

Janssen, A., Grützner, L., & Breitner, M. H. (2021). Why do Chatbots fail? A Critical Success Factors Analysis. *International Conference on Information Systems*.

Klein, S. H., Papies, D., & Utz, S. (2024). How Interaction Mechanism and Error Responses Influence Users' Responses to Customer Service Chatbots. *International Journal of Human–Computer Interaction*, 1–19. https://doi.org/10.1080/10447318.2024.2351707

Mariani, M., Hashemi, N., & Wirtz, J. (2023). Artificial intelligence empowered conversational agents: A systematic literature review and research agenda. *Journal of Business Research*, *161*. https://doi.org/10.1016/j.jbusres.2023.113838

Mollick, E., & Mollick, L. (2023, June 13). *Assigning AI: Seven Approaches for Students, with Prompts*. arXiv:2306.10052

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Poser, M., Singh, S., & Bittner, E. (2021). Hybrid Service Recovery: Design for Seamless Inquiry Handovers between Conversational Agents and Human Service Agents. *Hawaii International Conference on System Sciences*. Advance online publication. https://doi.org/10.24251/HICSS.2021.144

Rafaeli, A., Ziklik, L., & Doucet, L. (2008). The Impact of Call Center Employees' Customer Orientation Behaviors on Service Quality. *Journal of Service Research*, *10*(3), 239–255. https://doi.org/10.1177/1094670507306685

Schöbel, S., Schmitt, A., Benner, D., Saqr, M., Janson, A., & Leimeister, J. M. (2023). Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers. *Information Systems Frontiers*, *26*, 729–754. https://doi.org/10.1007/s10796-023-10375-9

Schuetzler, R. M., Grimes, G. M., Giboney, J. S., & Rosser, H. K. (2021). Deciding Whether and How to Deploy Chatbots. *MIS Quarterly Executive*, *20*(1), Article 4.

Sun, Y., Jang, E., Ma, F., & Wang, T. (2024). Generative AI in the Wild: Prospects, Challenges, and Strategies. In F. F. Mueller, P. Kyburz, J. R. Williamson, C. Sas, M. L. Wilson, P. T. Dugas, & I. Shklovski (Eds.), *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–16). ACM. https://doi.org/10.1145/3613904.3642160

van Vaerenbergh, Y., Larivière, B., & Vermeir, I. (2012). The Impact of Process Recovery Communication on Customer Satisfaction, Repurchase Intentions, and Word-of-Mouth Intentions. *Journal of Service Research*, *15*(3), 262–279. https://doi.org/10.1177/1094670512442786

Villarroel Ordenes, F., Packard, G., Hartmann, J., & Proserpio, D. (2025). Using Traditional Text Analysis and Large Language Models in Service Failure and Recovery. *Journal of Service Research*, Article 10946705241307678. Advance online publication. https://doi.org/10.1177/10946705241307678

Weiler, S., Matt, C., & Hess, T. (2022). Immunizing with information - Inoculation messages against conversational agents' response failures. *Electronic Markets*, *32*(1), 239–258. https://doi.org/10.1007/s12525-021-00509-9

Zeng, J., Fan, D., Zhou, X., & Tang, J [Jian] (2024). Chatbot with Resilience: The Impact of Repair Strategies on Customer Satisfaction in Conversational Breakdowns. In Y. P. Tu & M. Chi (Eds.), *Lecture Notes in Business Information Processing: Vol. 517, E-Business. New Challenges and Opportunities for Digital-Enabled Intelligent Future: 23rd Wuhan International Conference, WHICEB 2024, Wuhan, China, May 24–26, 2024, Proceedings, Part III* (1st ed. 2024, pp. 306–317). Springer Nature Switzerland; Imprint Springer. https://doi.org/10.1007/978-3-031-60324-2_26