A Preliminary Study of Hybrid Intelligence, Social Listening, and Chatbot Feedback

Janet RAFNER[a,1] , Ryan Q. GULOY [b], Eden W. WEN [b], Catherine M. CHIODO [b] and Jacob SHERSON[c]
[a] *University of Southern Denmark*
[b] *Adobe Inc.*
[c] *Center for Hybrid Intelligence, Aarhus University*

Abstract. Generative AI (GenAI) assistants are increasingly used in workplace and customer-facing contexts, yet their feedback mechanisms often rely on low-bandwidth signals such as thumbs up/down that offer limited actionable detail. We examine this design problem through the combined lenses of social listening and hybrid intelligence (HI). From a social listening perspective, feedback is not only an evaluative input but also a form of bi-directional communication that signals whether a system is attentive and responsive to its users. From an HI perspective, feedback can be framed as a collaborative contribution to iterative system refinement rather than as a simple end-user rating. We report a preliminary study comparing two feedback designs in a GenAI marketing-assistant prototype: (1) a conventional icon-based feedback affordance with optional comments, and (2) an HI-framed in-chat solicitation that explains how user feedback will be used to improve the system. Marketing professionals recruited via Maze.co completed 19 sessions. Likert-scale measures showed no significant differences in trust, willingness to provide feedback, or intended use. HI sessions produced significantly longer open-ended responses, and qualitative analysis suggests that the HI design reduced interpretation cost and made the feedback interaction feel more visible and collaborative. These preliminary findings suggest that designing feedback as social listening may improve the richness of user input while also surfacing trade-offs between low-friction interaction and expressive nuance.

Keywords. chatbots, GenAI, Hybrid Intelligence, organizational partnership, feedback, social listening

1. Introduction

GenAI assistants a common interface for workplace and customer-facing support, yet the mechanisms used to collect user feedback often lag behind the complexity of these systems. In many products, feedback is reduced to low- effort signals such as thumbs up/down, which are easy to deploy but difficult to translate into actionable system changes [1]. This creates a mismatch: GenAI systems benefit most from feedback that is specific, contextual, and actionable, but common feedback affordances often yield only limited diagnostic detail. User-feedback taxonomies show that evaluative inputs such as

[1] Corresponding Author: Janet Rafner, jraf@sam.sdu.dk.

ratings capture only a narrow slice of users' intent and may under-support reporting and requesting behaviors that are more useful for iterative improvement [2]. In conversational systems, feedback affordances can therefore shape not only whether users respond, but also what kind of feedback they produce.

We approach this problem through the lens of social listening and bi-directional communication. In this view, feedback is not merely a usability metric or post hoc evaluation signal, but a relational practice that communicates whether an organization is receptive and responsive [3,4,5]. Prior work on consumer–brand relationships and customer engagement argues that perceived responsiveness and interaction quality contribute to stronger long-term relationships [6,7,8]. Scholarship on organizational agents similarly shows that conversational systems can shape users' perceptions of attentiveness, responsiveness, and relational quality [9,10]. Organizational partnership research frames listening as inseparable from responsiveness and co-innovation, where stakeholder input becomes a resource for adaptation [11]. For GenAI systems, this means that feedback mechanisms are not only data-collection tools but also listening interfaces: design choices that make feedback visible, interpretable, and credibly used may support bi-directionality while improving the quality of input available for iterative refinement.

Recent HI work extends these ideas to GenAI assistants, emphasizing psychological safety, user involvement, and iterative refinement as key to effective human–AI collaboration [14,15]. We treat feedback as a site where HI can be operationalized: the prompt and explanatory framing can position the user either as an end evaluator of an AI tool or as a collaborator helping improve the system. We compare two feedback interactions in a GenAI assistant prototype for marketing tasks: (1) a conventional icon-based feedback affordance, and (2) an HI-framed in-chat solicitation that explains how user input will shape future system improvements. We ask: How do framing (AI tool vs. HI partnership) and feedback interaction design (embedded icons vs. in-chat solicitation) influence (a) feedback usability/discoverability and (b) the richness and actionability of the feedback users provide?

2. Preliminary Study Methods

We compared two interface conditions in a GenAI marketing-assistant prototype. Participants provided informed consent and completed a simulated workflow in which the assistant generated an email campaign and the participant then provided feedback through the interface assigned to that session. In the Control condition (C-C), feedback was collected via embedded thumbs up/down icons within assistant responses, with optional written comments submitted through a single feedback form. In the HI condition (HI-C), feedback was prompted through an in-chat message that also explained how the feedback would be used to improve the system. The HI-C was intended to foster a sense of partnership between the user and the AI assistant by making the improvement loop explicit. The study ran from December 23, 2024, to January 27, 2025. Participants were marketing professionals recruited via Maze.co. In total, we collected 19 sessions (8-C, 11-HI) from 14 unique participants; five participants completed both conditions. Participants were English-speaking, and their roles spanned common marketing positions such as marketer, copywriter, director, and brand manager. Maze.co did not provide demographic variables such as age, gender, or education.

3. Preliminary Measurements, Analysis and Results

Quantitative measures included Likert-scale questions assessing willingness to provide feedback, trust in the system, and satisfaction with the interaction. Descriptive statistics summarized survey results, and t-tests for independent samples were applied to compare conditions, with significance thresholds set at $p < 0.05$. Exploratory session-level *t*-tests showed no clear differences on survey outcomes and are interpreted descriptively given the small, partially repeated sample. The main quantitative difference emerged in the open-ended responses. Across comparable items, at the session level HI responses (averaged 16.12 words) were longer than Control responses (averaged 11.81 words). A mixed-effects model including participant as a random intercept showed a significant increase in response length in the HI-C (HI – Control = +4.32 words, SE = 1.86, 95% CI [0.67, 7.96], p = .020). A within-subject check on the 5 overlapping participants showed the same direction of effect, though significance varied with test choice, indicating sensitivity to the small paired sample.

For qualitative analysis, we applied a deductive coding scheme adapted from a user-feedback taxonomy [2], distinguishing topics, user experience, and intention. Then, we conducted thematic analysis to capture higher-order tensions in participants' reasoning, including interpretation cost, interruption, expressivity, and reciprocity. Qualitative feedback was categorized into Product Quality (e.g., "It needs improvement"; C-C), Product Context (e.g., "I'd need to see the quality" - C-C), Quality in Use (e.g., "It was straightforward"; HI-C), and Feedback Intention, including Informing (e.g., "I'd be more willing if answers were high quality"; C-C), Reporting ("The system froze" - HI-C), and Requesting ("I wish it said what each thumb meant" - C-C). User Experience was the most common theme, with HI-C participants noting intuitive feedback (e.g., "It was straightforward") and C-C participants reporting navigation difficulties (e.g., "I had to guess where to go"). In the C-C, several participants described uncertainty about where feedback "should" be in the interface, often relying on familiar UI conventions or guesswork. This suggests that discoverability was shaped by users' prior mental models and that misalignment with those expectations increased effort and confusion. In the HI-C, participants more often described the feedback prompt as visible and interpretable because it was embedded directly in the conversational flow. HI participants also noted opportunities for clearer labeling, suggesting that visibility alone is not enough.

4. Discussion, Limitations, and Future Work

The HI-C elicited longer open-ended responses and appeared to reduce interpretation cost by making the feedback mechanism more visible, legible, and collaborative. From an HI perspective, the explanatory in-chat prompt positioned users less as passive evaluators and more as contributors to iterative refinement [16]. The findings should be interpreted cautiously. The study is small-scale, exploratory, and underpowered, and the two conditions vary simultaneously in both feedback modality and explanatory framing, making it difficult to isolate the specific causal source of the observed effects. The sample was limited to marketing professionals recruited through Maze.co, which constrains generalizability. Future work should test these mechanisms in larger samples, disentangle framing from interface modality, and examine whether similar patterns hold across other conversational contexts and feedback tasks.

References

[1] Ghassempour S, Pershukova A, Gokgoz E, Nikulina I. Beyond thumbs up and thumbs down: a human-centered approach to evaluation design for LLM products [Internet]. Medium; 2025 Mar 4. Available from: https://medium.com/datascience- at-microsoft/beyond-thumbs-up-and-thumbs-down-a-human-centered-approach-to-evaluation-design-for-llmproducts- d2df5c821da5

[2] Santos R, Groen EC, Villela K. A taxonomy for user feedback classifications. In: REFSQ Workshops, Vol. 2376; 2019.

[3] Duncan T, Moriarty SE. A communication-based marketing model for managing relationships. J Mark. 1998;62(2):1-13

[4] Rosenblatt M, Curran T, Treiber J. Building brands through social listening. 2018.

[5] Zou Y. Social media review: the impact of social on brand-consumer relationships. In: Ideas in Marketing: Finding the New and Polishing the Old: Proceedings of the 2013 Academy of Marketing Science (AMS) Annual Conference. Springer; 2015. p. 624.

[6] Bowden JL. The process of customer engagement: a conceptual framework. J Mark Theory Pract. 2009;17:63-74.

[7] Fournier S. Consumers and their brands: developing relationship theory in consumer research. J Consum Res. 1998;24:343-73.

[8] Freeman K, Spenner P, Bird A. Three myths about what customers want [Internet]. Harvard Business Review; 2012 May 23. Available from: https://hbr.org/2012/05/three-myths-about-customer-eng

[9] Men LR, Zhou A, Sunny Tsai WH. Harnessing the power of chatbot social conversation for organizational listening: The impact on perceived transparency and organization-public relationships. Journal of Public Relations Research. 2022 Mar 4;34(1-2):20-44.

[10] Zhou A, Tsai WHS, Men LR. Optimizing AI social chatbots for relational outcomes: the effects of profile design, communication strategies, and message framing. Int J Bus Commun. 2024. doi: 10.1177/23294884241229223. [11] Desouza KC, Awazu Y, Jha S, Dombrowski C, Papagari S, Baloh P, et al. Customer-driven innovation. Res Tec

[11] Desouza KC, Awazu Y, Jha S, Dombrowski C, Papagari S, Baloh P, et al. Customer-driven innovation. Res Technol Manage. 2008;51(3):35-44.

[12] Akata Z, Balliet D, De Rijke M, Dignum F, Dignum V, Eiben G, et al. A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. Computer. 2020;53(8):18-28.

[13] Dellermann D, Ebel P, Söllner M, Leimeister JM. Hybrid intelligence. Bus Inf Syst Eng. 2019.

[14] Mao Y, Rafner J, Wang Y, Sherson J. HI-TAM, a hybrid intelligence framework for training and adoption of generative design assistants. Front Comput Sci. 2024;6:1460381.

[15] Sherson J, Rafner J, Büyükgüzel S. Operational criteria of hybrid intelligence for generative AI virtual assistants. In: HHAI 2024: Hybrid Human AI Systems for the Social Good. Amsterdam: IOS Press; 2024. p. 475-7.

[16] Sherson J, Vinchon F. Facilitating human feedback for GenAI prompt optimization. In: HHAI 2024: Hybrid Human-AI Systems for the Social Good (Frontiers in Artificial Intelligence and Applications, Vol. 386). Amsterdam: IOS Press; 2024. p. 478-80. doi: 10.3233/FAIA240230.