Visual Cues in Wrong-Path Interactions: An Exploratory Study of User Experience in Conversational AI

Jie Bai[1,*], Alessandra Zarcone[1]

[1] *Technische Hochschule Augsburg, Augsburg, Germany*

Abstract

Conversational AI agents are widely used in customer service, where wrong-path interactions often occur without clear signals of the agent's limitations. In such situations, users may lose trust and feel frustrated. Prior research has largely emphasized verbal strategies for handling wrong-path interactions, while the role of non-verbal visual cues in shaping user experience remains less explored. This paper investigates how non-verbal visual cues influence user experience during wrong-path interactions in an e-commerce scenario. We conducted a 2×2 experimental study with 32 participants. A running prototype was used, with and without visual cues such as avatars, emojis, and interactive buttons. We evaluated task efficiency (time, error rate) and emotional comfort using objective measures and the Chatbot Usability Questionnaire (CUQ). The initial results suggest that non-verbal visual cues can support task efficiency. However, they do not necessarily improve emotional comfort and may increase frustration. This study provides preliminary insights for HCI researchers, UX designers, and conversational designers, and points to opportunities for improving future experimental designs.

Keywords

Conversational interfaces, Wrong-path interactions, Non-verbal visual cues, User experience

1. Introduction

Conversational AI agents (CAs), commonly referred to as chatbots, are widely used in customer service, especially on e-commerce platforms, where they support tasks such as product search, order tracking, and customer inquiries [1, 2]. Their adoption is mainly driven by the potential to reduce operational costs and improve service efficiency [3]. When tasks cannot be successfully completed, conversational failures may occur, particularly in cases where the agent remains operational but progress is blocked. Such failures have been associated with user frustration and other negative emotional responses in customer service chatbot interactions [3, 4].

Prior research on conversational AI user experience has largely focused on verbal interaction, including dialogue quality, response style, and evaluation methods [5, 6, 7]. While these studies provide important insights into how users communicate with conversational agents, they offer limited understanding of how interface-level design shapes user experience when interactions do not proceed as expected.

Beyond language, conversational interfaces increasingly incorporate non-verbal visual cues such as emojis, avatars, and interactive buttons. Existing work suggests that these cues can support emotional expression, reduce ambiguity, and enhance social presence in conversational interactions [8, 9]. However, how such visual cues influence user experience in situations where task progress is disrupted remains insufficiently explored, particularly in goal-oriented e-commerce contexts.

To address this gap, this study examines how non-verbal visual cues affect user experience during failure scenarios in conversational AI agents. In this work, we refer to such situations as wrong-path interactions, where the system remains operational and the task can eventually be completed, but progress is temporarily blocked without explicit explanation. We focus on how visual interface design shapes task efficiency and emotional comfort under these conditions.

Joint Proceedings of the ACM Intelligent User Interfaces (IUI) Workshops 2026, March 23-26, 2026, Paphos, Cyprus

*Corresponding author.

†These authors contributed equally.

baijie.creates@gmail.com (J. Bai); alessandra.zarcone@tha.de (A. Zarcone)

2. Related Work

Prior research has shown that conversational AI agents often fail to complete user goals due to limited understanding, constrained system capabilities [10]. Such failures have been examined and shown to influence user trust and emotional responses in customer service contexts [11]. Expectation mismatch has been discussed as a potential mechanism underlying these effects, where users overestimate system capability and experience frustration when tasks cannot be completed [12]. These findings indicate that task failure is a common and consequential aspect of conversational agent interactions.

In response to these challenges, some studies focus on explicit failure handling and repair strategies, such as clarification prompts and error explanations, which aim to make system limitations explicit [10, 13]. In practice, however, these agents may fail silently or provide limited feedback, leaving users to infer system capabilities without clear guidance.

In parallel, non-verbal visual cues such as avatars, emojis, and interactive buttons have been shown to influence user perceptions of social presence, engagement, and emotional comfort in conversational interfaces [14, 15, 9]. However, prior work has primarily examined these cues in successful or neutral interactions, often relying on static or scripted designs, with limited attention to their roles during unresolved wrong-path interactions where task progress is silently blocked. This study addresses this gap by examining how non-verbal visual cues shape user experience during wrong-path interactions.

3. Study Design

3.1. Experimental Design

We conducted a 2×2 between-subjects experiment to examine the effects of non-verbal visual cues during failure scenarios. The experiment manipulated two factors: visual interaction cues (V: present vs. O: absent) and visual presence cues (A: robotic avatar vs. B: brand logo). This resulted in four conditions: VA, VB, OA, and the baseline OB. As shown in Figure 1, the first letter of each condition indicates the interaction cue, and the second indicates the presence cue. To avoid learning effects, each participant interacted with only one condition.

Figure 2 provides a concrete illustration of the interface design, contrasting the two extremes: VA (full visual cues) and OB (text-only baseline). To isolate the effects of these cues, the verbal content and conversational flow were kept strictly consistent across all conditions. Participants performed two e-commerce tasks, product tracking and coupon redemption, with a wrong-path interaction strategically embedded in the latter.

3.2. Wrong-Path Interaction Scenario

The study included a wrong-path interaction embedded in a coupon redemption task. Participants were initially provided with an incorrect coupon code and attempted to redeem it twice. During this process, a silent waiting period (5–10 seconds) was introduced. The correct coupon code was eventually presented, allowing the task to be completed.

Throughout the interaction, the agent did not explicitly signal any error, uncertainty, or limitation. Participants were therefore unaware that they had entered a wrong path and relied on the conversational flow and interface-level cues to interpret the system's behavior. The same wrong-path scenario was used across all conditions, with only the presence and configuration of non-verbal visual cues varying. Figure 3 illustrates the interaction flow used in the study.

3.3. Participants and Procedure

A total of $N = 34$ participants were recruited and randomly assigned to one of the four experimental conditions. Two responses from the VB condition were not included in the final analysis to ensure balanced group sizes across conditions, resulting in a final dataset of $N = 32$ ($n = 8$ per condition). The sample included 19 female and 13 male participants, with ages ranging from 18–24 to 45–54 years.

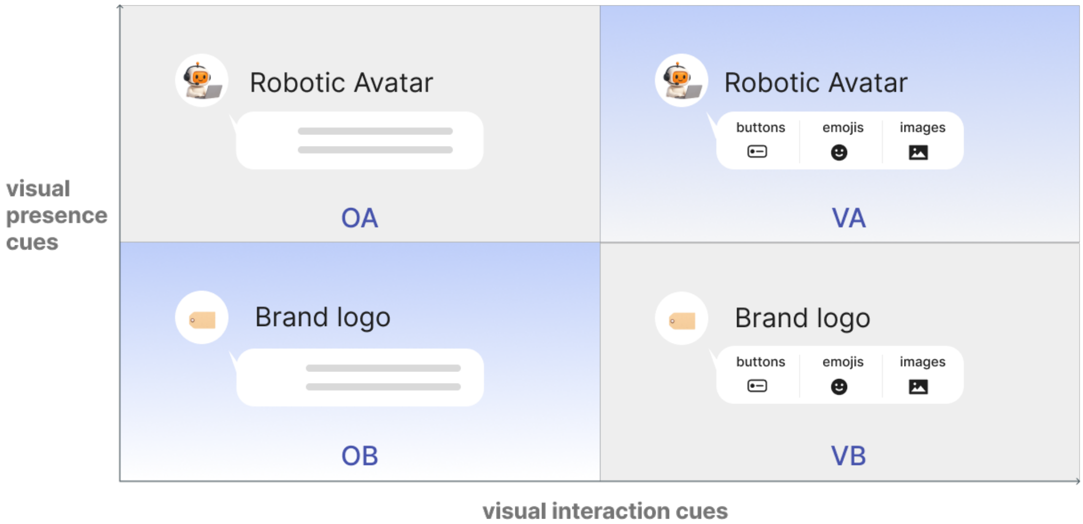

Figure 1: The 2 × 2 experimental design matrix showing the interface prototypes. Columns represent the presence (V) or absence (O) of visual interaction cues (e.g., emojis, buttons); rows represent the type of visual presence cue (A: Robotic Avatar, B: Brand logo).

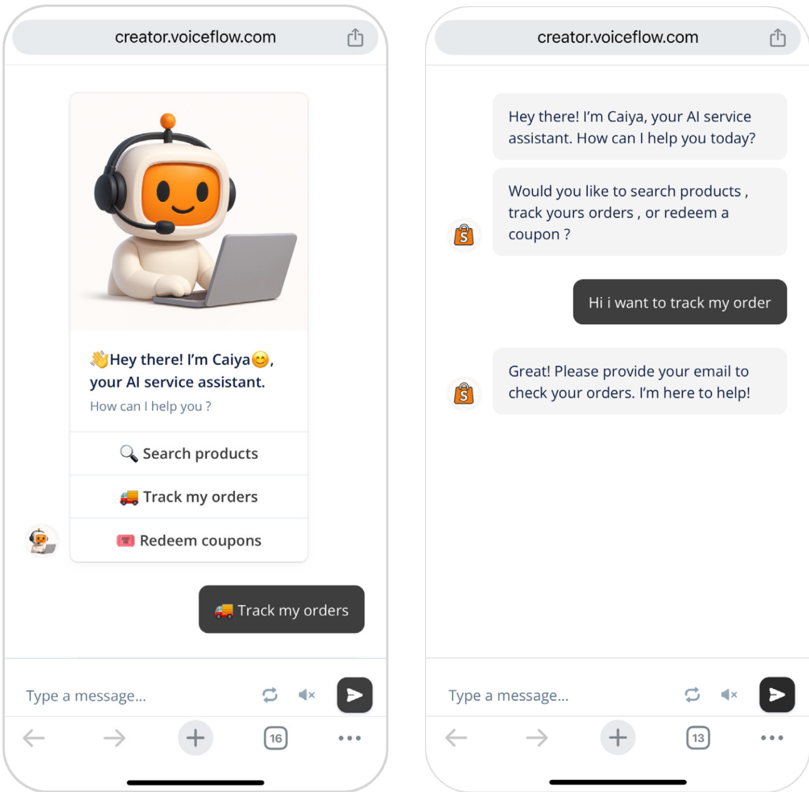

Figure 2: Prototype examples. Left: VA condition (full visual cues); Right: OB condition (text-only baseline).

Participants were recruited via in-person intercept sampling on campus (e.g., library and cafeteria). They completed the study on a device provided by the researchers (iPhone 15), and screen recordings were captured for subsequent analysis, ensuring consistent viewport and interaction constraints across participants.

Task order (product tracking vs. coupon redemption) was randomized to reduce ordering effects. The interaction lasted approximately 8 minutes in total. After completing the tasks, participants filled out a post-task questionnaire assessing their experience and provided brief qualitative comments.

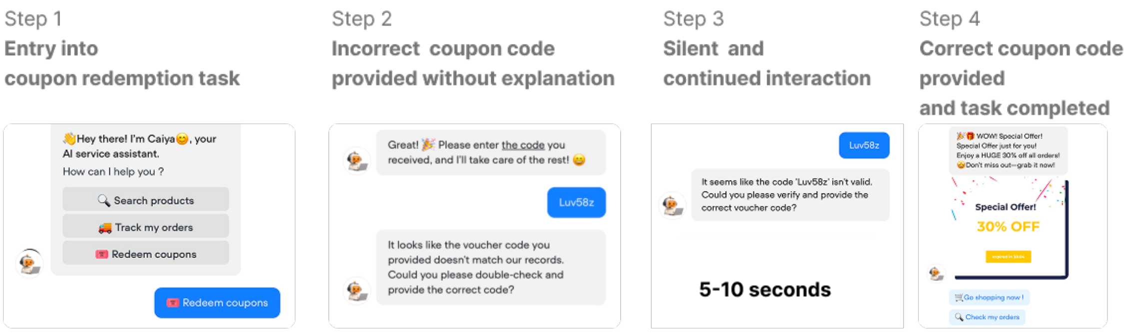

Figure 3: Wrong-path interaction flow.

3.4. Measurement

User experience was evaluated across task efficiency and emotional comfort. Task efficiency was measured using task completion time and error rates, capturing how participants navigated the interaction despite temporarily blocked task progress.

Emotional comfort was assessed using selected dimensions of the Chatbot Usability Questionnaire (CUQ) [16], focusing on trust and welcoming perception. In addition, participants were asked whether they felt frustrated during the interaction and to briefly describe the reason.

We report descriptive statistics (mean and SD) for each condition. Given the exploratory nature of the study and the sample size, non-parametric tests were used: Kruskal–Wallis tests followed by Dunn's post-hoc comparisons (Bonferroni correction) for completion time and error counts, and Fisher's exact test for binary yes/no frustration responses. For welcoming perception, we additionally compared avatar vs. non-avatar conditions by pooling groups (OA+VA vs. OB+VB).

4. Results

Table 1 reports descriptive statistics across conditions, while Table 2 summarizes the observed roles of visual cues.

4.1. Task efficiency

Overall, conditions with visual interaction cues (VA/VB) showed higher task efficiency than the text-only baseline (OA/OB). Participants in VA and VB completed the tasks faster than OA and OB, with mean completion times of 137.0s (SD = 44.6) and 139.0s (SD = 54.2), compared to 175.0s (SD = 26.6) and 188.0s (SD = 26.9), corresponding to an approximate 24% reduction in task duration. A Kruskal–Wallis test indicated significant differences in completion time across the four conditions ($p = 0.009$). Post-hoc pairwise comparisons (Dunn's test with Bonferroni correction) showed a significant difference between OB and VA ($p = 0.039$), while other comparisons did not reach significance.

Error counts followed a similar pattern. Participants in VA (M = 1.25, SD = 1.04) and VB (M = 1.38, SD = 1.19) made substantially fewer errors than those in OB (M = 4.00, SD = 2.39), representing reductions of 68.8% and 65.5%, respectively. A Kruskal–Wallis test found significant differences across conditions ($p = 0.011$). Dunn's post-hoc tests revealed that OB had significantly more errors than VA ($p = 0.030$) and VB ($p = 0.034$).

4.2. Emotional comfort

In contrast, visual interaction cues did not improve emotional comfort. Trust ratings were generally high across all conditions (OA: M = 4.00, SD = 0.76; OB: M = 3.50, SD = 0.93; VA: M = 3.75, SD = 0.71;

Table 1
Descriptive statistics (mean and SD) across the four prototype conditions ($N = 32$, $n = 8$ per condition). Conditions follow the naming rule: Interaction cue (V/O) and Presence cue (A/B).

Measure	OA	OB	VA	VB
Completion time (s)	175.0 (26.6)	188.0 (26.9)	137.0 (44.6)	139.0 (54.2)
Errors (count)	2.75 (1.83)	4.00 (2.39)	1.25 (1.04)	1.38 (1.19)
Trust (1–5)	4.00 (0.76)	3.50 (0.93)	3.75 (0.71)	4.12 (0.64)
Welcoming (1–5)	4.75 (0.46)	3.88 (0.99)	4.62 (0.52)	4.00 (0.93)

VB: M = 4.12, SD = 0.64), and no significant differences were observed ($p > 0.05$). Reported frustration also did not significantly differ across conditions (Fisher's exact test, $p = 0.524$), although VA showed the highest proportion of participants reporting frustration.

Visual presence cues (robotic avatar) did not lead to observable improvements in task efficiency. However, avatar presence was associated with higher perceived welcoming ratings (Avatar: M = 4.69, SD = 0.48; Non-avatar: M = 3.94, SD = 0.93), with a significant difference ($p = 0.008$). Given the small sample size ($N = 32$), these findings are interpreted as exploratory.

4.3. Uncertainty as a source of frustration in wrong-path interactions

Qualitative feedback suggested that frustration primarily stemmed from *uncertainty about the cause of failure* during wrong-path interactions. Several participants reported frustration when repeated input attempts were required without explanation, such as: *"Yes, when I entered the code twice and got a negative response."* and *"When it just said it's wrong, it should give further suggestions."* Others described attempting to infer system rules in the absence of feedback, for example: *"I was unsure if my error with the coupon was that I didn't capitalize all letters correctly."* These responses indicate that frustration was not driven by task difficulty itself, but by missing system feedback explaining why progress was blocked and what users should do next.

Table 2
Summary of observed roles of visual cues on efficiency and comfort during wrong-path interactions. Statistical significance is derived from results in Table 1.

Category	Cue Type	Task Efficiency (Time/Err)	Emotional Comfort	Frustr.
No Visuals (OB)	Text-only baseline	Baseline	Neutral / Lower	Baseline
Interaction (VB)	Emojis & Buttons	**Sig. Improved**	No sig. change	**Increased**
Presence (OA)	Robotic Avatar	No observable effect	**Sig. Welcoming**	No sig. change
Combined (VA)	Full Cues (VA)	**Highest Efficiency**	**High Welcoming**	**Highest Reported**

Note: "Improved" and "Higher" indicate shifts relative to OB baseline. Statistical significance ($p < 0.05$) follows Kruskal-Wallis and post-hoc Dunn's tests.

5. Discussion

Given the relatively small sample size ($N = 32$), the findings should be interpreted as exploratory. Nevertheless, the results point to three key observations regarding the role of visual cues in wrong-path conversational interactions, particularly in how they shape users' uncertainty and expectations during stalled progress.

First, visual interaction cues improved task efficiency but also appeared to increase users' implicit expectations about system capability. In wrong-path interactions, where failures persisted without clear explanation, these elevated expectations were associated with greater frustration. This suggests that efficiency-supporting cues may unintentionally intensify uncertainty and disappointment when progress stalls under silently unresolved failure conditions.

Second, visual presence cues in the form of a robotic avatar contributed to a more positive initial impression, as reflected in higher welcoming perceptions. However, avatar presence did not reduce frustration when unexpected or unresolved errors occurred. This indicates that presence-related cues may shape first impressions but offer limited emotional comfort when users face uncertainty about why the interaction is blocked.

Third, the findings suggest that in scenarios involving unknown failures, conversational AI agents may benefit from more explicit visual feedback to communicate system status and manage users' uncertainty (e.g., indicating that the system is processing, recognizing an issue, or providing clearer recovery options). Without such guidance, visual cues may signal progress without enabling resolution, which can increase user frustration even when tasks remain ultimately solvable.

6. Limitations and Future Work

This study has several limitations that should be considered when interpreting the findings. First, the sample size was relatively small, which may limit the generalizability of the results. In addition, participants' mobile device preferences (e.g., operating systems) and age distribution were not evenly balanced, which may have influenced interaction patterns and subjective perceptions. These factors may have introduced variability into the data and should be taken into account when drawing conclusions. At the same time, these limitations highlight opportunities for more systematic investigation in future research.

Building on these limitations, future studies could expand the participant pool to include a larger and more diverse population, allowing for more robust statistical analysis and improved generalizability. In addition, future research could further isolate individual design factors by systematically varying specific visual cues (e.g., avatars, emojis, and interaction buttons) and increasing the number of prototype conditions. Such an approach would enable a more fine-grained examination of how different visual cues contribute to user expectations, task efficiency, and emotional responses during wrong-path interactions.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT and Gemini in order to: Grammar and spelling check, paragraph and reword, improve writing style. After using the tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] E. Adamopoulou, L. Moussiades, An overview of chatbot technology, in: I. Maglogiannis, L. Iliadis, E. Pimenidis (Eds.), Artificial Intelligence Applications and Innovations. AIAI 2020, volume 584 of *IFIP Advances in Information and Communication Technology*, Springer, Cham, 2020. doi:10.1007/978-3-030-49186-4_31.

[2] A. Følstad, M. Skjuve, Chatbots for customer service: user experience and motivation, ACM Conference Proceedings (2019). doi:10.1145/3342775.3342784.

[3] R. Budiu, The user experience of chatbots, 2018. Available at: https://www.nngroup.com/articles/chatbots/.

[4] A. Zafar, V. Balavadhani, C. Le, S. Shahid, A. Iqbal, A. Shahid, Building trust in conversational AI: A comprehensive review and solution architecture for explainable, privacy-aware systems using llms and knowledge graph, 2023. doi:10.36227/techrxiv.24009351.

[5] R. J. Moore, M. H. Szymanski, R. Arar, G.-J. Ren (Eds.), Studies in Conversational UX Design, volume 157 of *Human-Computer Interaction Series*, Springer International Publishing AG, Cham, Switzerland, 2018. doi:10.1007/978-3-319-95579-7.

[6] A. B. Kocaballi, L. Laranjo, E. Coiera, Measuring user experience in conversational interfaces: a comparison of six questionnaires, in: Proceedings of the 32nd International BCS Human Computer Interaction Conference, HCI '18, BCS Learning & Development Ltd., Swindon, GBR, 2018. doi:10.14236/ewic/HCI2018.21.

[7] S. Holmes, A. Moorhead, R. Bond, H. Zheng, V. Coates, M. McTear, Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces?, in: Proceedings of the 31st European Conference on Cognitive Ergonomics, ECCE '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 207–214. doi:10.1145/3335082.3335094.

[8] S.-Y. Lee, G. Lee, S. Kim, J. Lee, Expressing personalities of conversational agents through visual and verbal feedback, Electronics 8 (2019) 794. doi:10.3390/electronics8070794.

[9] H. Meng, X. Lu, J. Xu, The impact of chatbot response strategies and emojis usage on customers' purchase intention: The mediating roles of psychological distance and performance expectancy, Behavioral Sciences 15 (2025) 117. doi:10.3390/bs15020117.

[10] H. J. Do, H.-K. Kong, P. Tetali, J. Lee, B. P. Bailey, To Err is AI: Imperfect Interventions and Repair in a Conversational Agent Facilitating Group Chat Discussions, Proc. ACM Hum.-Comput. Interact. 7 (2023). doi:10.1145/3579532.

[11] E. L.-C. Law, A. FØLstad, N. Van As, Effects of humanlikeness and conversational breakdown on trust in chatbots for customer service, in: Nordic Human-Computer Interaction Conference, NordiCHI '22, Association for Computing Machinery, New York, NY, USA, 2022. doi:10.1145/3546155.3546665.

[12] D. Saxena, J.-Y. Jung, J. Forlizzi, K. Holstein, J. Zimmerman, Ai mismatches: Identifying potential algorithmic harms before ai development, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25, Association for Computing Machinery, New York, NY, USA, 2025. doi:10.1145/3706598.3714098.

[13] S. Dey, Y.-J. Sun, G. Tur, D. Hakkani-Tur, Know your mistakes: Towards preventing overreliance on task-oriented conversational ai through accountability modeling, 2025. URL: https://arxiv.org/abs/2501.10316. arXiv:2501.10316.

[14] E. Go, S. S. Sundar, Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions, Computers in Human Behavior 97 (2019) 304–316. doi:10.1016/j.chb.2019.01.020.

[15] L. M. Assink, Exploring users' perception of chatbots in a mobile commerce environment : Creating a better user experience by implementing anthropomorphic visual and linguistic chatbot features, 2019. URL: http://essay.utwente.nl/78306/.

[16] S. Holmes, A. Moorhead, R. Bond, H. Zheng, V. Coates, M. McTear, The chatbot usability questionnaire: Development and validation, in: Proceedings of the 31st European Conference on Cognitive Ergonomics (ECCE '19), Association for Computing Machinery, New York, NY, USA, 2019, pp. 81–88. doi:10.1145/3335082.3335094.