

# Designing an AI Coach to Scaffold Self-Reflection in Customer Service Conversations

Kazuhiro Shidara  
Fujitsu Limited  
Kanagawa, Japan  
s-kazuhiro@fujitsu.com

Sho Iwasaki  
Fujitsu Limited  
Kanagawa, Japan  
iwasaki.sho@fujitsu.com

Toi Hirakawa  
Fujitsu Limited  
Kanagawa, Japan  
hirakawa.toi@fujitsu.com

Takahiro Yoshioka  
Fujitsu Limited  
Kanagawa, Japan  
y.takahiro@fujitsu.com

Masayuki Kiri  
Toyo University  
Tokyo, Japan  
kiri@toyo.jp

Takeshi Konno  
Fujitsu Limited  
Kanagawa, Japan  
konno.takeshi@fujitsu.com

## Abstract

In customer-facing operations such as call centers, advanced interpersonal skills are required not only to resolve issues but also to build rapport, a trust-based relationship that facilitates smooth interactions with customers. Because the success or failure of such skills hinges on subtle differences in context and expression, they are difficult to standardize uniformly through manuals or model examples, and simply presenting methods via AI has proven insufficient for skill acquisition. To address this challenge, this study positions generative AI not as a provider of correct answers but as a coach that supports reflection. We propose a self-reflection tool, Reflection AI Coach, that uses real customer interaction logs to guide operators through a three-stage reflection process: organizing situational judgments, constructing alternative utterances, and consolidating behavioral guidelines. As a preliminary investigation, we conducted a pilot study with two customer support practitioners and analyzed their experiences and challenges through semi-structured interviews. The results indicated that reflection aimed at organizing situations and judgments was perceived as useful; however, they also suggested that improvements in experience design are necessary to sustain reflection as an ongoing practice. Based on these findings, we discuss the design of human-AI interaction for supporting reflection and outline directions for future work.

## CCS Concepts

• **Human-centered computing** → **Interactive systems and tools.**

## Keywords

self-reflection; design and usage; AI coach; user interface; rapport; customer service

## ACM Reference Format:

Kazuhiro Shidara, Sho Iwasaki, Toi Hirakawa, Takahiro Yoshioka, Masayuki Kiri, and Takeshi Konno. 2026. Designing an AI Coach to Scaffold Self-Reflection in Customer Service Conversations. In *Proceedings of Tools for Thought Workshop*. ACM, New York, NY, USA, 5 pages.

## 1 Introduction

Interpersonal skills such as building trust and responding empathetically in dialogue are practiced as judgments and choices of expression adapted to the other party's state and context. A foundational element of such skills is rapport—a sense of trust that arises when the other party feels their circumstances and intentions are understood [9, 18]. However, general forms of training and guidance (e.g., manuals, exemplar responses, and abstract advice) do not always map cleanly to the moment-by-moment conversational cues in real interactions, which can hinder on-the-spot situational judgment even when learners later encounter a similar situation again [10, 16]. Consequently, learners need support for revisiting situations they could not handle well, together with conversational cues, and reconstructing the rationale for their judgments and actions in their own words. This challenge is particularly pronounced in customer service operations such as call centers [3, 7, 11], where agents must advance problem resolution within limited time while simultaneously building rapport.

This study addresses the problem that conceptual understanding of interpersonal skills often does not translate into practice. Specifically, two issues arise. First, it is difficult to identify which utterances in a conversation history are important and to read the other party's emotions and concerns from them [11]; abstract advice alone does not yield situation-specific judgments grounded in individual utterances and their context. Second, even when situational judgment is achieved, expressing it as a concrete utterance on the spot remains challenging [10, 16], as appropriate words may not be immediately recalled and learning tends to stay tied to specific cases, making it difficult to apply in similar situations.

We therefore propose **Reflection AI Coach**, a system that supports users in articulating the rationale for their situational judgments and organizing alternative utterances and action guidelines while referencing conversation histories. Unlike in-situ assistance that suggests responses during a dialogue, our focus is post-dialogue learning: reconstructing the rationale behind situational judgments from concrete conversational cues and externalizing it into reusable utterances and guidelines. We position generative AI

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*Tools for Thought Workshop, Barcelona, Spain*

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

not as a provider of correct answers but as a coach. In education, scaffolding is a well-established framework for guided learning, characterized by contingency, fading, and transfer of responsibility [19]. As a first step toward such adaptive scaffolding, this study focuses on verifying a full set of coaching functions before introducing individualized adaptation such as contingent adjustment and fading. Specifically, we adopt the design hypothesis that users’ judgment and articulation advance when the AI provides (i) selection of scenes for reflection, (ii) presentation of rapport-related behavioral perspectives, (iii) guidance through the reflection procedure, and (iv) feedback on written content. As a prototype, we integrated AI coach support into two web application functions: **reflection mode**, in which users organize the rationale for their judgments based on conversation histories, and **rehearsal mode**, in which users apply organized content to similar situations.

This paper focuses on organizing judgment rationale and articulating it as expressions usable in practice, a process difficult to address through summarization or advice alone. As a key contribution, we propose a three-stage reflection framework for this process and explore design principles for the AI coach’s presentation, questioning, and feedback that guide the connection from conversational evidence to transferable behavioral guidelines. Thus, Reflection AI Coach is positioned as a tool for thought that externalizes and scaffolds operators’ sensemaking and judgment reconstruction, rather than optimizing task efficiency or response correctness.

## 2 Related Work

### 2.1 Technical Support for Rapport Formation

Technologies have been proposed that treat rapport as a factor related to conversational smoothness and support relationship building through response generation and behavioral adjustment. For example, conversational agents have been used as facilitators of user-to-user interaction by providing topics and conversational flow [17], and AI-mediated communication that supports response expression through suggestions and rewriting has been reported [8]. Other work includes support for behavioral adjustment through nonverbal cue presentation [6], control design for human–AI collaboration in high-risk call environments [2], and field evaluations of how AI support is perceived by operators [15].

However, these studies primarily focus on in-situ assistance for behavior and responses during dialogue and do not sufficiently address the post-dialogue learning process in which users reconstruct their judgments and expressions to apply them in subsequent situations.

### 2.2 Reflection Support Systems

Various forms of support have been proposed to help people review work-related exchanges such as conversations and meetings after the fact, presenting records in accessible formats to facilitate learning. Examples include annotation for reviewing key moments during role-play afterward [4], support for clarifying reflection perspectives and goals [5], collaborative design of meeting minutes generation [1], and visualization-based support for grasping the overall flow of activities [13, 20].

While such support is effective for prompting awareness and eliciting descriptions, it often does not explicitly design for connecting insights to judgments and utterances usable in subsequent interactions. Record-presentation-based support concentrates the burden of interpretation on users, which can make reflection superficial [13, 20]. Moreover, the more directly generative AI summaries or standardized prompts present conclusions, the more users’ own reconstruction of judgment rationale may be weakened [12, 14, 21].

Building on these insights, while prior work has supported in-situ conversational assistance and retrospective reflection, frameworks that explicitly scaffold the reconstruction and transfer of situated professional judgment remain relatively underexplored. We therefore propose Reflection AI Coach, which structures reflection as a process of articulating judgment grounded in conversational evidence, abstracting it into reusable guidelines, and rehearsing its transfer to similar situations.

## 3 System Design

This section describes the structure of Reflection AI Coach, the points at which the AI coach intervenes, and an overview of the implementation. The proposed system comprises two web application functions: **reflection mode**, in which users organize the rationale for their judgments based on conversation histories, and **rehearsal mode**, in which users apply organized content to similar situations. Within each function, the AI coach is responsible for (i) selecting scenes to use as material for reflection, (ii) presenting behavioral perspectives that contribute to rapport formation, (iii) guiding the reflection procedure, and (iv) providing feedback on written content, thereby supporting users in advancing their examination grounded in conversation histories. The system was implemented as a web application, using Next.js (React) for the frontend and FastAPI (Python) for the backend. All conversation logs used in the system were fully anonymized (with personally identifiable information removed) before any processing or display.

### 3.1 Reflection Mode

Reflection mode aims not merely to confirm understanding but to organize judgments and utterances based on conversation histories into a form referenceable in similar situations. To align history reference with reflection input, the interface places AI coach suggestions and conversation histories on the left, and a guided work area on the right (Figure 1).

The left side displays the AI coach’s suggestions, relevant portions of the conversation history, and a summary of the full history. The suggestions indicate rapport-building behaviors recommended for improvement and specific scenes and utterances to focus on, with corresponding sections highlighted in the conversation history for reference. The suggestions (“Reflection Points for This Session”) are generated prior to the session through a three-stage process using Google’s Gemini 2.5 Pro [?], based on eight rapport-building behaviors defined from prior literature and practitioner review (attentive empathy, emotional empathy, cognitive empathy, apology, addressing by name, quantitative expressions, polite expressions, and easy-to-understand utterances) [3, 7, 11]. In the first stage, each operator utterance is classified according to these behavior types and assigned a priority score reflecting the degree of room

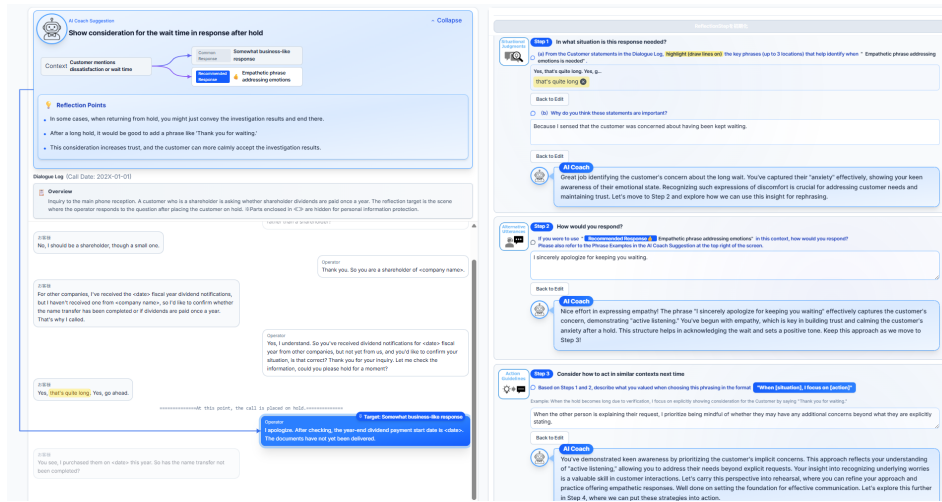


Figure 1: Example of the reflection support interface in Reflection AI Coach

for improvement and potential contribution to rapport (not as a performance rating). In the second stage, scenes and utterances for reflection are selected from the prioritized candidates, together with the recommended rapport-building behavior. In the third stage, feedback text is generated from the selected utterances and surrounding context, including a brief rationale for the recommendation and concrete directions for improvement. We deliberately present these suggestions as tentative cues for examining judgments and expressions, rather than as conclusions, so that users can articulate their own reasoning grounded in the conversation history.

In the right-side work area, users write their reflections following guided prompts in three steps: Step 1, identifying the rationale for situational judgment; Step 2, composing alternative utterances; and Step 3, organizing action guidelines. The details of each step are described below.

**Step 1: Identifying the Rationale for Situational Judgments.** In Step 1, users organize why a particular response was warranted, using customer utterances as evidence. The AI coach highlights sections related to the recommended rapport-building behavior in the conversation history, indicating utterances to focus on. Users select utterances and describe which phrases led them to read a particular situation (e.g., dissatisfaction, concern, misunderstanding, urgency). Narrowing down the evidential phrases at this stage helps users proceed through subsequent steps while continuing to reference the conversation history.

**Step 2: Composing Alternative Utterances.** In Step 2, based on the rationale organized in Step 1, users compose concrete alternative utterances from the perspective of “what to convey (e.g., reassure, resolve a misunderstanding, clarify the next action) and how to express it.”

**Step 3: Organizing Action Guidelines.** In Step 3, users summarize the content from Steps 1 and 2 as key points applicable to similar situations, briefly articulating (i) the situational characteristics (what customer state or context was present) and (ii) the appropriate response (what to prioritize and how to communicate). These

organized guidelines can then be referenced when facing similar situations in the future.

### 3.2 Rehearsal Mode

Rehearsal mode aims to connect the action guidelines developed during reflection to practical application by having users apply them to different similar situations. The system presents situations requiring the same type of improvement as the one addressed during reflection, and users compose alternative responses while referencing their Step 3 guidelines. Three similar situations are presented as rehearsal exercises.

Scene selection follows a process similar to the AI coach suggestion generation: the process is applied to multiple anonymized conversation records of the focal operator from a given period, and situations requiring the same type of improvement are retrieved from both the focal operator’s history and other operators’ histories.

In both modes, the AI coach presents a brief follow-up message generated by OpenAI’s GPT-4o [?] each time an input field is completed, acknowledging which aspects of the suggestions are reflected in the user’s input, with positive feedback as the default principle. When input is extremely brief or off-topic, a message prompting further description is generated.

## 4 Pilot Study

In the pilot study, we explored how well the proposed system supports reflection based on conversation histories and identified potential barriers to continued use regarding acceptability, burden, and feasibility under field constraints.

### 4.1 Methods

**Participants and Procedure.** Participants were two customer support operators (Participant A: female, 30s, over 10 years of experience; Participant B: female, 20s, less than 3 years). For each participant, the AI coach recommended a rapport-building behavior based on their anonymized conversation history, drawn from the eight types defined in Section 3. The three-stage analysis (Section 3.1)

selected emotional empathy as the behavior with the greatest room for improvement for both. AI coach suggestions were generated accordingly and reviewed by each participant’s supervisor for appropriateness. Each session lasted approximately 60 minutes: informed consent, pre-session questionnaire, operational instructions, reflection mode (Steps 1–3), rehearsal mode, post-session questionnaire, and semi-structured interview. This study was approved by the Research Ethics Committee of Toyo University (No.: P250072).

*Measures and Analysis.* A post-session 7-point Likert scale questionnaire separately measured the usefulness of the AI coach suggestions (including the AI-selected reflection scenes and improvement directions), the perceived usefulness of each reflection step (Step 1: situational understanding, Step 2: alternative utterances, Step 3: action guidelines), rehearsal mode, and behavioral change intention. A semi-structured interview elicited perceptions of insights, input burden, feasibility, and the specificity and perceived convincingness of AI suggestions.

A follow-up questionnaire one week later assessed changes in awareness and behavior during work (“Did you make an effort to improve?” “Did your behavior change?” “Was the reflection content useful during work?”).

Due to the small sample, no inferential statistics were conducted; scores were descriptively summarized. Interview transcripts were organized around acceptability (usefulness, convincingness) and barriers (time burden, etc.).

## 4.2 Results

In the post-session questionnaire, both participants rated the appropriateness of the AI coach suggestions as 6, and usefulness was rated 6 by Participant A and 5 by Participant B. For the reflection support steps, Step 1 was rated 6 by Participant A and 7 by Participant B, Step 2 was rated 6 by Participant A and 5 by Participant B, and Step 3 was rated 6 by Participant A and 5 by Participant B. Rehearsal mode was rated 6 by Participant A and 3 by Participant B, and behavioral change intention was rated 7 by Participant A and 5 by Participant B. These results suggest that the procedure of organizing situational judgments from conversation histories and connecting them to alternative utterances and action guidelines was generally well-received.

In the interviews, Participant B, who had less practical experience, noted primarily regarding Step 1 that “the highlighting of important sections helps me notice things I would have overlooked,” suggesting that cue presentation on conversation histories can support situational understanding. However, the same participant remarked about Step 3’s action guidelines that “it takes time to think through, making it difficult to reflect during short breaks between tasks,” indicating that time requirements could be a major barrier to continued use. Participant A, who had extensive practical experience, noted that reflection mode overlaps in some respects with the review methods she independently practices, suggesting that additional insights may be limited for users with existing reflection habits. Furthermore, the comment that “AI suggestions are easier to make use of when they are specific” suggested that the level of abstraction of suggestions may affect acceptability and connection to practice.

In the one-week follow-up survey, “Did you make an effort to improve?” and “Did your behavior change?” were both rated 5 by Participant A and 6 by Participant B. “Was the reflection content useful during work?” was rated 6 by both participants, suggesting a degree of workplace applicability.

## 5 Discussion

While the questionnaire ratings for Steps 1–3 were all within the 5–7 range, suggesting overall favorable evaluation, the interviews revealed that perceived burden differs across stages: Step 1’s highlighting was valued for “helping notice overlooked points,” whereas Step 3’s action guidelines were noted to be “time-consuming to think through.”

This difficulty likely stems from the fact that Step 3 requires abstracting one level up from concrete conversation histories to summarize “situational characteristics” and “appropriate responses” as key points. Furthermore, in the workplaces examined in this study, while concepts such as rapport formation and emotional empathy are understood as frameworks for reflection by supervisors, they are not necessarily used as explicit terminology in daily operations. For some operators, these concepts may be unfamiliar, making it difficult to form a concrete image of what to write and from what perspective when summarizing in Step 3, potentially leading to cognitive stagnation and a sense of burden. For sustained use, it therefore appears necessary to enhance UI-level instructions for the more abstract steps. It is also important to first clarify users’ own work-related goals and guide them to understand how the relevant behavior (e.g., emotional empathy) contributes to those goals.

This study is a preliminary investigation with only two participants, and the generalizability of its conclusions remains limited. Future work should explore enhanced guidance with stronger goal-connection and introduce contingent adjustment and fading of support. In evaluating effectiveness, it will be important to examine how these changes affect reflection content, burden, workload, and transfer to workplace practice, with more diverse participants and continued-use conditions.

## 6 Conclusion

This study proposed Reflection AI Coach, a reflection support system focusing on articulating the rationale for judgments and translating them into expressions usable in practice—processes difficult to address through summarization or advice alone. By positioning generative AI as a coach rather than a provider of correct answers, we presented a framework for three-stage reflection—identifying the rationale, composing alternative utterances, and organizing action guidelines—based on conversation histories. A small-scale pilot study suggested that this framework can be acceptable for reflection on conversation histories. These findings are expected to contribute to learning support design not only in customer service but also in medical consultations, educational counseling, negotiation, and other domains requiring interpersonal skills.

## References

- [1] Riku Arakawa, Hiromu Yakura, Kei Akuzawa, and Shizuma Kubo. 2025. AI for meeting minutes: Promises and challenges in designing human-AI collaboration on a production SaaS platform. In *Proceedings of the Extended Abstracts of the*

- CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–6.
- [2] Arngeir Berge, Frode Guriby, Siri-Linn Schmidt Fotland, Gro Fonnes, Ingrid H Johansen, and Christoph Trattner. 2023. Designing for control in nurse-AI collaboration during emergency medical calls. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA.
- [3] Jan Hendrik Blümel, Mohamed Zaki, and Thomas Bohné. 2024. Personal touch in digital customer service: a conceptual framework of relational personalization for conversational AI. (Jan. 2024).
- [4] Tianying Chen, Michael Xieyang Liu, Emily Ding, Emma O’Neil, Mansi Agarwal, Robert E Kraut, and Laura Dabbish. 2023. Facilitating counselor reflective learning with a real-time annotation tool. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Vol. 44. ACM, New York, NY, USA, 1–17.
- [5] Xinyue Chen, Lev Tankelevitch, Rishi Vanukuru, Ava Elizabeth Scott, Payod Panda, and Sean Rintel. 2025. Are we on track? AI-assisted active and passive goal reflection during meetings. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–22.
- [6] Kevin Chow, Roy Rutishauser, André N Meyer, Joanna McGrenere, and Thomas Fritz. 2025. Exploring a real-time feedback display of non-verbal cues in online work meetings to support self-presentation. *Proc. ACM Hum. Comput. Interact.* 9, 2 (May 2025), 1–28.
- [7] Colin Mackinnon Clark, Ulrike Marianne Murfett, Priscilla S Rogers, and Soon Ang. 2013. Is empathy effective for customer service? Evidence from call center interactions. *J. Bus. Tech. Commun.* 27, 2 (April 2013), 123–153.
- [8] Yue Fu, Sami Foell, Xuhai Xu, and Alexis Hiniker. 2023. From text to self: Users’ perceptions of potential of AI on interpersonal communication and self. *arXiv [cs.HC]* (Oct. 2023).
- [9] Amie M Gordon and Emily Diamond. 2023. Feeling understood and appreciated in relationships: Where do these perceptions come from and why do they matter? *Curr. Opin. Psychol.* 53, 101687 (Oct. 2023), 101687.
- [10] Silvia Mamede and Henk G Schmidt. 2004. The structure of reflective practice in medicine. *Med. Educ.* 38, 12 (Dec. 2004), 1302–1308.
- [11] John Mendonça, Patrícia Pereira, Miguel Menezes, Vera Cabarrão, Ana C Farinha, Helena Moniz, João Paulo Carvalho, Alon Lavie, and Isabel Trancoso. 2023. Dialogue quality and emotion annotations for customer support conversations. *arXiv [cs.CL]* (Nov. 2023).
- [12] K L Mosier, L J Skitka, S Heers, and M Burdick. 1997. Automation bias: decision making and performance in high-tech cockpits. *Int. J. Aviat. Psychol.* 8, 1 (1997), 47–63.
- [13] Tricia J Ngoon, S Sushil, Angela E B Stewart, Ung-Sang Lee, Saranya Venkatraman, Neil Thawani, Prasenjit Mitra, Sherice Clarke, John Zimmerman, and Amy Ogan. 2024. ClassInSight: Designing conversation support tools to visualize classroom discussion for personalized teacher professional development. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Vol. 5. ACM, New York, NY, USA, 1–15.
- [14] Raja Parasuraman and Dietrich H Manzey. 2010. Complacency and bias in human use of automation: an attentional integration. *Hum. Factors* 52, 3 (June 2010), 381–410.
- [15] Kai Qin, Kexin Du, Yimeng Chen, Yueyan Liu, Jie Cai, Zhiqiang Nie, Nan Gao, Guohui Wei, Shengzhu Wang, and Chun Yu. 2025. Customer service representative’s perception of the AI assistant in an organization’s call center. In *Companion Publication of the 2025 Conference on Computer-Supported Cooperative Work and Social Computing*. ACM, New York, NY, USA, 399–404.
- [16] Donald A Schoen. 2017. *The reflective practitioner: How professionals think in action*. Routledge, London, England.
- [17] Donghoon Shin, Soomin Kim, Ruoxi Shang, Joonhwan Lee, and Gary Hsieh. 2023. IntroBot: Exploring the use of chatbot-assisted familiarization in online collaborative groups. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Vol. 12. ACM, New York, NY, USA, 1–13.
- [18] Linda Tickle-Degnen and Robert Rosenthal. 1990. The nature of rapport and its nonverbal correlates. *Psychol. Inq.* 1, 4 (Oct. 1990), 285–293.
- [19] Janneke van de Pol, Monique Volman, and Jos Beishuizen. 2010. Scaffolding in teacher–student interaction: A decade of research. *Educ. Psychol. Rev.* 22, 3 (Sept. 2010), 271–296.
- [20] Xiaotong (tone) Xu, Arina Konnova, Bianca Gao, Cindy Peng, Dave Vo, and Steven P Dow. 2025. Productive vs. Reflective: How different ways of integrating AI into design workflows affect cognition and motivation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–15.
- [21] Xiaotong (tone) Xu, Jiayu Yin, Catherine Gu, Jenny Mar, Sydney Zhang, Jane L., and Steven P Dow. 2024. Jamplate: Exploring LLM-enhanced templates for idea reflection. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. ACM, New York, NY, USA, 907–921.