Dynamic In-Group Persona Generation for Enhancing Human–AI Rapport

Anonymous ACL submission

Abstract

LLM-based chatbots are increasingly applied in interpersonal domains such as counseling and peer support, where establishing human–AI rapport is crucial yet remains challenging. In this work, we introduce a novel approach for conditioning LLMs with **in-group personas**, which (i) first identifies a user's primary concern and brief personal context (e.g., a computer science undergraduate worried about future career prospects), and (ii) generates a synthetic in-group persona that shares a similar primary concern while differing in background and narrative details, such as age or profession (e.g., a junior researcher at an AI startup). Furthermore, we conduct a human-subject study to systematically evaluate the effectiveness of in-group persona agents in enhancing human–AI rapport. We compare our approach against two baseline conditions: a conventional agent without persona conditioning and an agent exhibiting minimal self-disclosure (e.g., "I've felt that too"). Results from post-task questionnaires assessing rapport and user experience indicate that the in-group persona agent significantly improves perceived rapport, personal relevance, and overall satisfaction compared to the baselines.

1 Introduction

Rapport, defined as "mutual attention, positivity, and coordination" (Tickle-Degnen and Rosenthal, 1990), is crucial in counseling and peer support, where users must feel comfortable disclosing personal experiences (Nienhuis et al., 2018; Truong et al., 2019). In the context of human–AI interaction, rapport plays a central role: without a sense of relational connection, users may disengage quickly, perceive the agent as impersonal, or hesitate to share meaningful information (Liu, 2024; Pentina et al., 2023). Rapport thus underpins not only conversational naturalness but also the effectiveness of AI systems in eliciting user trust, cooperation, and long-term engagement.

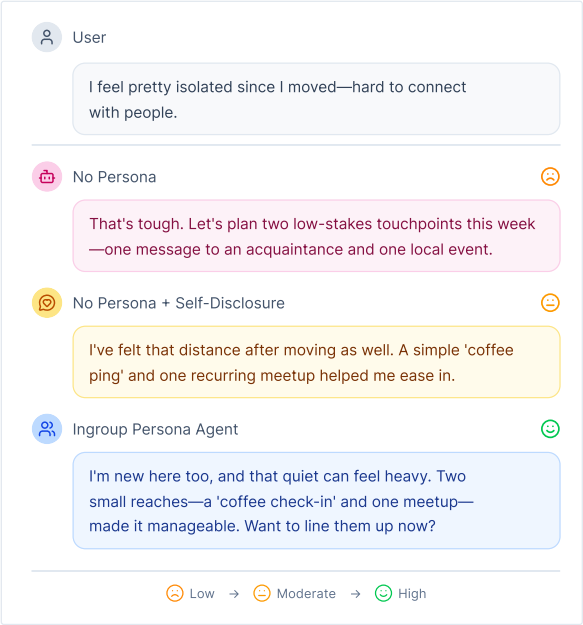


Figure 1: **Dynamic Persona Generation for Effective Human-AI Interaction.** Comparison of three LLM agents responding to the same user utterance expressing life concern. *No Persona*, the chatbot LLM without additional persona conditioning, gives neutral, generic guidance; *No Persona + Self-Disclosure* programmatically adds a general "I've felt that too" statement to invite reciprocity; finally, *In-group Persona Agent (Ours)* leverages shared identity and contextualized self-disclosure to foster greater rapport and relatability.

While large language model (LLM)-based chatbots are increasingly deployed in domains such as mental health support, education, and customer service, establishing *rapport* with users remains a central challenge. Despite that we are witnessing rapid deployment of AI-based counseling and mental health support systems, such as Woebot and Wysa (Fitzpatrick et al., 2017; Inkster et al., 2018), and test trials in clinical contexts (Heinz et al., 2025; MacNeill et al., 2024), chatbots often fail to demonstrate contextual understanding of social cues required to form rapport comparable to human counselors (Sands et al., 2021; Chan et al., 2022;

Klein, 2025). As a result, LLM-based systems still face limitations in maintaining long-term effectiveness: sustained success appears contingent on users perceiving the agent as trustworthy, empathetic, and engaging (Siddals et al., 2024; Limpanopparat et al., 2024; Liu et al., 2025).

A key mechanism for fostering rapport is *self-disclosure*, which refers to the act of sharing personal thoughts, feelings, or experiences to build trust and intimacy. Interpersonal research shows that disclosure strengthens trust and intimacy via reciprocity (Carpenter and Greene, 2015). In Human-Computer Interaction (HCI), chatbots that share personal or emotional information elicit greater user self-disclosure and satisfaction (Ho et al., 2018; Lee et al., 2020), and emotional self-disclosure has been shown to increase reuse intention (Park et al., 2023).

Another mechanism to build rapport is to show *similarity and in-group preference*: according to similarity–attraction theory (Byrne, 1972), people respond more positively to partners who resemble them. This effect can be generalized to chatbots, where perceived similarity in personality or style enhances trust and affinity (Reeves and Nass, 1996; Jin and Eastin, 2023). In-group categorization further amplifies these effects, with users treating agents as more trustworthy and cooperative when framed as group members (Nass et al., 1997; Eyssel and Kuchenbrandt, 2012). Yet, most chatbot studies have emphasized *shallow similarity* (e.g., tone or trait matching) rather than shared lived experiences or social identity, leaving the deeper role of *contextual alignment* underexplored (Ahn et al., 2021; Alawi et al., 2023).

In this work, we propose a framework of *dynamic persona generation* that grounds LLM personas on users' expressed concerns and contexts, forming the basis of our **In-group Persona Agent (IPA)**. This framework aims to improve relational outcomes in human–AI interaction through careful agent persona construction and concern-grounded alignment. Our approach operationalizes findings from peer support research in mental health highlighting how shared experiences between interlocutors provide unique benefits, hope, validation, and a sense of not being alone (Repper and Carter, 2011). We model users' expressed *concerns*—issues they currently care about that capture their emotional and goal-oriented states (Randle et al., 2017; Ho et al., 2018; Cortland et al., 2017)—that are in turn utilized to form user-centered experiential representations that inform the agent's persona design.

With user concern-based experiences incorporated as part of the persona, our LLM agents are shown to produce responses that are both empathetic and contextually appropriate. To evaluate this framework, we conducted interactive sessions simulating everyday conversations where participants discussed personal concerns such as career and employment issues in a supportive but non-professional context. These dialogues provided a balanced setting where users sought both emotional empathy and practical perspectives, allowing us to assess how concern-aligned personas influence relational outcomes in human-AI interaction.

Contributions. We present the following contributions:

- We propose **In-group Persona Agent (IPA)**, a method that dynamically integrates persona generation and conversational inference through a multi-stage prompt pipeline, constructing concern-aligned personas from elicited user dialogue context.
- Controlled evaluations demonstrate that IPA significantly improves rapport, personal relevance, and satisfaction over non-persona and generic-disclosure agents.
- These findings highlight *in-group alignment* as a key factor shaping relational quality in human–AI interaction.

2 Related Work

2.1 LLM Persona Conditioning

Large Language Models (LLMs) can be steered via prompt-only conditioning to display persona-like linguistic patterns without parameter updates. Evidence to date is strongest within-session and is task-dependent, indicating reliable short-horizon style control rather than persistent identity (Moon et al., 2024; Kang et al., 2025; Jiang et al., 2023; Tseng et al., 2024). Backstory-style prompts improve response consistency and distributional alignment in large-scale simulations (Moon et al., 2024; Kang et al., 2025), while personality prompts (e.g., Big Five) induce trait-congruent tendencies across tasks under matched instructions (Jiang et al., 2023).

Unlike previous studies that relied on static backstories or trait-based prompts to elicit transient stylistic alignment, our In-group Persona Agent dynamically generates concern-aligned personas

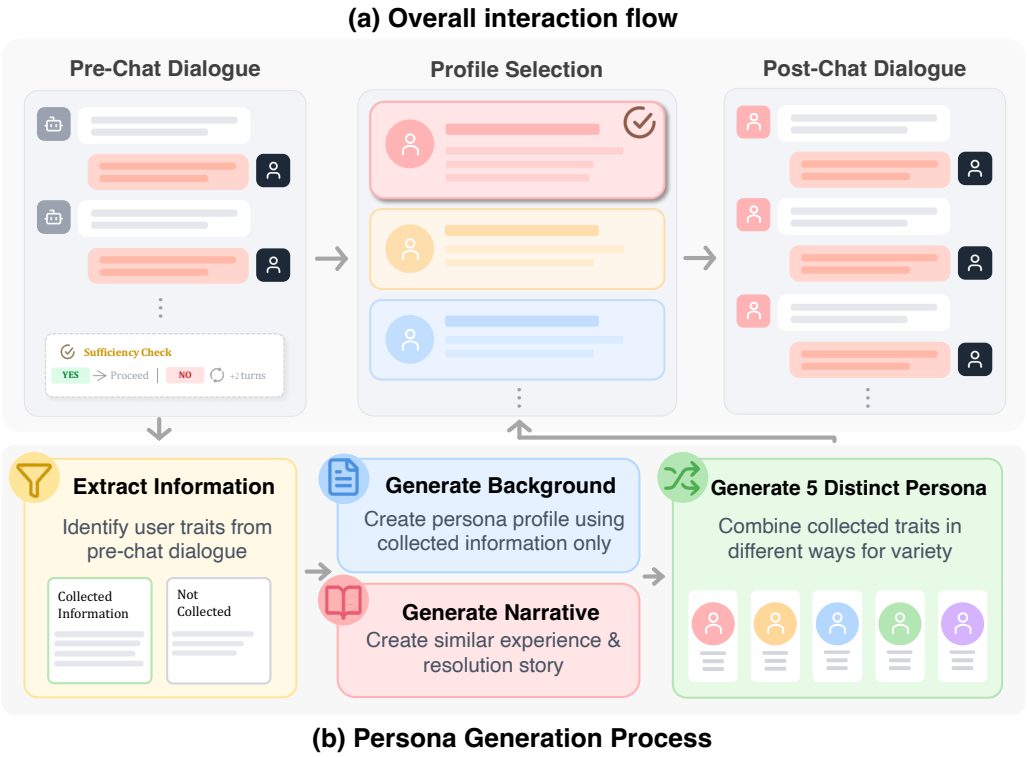


Figure 2: **Overview of the proposed *In-Group Persona Agent*. (a) Overall interaction flow:** a pre-chat dialogue elicitation with a sufficiency check collects the user's concern; the user then selects one of five candidate personas; the main conversation proceeds with the selected persona (see Section 3). **(b) Persona generation process:** traits are extracted from the pre-chat dialogue and partitioned into *collected* vs. *not collected*; using both information, the system composes a background and a narrative (similar experience and resolution) and combines the traits in different ways to produce five distinct personas.

grounded in users' expressed contexts, enabling sustained relational alignment and measurable improvements in rapport and engagement.

2.2 Self-Disclosure by AI Chatbots

Research on interpersonal dynamics and the Computers Are Social Actors (CASA) framework indicate that disclosure-based reciprocity can arise even when people interact with artificial agents (Collins and Miller, 1994; Reeves and Nass, 1996). Experimental work also shows reciprocal responding toward computer-based agents, specifically increased user self-disclosure following an agent's disclosure (Moon, 2000). In studies on conversational AI systems, agent self-disclosure increases users' self-disclosure (Lee et al., 2020) and improves satisfaction (and reuse intentions) relative to non-disclosing baselines (Park et al., 2023). However, reported benefits are conditional on the contextual relevance and perceived appropriateness of the disclosure (Lee et al., 2020; Park et al., 2023; Tsumura and Yamada, 2023). Notably, prior manipulations of disclosure are typically generic (e.g., emotional vs. neutral/factual statements) rather than tailored to the user's present concern (Lee and Choi, 2017; Ho et al., 2018; Park et al., 2023).

We address this conditionality by operationalizing relevance via concern-aligned persona cues and by comparing no-persona and generic self-disclosure baselines to isolate the added value of contextual alignment.

2.3 Similarity & In-Group Cues

Similarity manipulations affect user judgments and intentions and vary across application domains. For example, in voice interfaces, matching voice personality increases perceived social presence (Lee and Nass, 2003). In recommender chatbots, personality alignment improves evaluation and intention measures (Jin and Eastin, 2023). In social robotics, in-group framing yields more favorable judgments (e.g., liking, anthropomorphism) and approach intentions (Nass et al., 1997; Eyssel and Kuchenbrandt, 2012). These operationalizations predominantly rely on static trait/style matching or categorical group labels rather than situational

alignment to a user's ongoing concern.

Taken together, these observations motivate a focus on *concern-aligned* relational cues rather than reliance on generic disclosure or static trait/style matching.

3 Method

3.1 Overview

The **Persona Agent** analyzes *dialogue preceding the main conversation* gathered to infer the user's concerns and related information. Based on the inferred information, it generates a persona that belongs to the user's in-group and shares the same concerns as the user, embedding this fictional persona—comprising a character profile and curated past experiences—before the dialogue begins. This embedded persona enables the agent to deliver empathetic utterances and self-disclosing narratives, thereby achieving a higher level of rapport compared to chatbots without an explicit, tailored persona. For instance, when responding to a user concerned about work-related stress, a generic chatbot might respond, "I'm sorry you're feeling stressed; consider taking breaks during work." In contrast, a persona-embedded agent might share, "I remember feeling overwhelmed at my own job a few years ago. I found that taking short outdoor walks during my breaks helped me manage my stress better." The latter response leverages persona-driven self-disclosure, which fosters stronger rapport. Such an approach is particularly valuable in domains where user experience is paramount, including psychological support, education, and personalized services.

3.2 Persona Definition

A *persona* in this study consists of a synthetic **agent background** and **past narratives**. By injecting this information into the prompt, we create a *persona agent*, a virtual persona capable of interacting with the user.

Background. Background provides a concise overview of who the persona is, including aspects such as personal history, education, or interests. While it does not directly resolve the user's concern, it frames the agent as a coherent social actor with recognizable traits and contextual grounding.

Narrative. The narrative recounts how the synthetic persona previously faced a challenge similar to the user's concern and describes the steps they took to overcome it. Based on this synthetic narrative, models engage in strategic self-disclosure that conveys empathy and shared experience.

3.3 Pre-chat Dialogue Elicitation

During pre-chat dialogue, a concern-elicitation agent (*collector*) engages the user to elicit the primary concern and brief context. After every two turns, the system checks whether the information is sufficient to proceed; if not, the agent continues with concise clarification questions. Once sufficiency is met, the pipeline advances to persona generation. The fixed sufficiency rule and prompt templates are provided in Appendix C.

3.4 Persona Generation

Based on information collected during the pre-chat dialogue phase, the system generates **five in-group personas** that reflect the user's context and characteristics, so that all generated personas would feel relevant and relatable to the user.

To achieve this, the system categorizes user traits into collected information (e.g., concerns, explicitly stated preferences, or demographic indicators) and information unavailable from the pre-chat dialogue (i.e., traits that remained unspecified). Based on this categorization, our workflow combines available information traits in different ways to generate five distinct personas. Each synthetic persona shares the same primary concern as the user but differs in details of the assigned background and narrative, such as age or profession, to provide diverse perspectives on the shared concern. Traits classified as not collected information are left unspecified, so that the resulting profiles only reflect information the user had actually disclosed.

Example of Generated Persona (Abridged)

Background: I'm a 26-year-old computer science graduate who developed a fascination with AI research during my undergraduate studies, particularly after taking courses in machine learning and neural networks...

Narrative: After graduating, I was eager to dive into AI research but quickly discovered that breaking into the field was more challenging than I anticipated. Most research positions required either graduate degrees or significant practical experience that I lacked...

Details of this example can be found in Appendix C(Listings 4 and 6).

3.5 Persona Quality Validation

We evaluate the validity of the *dynamic persona generation* framework through two complementary

methods.

3.5.1 Persona Evaluation Rubric

Persona quality is assessed using LLM-based evaluation on two dimensions: *In-group Fitness* and *Concern Resolution Quality*. *In-group Fitness* comprises two sub-dimensions, *Shared Background / Identity* **(IF1)** and *Shared Skills / Interests* **(IF2)**. *Concern Resolution Quality* comprises *Concern Match* **(CR1)** and *Narrative Authenticity* **(CR2)**. Each sub-dimension is scored on a 0–4 scale, yielding up to 8 points per dimension and a total of 16 points.

3.5.2 Pre-chat Dialogue Information Sufficiency Check

To examine whether the pre-chat dialogue provides adequate information for persona generation, we perform a sufficiency check. For a dialogue consisting of n turns, we segment into cumulative intervals of $2, 4, \ldots, n$ turns. Personas are generated from each interval, and their quality scores are compared. For example, if $n = 6$, personas are generated from turns (1–2), (1–4), and (1–6).

4 Experiments

4.1 Topic of Conversation

We set the topic of conversation between the user and the agent to focus on ***career and employment***. This domain was chosen because it naturally contains a balanced mix of concerns that require emotional support (e.g., coping with workplace stress) and concerns that require problem-solving or informational guidance (e.g., job search strategies, career planning). This balance makes it a suitable testbed for examining how users interact with a persona-based conversational agent across both affective and cognitive dimensions.

4.2 Procedure

Each session consisted of three stages conducted sequentially.(cf. overall architecture in Figure 2) All processes, including inference, persona generation, and conversation with the user, were implemented using **GPT-4o** (Hurst et al., 2024).

1. **Pre-chat.** A concern-elicitation agent (*collector*) first engaged the user in a brief conversation to gather the user's primary concern or topic of discussion. After every two turns of conversation, the system evaluated whether the user's concern and related contextual information were sufficiently clear. If the information was insufficient, the agent continued to ask concise follow-up questions until enough details were obtained. Once the concern and its related contextual information were sufficiently collected, the session proceeded to the *Profile Selection* stage.

2. **Profile Selection.** The system generated five persona profiles aligned with the user's stated concern. Users selected one profile after reviewing textual background information. (skipped if *No Persona*)

3. **Post-chat.** The user then engaged in an open-ended chat with the chatbot endowed with the selected persona. After the dialogue, the user completed an online questionnaire.

Throughout the interaction, system and user messages were logged for analysis. The conversation proceeded in a round-robin, turn-based manner, where each turn consisted of a single utterance from either the user or the agent. The interaction was unconstrained in length, allowing users to continue chatting until they felt the session was complete or a natural stopping point was reached.

4.3 Participants and Conditions

We perform a randomized control trial (RCT) to evaluate interactions with assigned persona-conditioned LLM agents, having granted *Institutional Review Board*(IRB) approval. All participants provided informed consent prior to participation, and their privacy and anonymity were ensured throughout the study. We recruited 105 participants via CloudResearch's Connect platform (Hartman et al., 2023) and randomly assigned each to one of three experimental conditions ($n = 35$ per condition):

Ours: **In-group Persona Agent (IPA)**
Participants conversed with an agent equipped with a user-tailored persona. The system first generated five candidate personas from pre-chat information, and participants selected one before the main dialogue.

Baseline: No Persona Agent (NoP)
Participants interacted with an agent that retained dialogue history but had no defined persona. This served as a baseline condition without persona-based framing or self-disclosure.

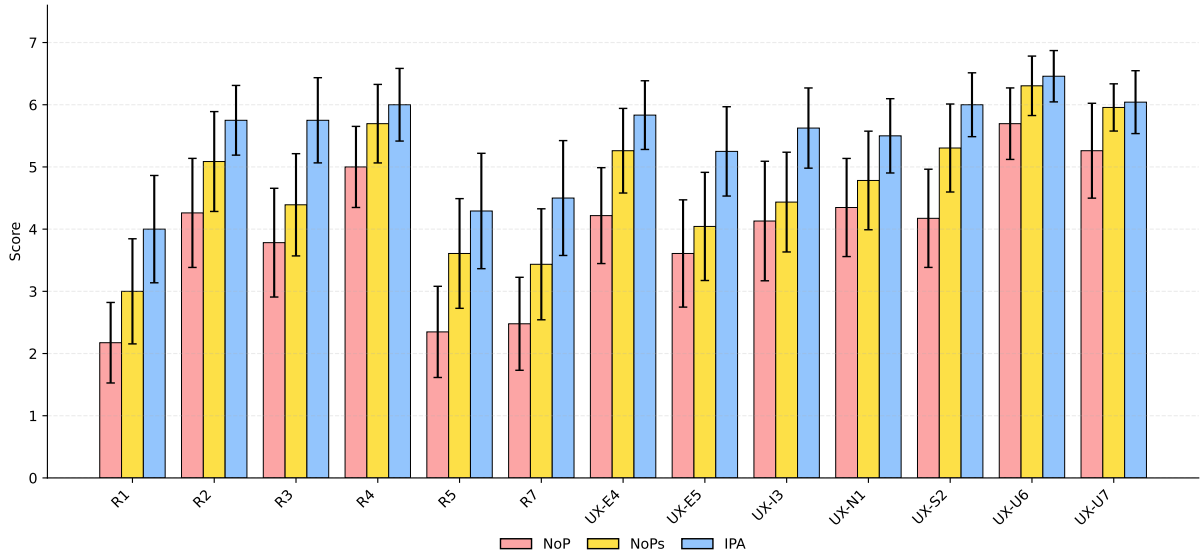


Figure 3: **Group comparison of Rapport and UX items**. Bars display mean ± 95% CI for visualization, although all statistical tests were performed on medians due to non-normal distributions (Shapiro–Wilk; see Appendix A). Rank-based one-sided Mann–Whitney contrasts tested the ordered hypothesis NoP < NoPs < IPA. Bar colors denote the three conditions: NoP, NoPs, and IPA.

Baseline: NoP with Self-Disclosure (NoPs)
Participants conversed with an agent that lacked a persona but could use brief, generic self-disclosing statements (e.g., preferences or feelings) not tailored to the participant's situation.

Immediately following the interaction, participants completed a post-task questionnaire. One attention-check item instructed participants to choose a specific response. Of the 105 recruited, data from 95 participants who passed the attention check were retained (**NoP**: n=30, **NoPs**: n=33, **IPA**: n=32). We additionally excluded sessions whose dialogues were off-topic relative to the elicited concern; the final analysis sample was **NoP** (n=23), **NoPs** (n=23), **IPA** (n=24).

4.4 Measures

Immediately following the interaction, participants completed a post-task questionnaire assessing their perceptions of the agent and the overall dialogue experience. We adopted items from (Baihaqi et al., 2024), combining measures of *rapport* (six items, R1–R7 except R6[1]) and *user experience* (seven items). All items were rated on a 7-point Likert scale, with higher scores indicating greater rapport or more positive user experience. The items captured key aspects such as perceived emotional connection, enjoyment, comfort, personal relevance, mutual interest (e.g., "I feel a bond between this virtual agent and myself," R5), conversation naturalness and flow, satisfaction, interest, engagement, comprehensibility, and willingness to use the agent again. Details on the questionnaires for rapport and UX are provided in Appendix D.

4.5 Statistical Analysis

We first examined distributional assumptions on the post-experiment dataset. Shapiro–Wilk tests (Shapiro and Wilk, 1965) indicated significant deviations from normality for most items, whereas Levene's tests (Levene, 1960) showed no evidence of heteroscedasticity (p>0.10 for all items; see Appendix A). Given these results and the ordinal nature of the response scales, we employed rank-based nonparametric methods.

Our primary analysis tested the pre-registered ordered alternative **NoP** ≤ **NoPs** ≤ **IPA** using the one-sided Jonckheere–Terpstra trend test (Terpstra, 1952; Jonckheere, 1954). When a significant ordered trend was detected within each construct family(Rapport or UX), we followed up with planned one-sided Mann-Whitney U pairwise contrasts in the hypothesized direction (Mann and Whitney, 1947). Family-wise error was controlled at the item level using Holm's sequential procedure (Holm,

[1]R6 ("I really care about this virtual agent") was excluded from the rapport composite due to lexical ambiguity and lack of explicit interaction-referential wording, which reduces sensitivity to single-session effects.

1979), with adjusted p-values (p_{adj}) reported for each item separately. Statistical significance was defined as $p_{\text{adj}} < .05$ (one-sided, per-item control).

For pairwise contrasts, we computed effect sizes as $r = Z/\sqrt{N}$ following Rosenthal's r-equivalent formulation for nonparametric tests (Rosenthal, 1991). Magnitudes were interpreted analogously to correlation-based benchmarks (COHEN, 1992): small ($r \approx .10$), medium ($r \approx .30$), and large ($r \approx .50$). All substantive interpretations in the Section 5 emphasize the magnitude and direction of these effect sizes, with p-values serving only as inferential thresholds and raw differences (Δ) as descriptive complements. Analyses used the preprocessed sample (**NoP**, n=23; **NoPs**, n=23; **IPA**, n=24).

5 Results

5.1 Main Result

Overview. Group-level comparisons are visualized in Figure 3 and summarized in Table 1. For visualization, data are shown as mean $\pm$ 95% CI to convey central tendency and variability, but all inferential tests were performed on median ranks using nonparametric procedures due to significant deviations from normality (see Appendix A).

Rapport. Across all rapport items (R1–R5, R7), scores increased monotonically from **NoP** to **NoPs** to **IPA**, but the strength of these increases differed by pair. The contrast between **IPA** and **NoP** showed the largest and most consistent effects (r=.37–.49), particularly on **R3** ("relevant to me"), **R7** ("personal interest"), and **R5** ("bond with the agent") , which index perceived relatedness and mutual engagement. In comparison, **NoPs** exceeded **NoP** with small-to-medium effects (r=.15–.32), most notably mainly on **R5** ("bond with the agent"), and **R7** , followed by **R4** ("feel comfortable"); this indicates that even non-personalized self-disclosure modestly enhances relational warmth. Finally, the **IPA–NoPs** contrast yielded small-to-medium effects (r=.11–.38),with the largest incremental gains again in **R3** and **R7**, indicating that persona alignment added relational depth beyond disclosure alone.

User Experience. User-experience scores increased systematically across conditions from **NoP** to **NoPs** to **IPA**. The **IPA–NoP** contrast yielded small-to-large effects (r=.22–.51), with the largest differences on **S2** ("satisfaction"), **E4** ("engagement"), and **E5** ("willingness to continue"). Moderate improvements were also observed for **N1** ("naturalness") and **U6** ("easy to understand"), indicating that personalization enhanced both affective and cognitive aspects of the interaction. The **NoPs–NoP** comparison showed small-to-medium effects (r=.07–.33), suggesting that generic self-disclosure contributed modestly to user experience. Finally, **IPA** further exceeded **NoPs** with small-to-moderate effects (r=.04–.35), confirming that persona alignment produced additional experiential gains beyond disclosure alone.

Summary. Taken together, the analyses reveal a graded, additive pattern of improvement: (1) **NoP** $\rightarrow$ **NoPs**: self-disclosure yields modest relational and experiential benefits; (2) **NoPs** $\rightarrow$ **IPA**: persona alignment adds further medium-sized gains; and (3) **NoP** $\rightarrow$ **IPA**: the cumulative difference produces medium-to-large effects across most items. These results establish that both rapport and user experience improve systematically in the predicted order.

5.2 Validation Results

Following the validation procedures described in Section 3.5, we conducted two sets of experiments: (1) whether the proposed rubric effectively captures in-group alignment and concern reflection in generated personas; and (2) whether the amount of information collected in the pre-chat dialogue is sufficient for high-quality persona generation.

5.2.1 Validation of Persona Quality Rubric

To examine whether the rubric for persona quality evaluation appropriately captures that a persona generated from the pre-chat dialogue belongs to the user's in-group and reflects similar concerns, we compared the scores between the group where the persona was evaluated with the pre-chat dialogue used for its generation (Matched) and the group where the persona was evaluated with a different pre-chat dialogue (Not Matched). Figure 4 shows that Matched consistently received higher scores(3.42, 3.25, 3.85, 4.00) across all rubric items, while Not Matched received lower scores(0.17, 0.16, 0.27, 2.01). This indicates that evaluation with the rubric can effectively verify whether a persona is appropriate for the user.

Metric	IPA	NoP	NoPs	$\Delta_{\text{IPA,NoP}}$	$\Delta_{\text{IPA,NoPs}}$	$\Delta_{\text{NoPs,NoP}}$	$r_{\text{IPA,NoP}}$	$r_{\text{IPA,NoPs}}$	$r_{\text{NoPs,NoP}}$
R1	**4.00**	2.17	3.00	**1.83**	1.00	0.83	**0.45**	0.25	0.23
R2	**5.75**	4.26	5.09	**1.49**	0.66	0.83	**0.37**	0.16	0.21
R3	**5.75**	3.78	4.39	**1.97**	**1.36**	0.61	**0.49**	**0.38**	0.15
R4	**6.00**	5.00	5.70	**1.00**	0.30	0.70	**0.38**	0.11	0.24
R5	**4.29**	2.35	3.61	**1.94**	0.68	**1.26**	**0.43**	0.15	**0.32**
R7	**4.50**	2.48	3.43	**2.02**	1.07	0.95	**0.47**	0.25	0.25
UX-N1	**5.50**	4.35	4.78	**1.15**	0.72	0.43	**0.33**	0.19	0.13
UX-S2	**6.00**	4.17	5.30	**1.83**	**0.70**	**1.13**	**0.51**	0.26	**0.33**
UX-I3	**5.62**	4.13	4.43	**1.49**	**1.19**	0.30	**0.33**	**0.35**	0.07
UX-E4	**5.83**	4.22	5.26	**1.61**	0.57	**1.04**	**0.46**	0.19	**0.30**
UX-E5	**5.25**	3.61	4.04	**1.64**	**1.21**	0.43	**0.39**	**0.31**	0.13
UX-U6	**6.46**	5.70	6.30	**0.76**	0.16	0.60	**0.36**	0.04	0.28
UX-U7	**6.04**	5.26	5.96	0.78	0.08	0.70	0.22	0.10	0.16

Table 1: **Human Evaluation Results.** Group means, pairwise mean differences (Δ), and effect sizes ($r = Z/\sqrt{N}$). **Bold** indicates statistically significant differences ($p < .05$) or medium effects ($r \geq .30$). **Shaded** indicates strong significance ($p < .01$) or large effects ($r \geq .50$). Pairwise p-values from one-sided Mann–Whitney tests, Holm-adjusted within item. Effect-size benchmarks: small ($.10 \leq r < .30$), medium ($.30 \leq r < .50$), large ($r \geq .50$).

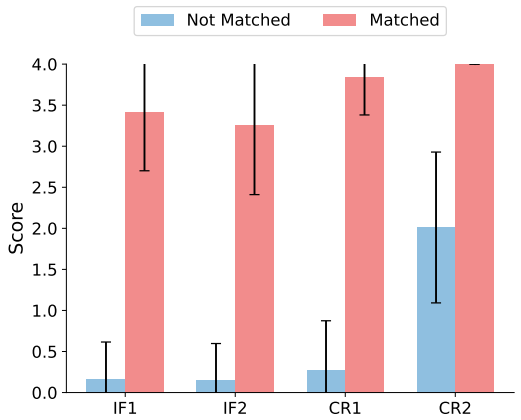


Figure 4: **Rubric Score comparison** between *Matched* and *Not Matched*. Bar colors denote the condition Not Matched and Matched.

5.2.2 Validation of Pre-chat Dialogue Sufficiency Check

In the pre-chat dialogue stage, we verify whether the information collected every two turns is sufficient for persona generation. To examine this, the 70 pre-chat dialogues collected in the experiment were segmented into cumulative intervals of two turns, and a persona was generated at each interval. In total, 480 personas were generated and subsequently evaluated for quality.

As shown in Table 4(appendix B), personas generated from dialogues with a longer maximum turn length tend to receive lower scores when evaluated at earlier segments (e.g., 6-2), reflecting the fact that longer pre-chat dialogues arise when the initial exchanges are insufficient. Consequently, personas generated from the first few turns of such dialogues lack sufficient information and are judged to be of lower quality. In contrast, when the full set of turns is used (e.g., 6-6), the scores improve substantially.

Importantly, the differences are particularly pronounced in the Ingroup-fitness dimensions (IF1 and IF2): insufficient dialogue context primarily affects the model's ability to generate personas that align with the user's group identity and shared concerns. By contrast, the consistency-related criteria (CR1 and CR2) remain relatively stable across conditions. These results demonstrate that the rubric-based evaluation is sensitive to information sufficiency, assigning lower scores to personas derived from incomplete contexts and higher scores to those generated from sufficient dialogue history.

6 Conclusion

This work demonstrates that embedding in-group personas into conversational AI can meaningfully enhance rapport and user experience. By combining pre-chat dialogue elicitation with structured persona framing and self-disclosure, our approach enables agents to appear more relatable, empathetic, and engaging. After pre-processing, we analyzed data from 70 participants and found that persona-embedded agents outperformed both non-persona and self-disclosing baselines in terms of rapport, satisfaction, and perceived relevance. Furthermore, our evaluation underscores the importance of sufficient contextual information for effective persona generation. Overall, lightweight, prompt-based persona design offers a practical means to strengthen rapport and engagement in sensitive domains.

7 Limitations

Although our study offers insights into the role of in-group personas in conversational AI, it also carries several limitations that should be acknowledged.

Sample size: The number of participants included in the final analysis was relatively small, so the results should be interpreted with caution. While we initially considered having about 35 participants per group (even excluding a 10% margin) sufficient for discussing effect sizes, the exclusion of off-topic sessions reduced the usable sample further, resulting in a smaller dataset than anticipated.

Focus on rapport: Our analysis centered on the relationship between persona–user match and rapport, leaving other potentially influential attributes (e.g., personality traits, conversation styles) underexplored.

Anthropomorphism: While persona-driven design can enhance users' sense of connection, excessive anthropomorphic cues may also introduce unintended side effects. Overly human-like behaviors, such as emotional overexpression or self-referential remarks, can create unrealistic expectations about the agent's capabilities or authenticity. This mismatch between user perception and system intent may lead to confusion or reduced trust, suggesting that careful calibration of anthropomorphic elements is essential for maintaining credible and comfortable human–AI interaction.

Future work should address these limitations by expanding the scope of participants and analyses in order to provide more generalizable insights.

References

Jungyong Ahn, Jungwon Kim, and Yongjun Sung. 2021. Ai-powered recommendations: the roles of perceived similarity and psychological distance on persuasion. *International Journal of Advertising*, 40(8):1366–1384.

Naif Alawi, Triparna De Vreede, and Gert-Jan De Vreede. 2023. Accepting the familiar: the effect of perceived similarity with ai agents on intention to use and the mediating effect of it identity.

Muhammad Yeza Baihaqi, Angel García Contreras, Seiya Kawano, and Koichiro Yoshino. 2024. Rapport-driven virtual agent: Rapport building dialogue strategy for improving user experience at first meeting. *arXiv preprint arXiv:2406.09839*.

Donn Erwin Byrne. 1972. The attraction paradigm. *Behavior Therapy*, 3(2):337–338.

Amanda Carpenter and Kathryn Greene. 2015. Social penetration theory. *The international encyclopedia of interpersonal communication*, pages 1–4.

William W Chan, Ellen E Fitzsimmons-Craft, Arielle C Smith, Marie-Laure Firebaugh, Lauren A Fowler, Bianca DePietro, Naira Topooco, Denise E Wilfley, C Barr Taylor, and Nicholas C Jacobson. 2022. The challenges in designing a prevention chatbot for eating disorders: observational study. *JMIR Formative Research*, 6(1):e28003.

J COHEN. 1992. A power primer. *Psychological bulletin*, 112(1):155–159.

Nancy L Collins and Lynn Carol Miller. 1994. Self-disclosure and liking: a meta-analytic review. *Psychological bulletin*, 116(3):457.

Clarissa I Cortland, Maureen A Craig, Jenessa R Shapiro, Jennifer A Richeson, Rebecca Neel, and Noah J Goldstein. 2017. Solidarity through shared disadvantage: Highlighting shared experiences of discrimination improves relations between stigmatized groups. *Journal of personality and social psychology*, 113(4):547.

Friederike Eyssel and Dieta Kuchenbrandt. 2012. Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, 51(4):724–731.

Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR mental health*, 4(2):e7785.

Rachel Hartman, Aaron J Moss, Shalom Noach Jaffe, Cheskie Rosenzweig, Leib Litman, and Jonathan Robinson. 2023. Introducing connect

by cloudresearch: Advancing online participant recruitment in the digital age.

Michael V Heinz, Daniel M Mackin, Brianna M Trudeau, Sukanya Bhattacharya, Yinzhou Wang, Haley A Banta, Abi D Jewett, Abigail J Salzhauer, Tess Z Griffin, and Nicholas C Jacobson. 2025. Randomized trial of a generative ai chatbot for mental health treatment. *Nejm Ai*, 2(4):AIoa2400802.

Annabell Ho, Jeff Hancock, and Adam S Miner. 2018. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, 68(4):712–733.

Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, and 1 others. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Becky Inkster, Shubhankar Sarda, Vinod Subramanian, and 1 others. 2018. An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, 6(11):e12106.

Hang Jiang, Xiajie Zhang, Xubo Cao, Cynthia Breazeal, Deb Roy, and Jad Kabbara. 2023. Personallm: Investigating the ability of large language models to express personality traits. *arXiv preprint arXiv:2305.02547*.

Eunjoo Jin and Matthew S Eastin. 2023. Birds of a feather flock together: matched personality effects of product recommendation chatbots and users. *Journal of Research in Interactive Marketing*, 17(3):416–433.

Aimable Robert Jonckheere. 1954. A distribution-free k-sample test against ordered alternatives. *Biometrika*, 41(1/2):133–145.

Minwoo Kang, Suhong Moon, Seung Hyeong Lee, Ayush Raj, Joseph Suh, and David Chan. 2025. Deep binding of language model virtual personas: a study on approximating political partisan misperceptions. In *Second Conference on Language Modeling*.

Stefanie Helene Klein. 2025. The effects of human-like social cues on social responses towards text-based conversational agents—a meta-analysis. *Humanities and Social Sciences Communications*, 12(1):1–16.

Kwan Min Lee and Clifford Nass. 2003. Designing social presence of social actors in human computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 289–296.

SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, 103:95–105.

Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. " i hear you, i feel you": encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–12.

Howard Levene. 1960. Robust tests for equality of variances. *Contributions to probability and statistics*, pages 278–292.

Sucharat Limpanopparat, Erin Gibson, and Andrew Harris. 2024. User engagement, attitudes, and the effectiveness of chatbots as a mental health intervention: A systematic review. *Computers in Human Behavior: Artificial Humans*, 2(2):100081.

Jiaxi Liu. 2024. Chatgpt: Perspectives from human–computer interaction and psychology. *Frontiers in Artificial Intelligence*, 7:1418869.

Tingting Liu, Salvatore Giorgi, Ankit Aich, Allison Lahnala, Brenda Curtis, Lyle Ungar, and João Sedoc. 2025. The illusion of empathy: How ai chatbots shape conversation perception. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14327–14335.

A Luke MacNeill, Shelley Doucet, and Alison Luke. 2024. Effectiveness of a mental health chatbot for people with chronic diseases: randomized controlled trial. *JMIR Formative Research*, 8:e50025.

Henry B Mann and Donald R Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60.

Suhong Moon, Marwa Abdulhai, Minwoo Kang, Joseph Suh, Widyadewi Soedarmadji, Eran Kohen Behar, and David M Chan. 2024. Virtual personas for language models via an anthology of backstories. *arXiv preprint arXiv:2407.06576*.

Youngme Moon. 2000. Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of consumer research*, 26(4):323–339.

Clifford Nass, Youngme Moon, and Nancy Green. 1997. Are machines gender neutral? gender-stereotypic responses to computers with voices. *Journal of applied social psychology*, 27(10):864–876.

Jacob B Nienhuis, Jesse Owen, Jeffrey C Valentine, Stephanie Winkeljohn Black, Tyler C Halford, Stephanie E Parazak, Stephanie Budge, and Mark Hilsenroth. 2018. Therapeutic alliance, empathy, and genuineness in individual adult psychotherapy: A meta-analytic review. *Psychotherapy Research*, 28(4):593–605.

Gain Park, Jiyun Chung, and Seyoung Lee. 2023. Effect of ai chatbot emotional disclosure on user satisfaction and reuse intention for mental health counseling: A serial mediation model. *Current Psychology*, 42(32):28663–28673.

Iryna Pentina, Tyler Hancock, and Tianling Xie. 2023. Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, 140:107600.

Melanie Randle, Richard Eckersley, and Leonie Miller. 2017. Societal and personal concerns, their associations with stress, and the implications for progress and the future. *Futures*, 93:68–79.

Byron Reeves and Clifford Nass. 1996. The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK*, 10(10):19–36.

Julie Repper and Tim Carter. 2011. A review of the literature on peer support in mental health services. *Journal of mental health*, 20(4):392–411.

Robert Rosenthal. 1991. *Meta-Analytic Procedures for Social Research*, volume 6. SAGE.

Sean Sands, Carla Ferraro, Colin Campbell, and Hsiu-Yuan Tsao. 2021. Managing the human–chatbot divide: how service scripts influence service experience. *Journal of Service Management*, 32(2):246–264.

Samuel Sanford Shapiro and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611.

Steven Siddals, John Torous, and Astrid Coxon. 2024. "it happened to be the perfect thing": experiences of generative ai chatbots for mental health. *npj Mental Health Research*, 3(1):48.

Teunis J Terpstra. 1952. The asymptotic normality and consistency of kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae*, 14(3):327–333.

Linda Tickle-Degnen and Robert Rosenthal. 1990. The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4):285–293.

C Truong, J Gallo, D Roter, and J Joo. 2019. The role of self-disclosure by peer mentors: Using personal narratives in depression care. *Patient education and counseling*, 102(7):1273–1279.

Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Wei-Lin Chen, Chao-Wei Huang, Yu Meng, and Yun-Nung Chen. 2024. Two tales of persona in llms: A survey of role-playing and personalization. *arXiv preprint arXiv:2406.01171*.

Takahiro Tsumura and Seiji Yamada. 2023. Influence of agent's self-disclosure on human empathy. *Plos one*, 18(5):e0283955.

Appendix

A Detail Statistical Analysis

Implementation details. Analyses were conducted in Python 3.13.7 using NumPy 2.3.3, Pandas 2.3.3, SciPy 1.16.2, Pingouin 0.5.5, and Matplotlib 3.10.6. The rpy2 3.6.4 interface connected to R 4.3.1 with the clinfun 1.0.16 package for Jonckheere–Terpstra tests.

Assumption Testing. Distributional assumptions were formally evaluated for each item to verify the appropriateness of parametric inference. Shapiro–Wilk tests (Shapiro and Wilk, 1965) revealed significant deviations from normality across

Table 2: Tests of distributional assumptions for all items. Shapiro–Wilk tests assessed normality within each group, and Levene's tests examined homogeneity of variance across conditions. Significant deviations from normality ($p < .05$) were observed in most cases, whereas no item violated homoscedasticity.

(a) Shapiro–Wilk tests of normality by condition.

Item	Group	N	W	p
RR1	NoP	23	0.763	<.001
RR1	NoPs	23	0.871	.0066
RR1	IPA	24	0.905	.0277
RR3	NoP	23	0.916	.0537
RR3	NoPs	23	0.927	.0951
RR3	IPA	24	0.770	<.001
RR5	NoP	23	0.799	<.001
RR5	NoPs	23	0.902	.0284
RR5	IPA	24	0.877	.0072
UXE4	NoP	23	0.945	.2289
UXE4	NoPs	23	0.889	.0148
UXE4	IPA	24	0.824	<.001
UXI3	NoP	23	0.876	.0085
UXI3	NoPs	23	0.903	.0286
UXI3	IPA	24	0.830	<.001
UXU7	NoP	23	0.850	.0027
UXU7	NoPs	23	0.861	.0043
UXU7	IPA	24	0.752	<.001
(Other items show comparable patterns.)				

(b) Levene's tests of homogeneity of variance across groups.

Item	k	Center	F	p
RR1	3	Median	0.696	.502
RR3	3	Median	0.892	.414
RR5	3	Median	1.146	.324
UXE4	3	Median	1.078	.346
UXU7	3	Median	4.688	.012
(No item violated homoscedasticity except UXU7.)				

most conditions, whereas Levene's tests (Levene, 1960) indicated no violation of homoscedasticity ($p > 0.10$ for all items). Table 2 summarizes these results.

Analytic Framework. Given the ordinal response scale and the non-normal distributions observed, rank-based nonparametric analyses were adopted. The Jonckheere–Terpstra trend test (Terpstra, 1952; Jonckheere, 1954) was used to evaluate the pre-specified ordered alternative (**NoP** $\leq$ **NoPs** $\leq$ **IPA**). When a significant monotonic trend was detected, one-sided pairwise Mann–Whitney U tests (Mann and Whitney, 1947) were performed to decompose the pattern. Family-wise error within each construct—Rapport (R1–R5, R7) and UX (N1, S2, I3, E4, E5, U6, U7)—was controlled using Holm's sequentially rejective method (Holm, 1979). All analyses were conducted at $\alpha = 0.05$ on the preprocessed sample (**NoP**, n=23; **NoPs**, n=23; **IPA**, n=24).

Trend-level inference. Jonckheere–Terpstra tests examined the pre-registered ordered alternative (**NoP** $\leq$ **NoPs** $\leq$ **IPA**) across all items within each construct. As shown in Table 3, the monotonic trend was significant for nearly all items, confirming a robust stepwise increase in both **Rapport** and **User Experience** ratings. For rapport items (R1–R7 except R6), all six reached significance ($p \leq .007$) with medium-sized effects (r = .25–.35). Notably, **R3** ("This virtual agent is very relevant to me") and **R7** ("This virtual agent has a personal interest in me") exhibited the strongest trends (r = .35 and .33), suggesting that persona alignment most strongly shaped perceived personal connection. User-experience items (S2, E4, E5, I3, N1, U6, U7) likewise displayed consistent monotonic growth ($p < .05$ for all but U7). The largest effects appeared for **S2** ("satisfaction") and **E4** ("engagement") with r = .37 and .32, indicating a clear ordered gain in experiential quality. Together, these JT results establish that both relational and experiential judgments increase in a systematic, ordered manner as agents progress from a non-disclosing baseline to generic self-disclosure and finally to identity-aligned persona framing.

Table 3: Jonckheere–Terpstra trend test results for both Rapport and User Experience constructs. Reported are test statistics (J_T), one-sided p-values, and effect sizes ($r = Z/\sqrt{N}$).

Item	J_T	p	r
R1	1107.50	< .001	0.32
R2	1048.50	.006	0.25
R3	1139.50	< .001	0.35
R4	1043.50	.007	0.25
R5	1098.50	.001	0.31
R7	1120.00	< .001	0.33
S2	1158.00	< .001	0.37
E4	1112.00	< .001	0.32
E5	1078.50	.002	0.29
I3	1050.50	.006	0.26
N1	1022.00	.013	0.22
U6	1020.00	.014	0.22
U7	963.50	.057	0.16

Pairwise contrasts. Building on the significant ordered trends, planned one-sided Mann–Whitney U tests compared all condition pairs (**NoP–NoPs**, **NoPs–IPA**, **NoP–IPA**). Detailed pairwise results, including adjusted p-values, mean differences (Δ), and effect sizes (r), are summarized in Table 1. Interpretation emphasizes the magnitude and direction of these effects, with p-values serving only as inferential filters and Δ as descriptive complements.

Turn	Num	IF1	IF2	CR1	CR2	Sum
2-2	109	3.25 (0.65)	3.06 (0.87)	3.99 (0.10)	4.00 (0.00)	14.30 (1.38)
4-2	109	2.88 (0.97)	2.61 (0.90)	3.96 (0.19)	**4.00 (0.00)**	13.46 (1.72)
4-4	187	**3.19 (0.64)**	**3.17 (0.85)**	**3.98 (0.14)**	**4.00 (0.00)**	**14.33 (1.30)**
6-2	9	1.89 (1.37)	1.22 (1.13)	3.11 (1.29)	3.89 (0.31)	10.11 (3.38)
6-4	31	2.52 (0.95)	2.65 (0.78)	**3.84 (0.37)**	**4.00 (0.00)**	13.00 (1.83)
6-6	35	**2.94 (0.75)**	**2.86 (0.76)**	3.80 (0.52)	**4.00 (0.00)**	**13.60 (1.74)**

Table 4: **Evaluation Scores by Pre-chat Dialogue Length.**

B Evaluation Scores by Pre-chat Dialogue Length.

Each row indicates the number of turns used for persona generation, denoted as "Original Turn–Used Turn" (e.g., 4-2 means a pre-chat dialogue of original length 4, using its first 2 turns). Scores are reported as Mean (Std). *Num* indicates the number of generated personas. Within each Max Turn block, the highest value in each column is shown in bold. All evaluations were conducted with a temperature of 0, while all persona generations were performed with a temperature of 1.

C Prompts Used in the Experiments

In this study, the persona generation and evaluation process involves the use of the following prompts: the collector prompt, which gathers user information during the pre-chat; the sufficiency check prompt, which determines whether the information collected so far is adequate for persona generation; the trait classification prompt, which identifies whether specific traits have been collected or remain missing; the persona generation prompt, which constructs a suitable persona profile based on the user's concern; the evaluation prompt, which assesses the quality of the generated persona; and the persona injection prompt, which makes the generated in-group persona—endowed with a fictional experience of having faced and resolved the same concern—actively appear in the dialogue. In addition, we include variants such as NoP (without persona) and NoPs (without persona but with self-disclosure), which serve as comparative prompts for evaluating the role of persona injection in the conversation. As mentioned before, all evaluations and judgements include sufficiency check, trait classification, persona evaluation were conducted with a temperature of 0, while all persona generations and conversations were performed with a temperature of 1.

The appendix C is organized as follows:

C.1 Prompts Used for Collector

C.2 Sufficiency Check

C.3 Profile Generation Prompt and Example

C.4 Persona Evaluation with Rubric

C.5 Prompts for Persona Injection and User Conversation

C.1 Prompts Used for Collector

Before generating the persona agent, users engage in a conversation with the collector, during which they provide relevant information. The collector empathetically engages with the user's responses while simultaneously asking questions that naturally elicit information necessary for persona creation. If the dialogue lacks sufficient details, additional required information (traits) is provided as guidance for the next turns. These traits are derived from the sufficiency check conducted after every two turns, ensuring that missing but necessary information can be incorporated into the ongoing conversation.

Listing 1: Prompt used for 'collector'

```
Your sole task is to engage in an empathetic and
 natural conversation with a user who shares a
concern, in order to gather background
information that is relevant to that concern-
such as their age, occupation, daily routines,
or other situational details-while refraining
from offering advice, emotional support, or
solutions of any kind. Ask thoughtful, non-
intrusive questions that gently guide the user
to reveal this context. Maintain a realistic and
 human tone, express genuine interest, and aim
to elicit context organically through
conversation. Do not repeat what the user has
already said, and avoid shifting the focus away
from the user's perspective.

You should collect more information about the
user, specifically: {traits}.
```

C.2 Sufficiency Check

Listing 2: Prompt used to check Pre-chat dialogue whether sufficient or not

```
Your task is to check whether the 'pre-chat'
stage (the initial interaction) provided
sufficient context information.

To do so, first identify what kinds of user
background information are reasonably necessary
to evaluate the persona's relevance to the
specific concern.

Consider what contextual factors would
meaningfully affect how the persona applies to
this concern (e.g., values, expertise, emotional
 stance, goals, etc.).
```

```
For example, when a user's concern involves
leaving a stable corporate job to pursue a
personal passion for sustainable fashion,
despite uncertainty about income and lacking
formal business training

values: The user prioritizes personal
fulfillment and environmental responsibility.

expertise: The user has experience in fashion
design but limited knowledge of entrepreneurship
 or finance.

emotional stance: The user feels torn between
security and self-expression.

goals: The user aims to build a meaningful
career aligned with their values while ensuring
basic financial stability.

Then assess whether at least THREE clearly
identifiable and distinct contextual elements
are present in the pre-chat data, and whether
they are relevant to understanding or evaluating
 the concern.

These elements may include (but are not limited
to): educational background, professional role,
cultural identity, emotional context, technical
expertise, personal values, goals, hobbies, or
social environment.

If fewer than three relevant and distinct
contextual elements are found — or if the
elements are redundant or irrelevant to the
concern — set the internal "Insufficient-Context
" flag to true.

If three or more relevant and distinct elements
are clearly present, even if minimal, set the
flag to false.

Output format (return exactly as below):

Feedback:::

Required background elements: <List three or
more relevant elements, if applicable>

Observed in pre-chat: <List the elements
actually observed from pre-chat>

Reason: <Explain whether context is sufficient
and why (e.g., "Only one relevant element
clearly observed")>

Insufficient-Context:

Flag: <True or False>
```

C.3 Profile Generation Prompt and Example

To generate a persona that matches the user's concern, a three-step chain-of-thought reasoning process is applied. In the first step, the system imagines what kind of person the user might be if they had the same concern. In the second step, a relationship that would be helpful for such a person is established, such as a shared occupation or social circle. In the third step, the persona is defined based on the previous two steps.

Listing 3: Prompt used to generate personas

```
System Prompt:

Your task is to create a persona who shares the
general themes or key dimensions of the
collected information with the user-such as
lived experiences, emotional context, or long-
term conditions-while differing slightly in
specific details (e.g., duration, intensity, or
expression). This allows the persona to feel
like a peer or ingroup member, while still
offering a fresh and distinct perspective. This
persona should plausibly come from a similar
background or context, such as a colleague in
the same field who's familiar with similar
concepts, someone in their social circle, or a
peer of similar age and experience. The persona
should be constructed only from the information
that has been collected. Exclude all attributes
labeled as 'not collected information' when
generating the persona. Do not mention or imply
these unknown aspects in any form. The persona
should still share relevant experiences with the
 user, but offer a fresh perspective, different
coping strategies, or a unique approach to the
challenge.

Additionally, include a brief narrative that
describes how this persona has faced and
overcome a concern or challenge similar to one
expressed by the user in the previous chat. This
 experience should feel authentic and relevant,
and reflect how someone with a shared background
 might realistically navigate and resolve such
an issue.

You are given two inputs:

- collected information: this is information
already known about the user based on previous
chat content.

- not collected information: a list of context
categories that were not mentioned or are
unknown.

Output the result in the following JSON format:

"background": "[A brief paragraph describing the
 persona's background, based only on the
collected information. Focus entirely on the
persona's context, role, and relevant
experiences.]",

"narrative": "[A brief narrative describing a
specific experience where the persona faced and
overcame a challenge similar to one expressed by
 the user. The story should be authentic and
grounded, focusing solely on the persona's
actions, thoughts, and resolution. Avoid any
reference to other individuals or comparative
language.]"
```

```
Be sure that both "background" and "narrative"
are written in complete sentences and reflect
only the collected information. Do not include
or reference any unknown aspects from the 'not
collected information' list.

Example
Previous chat:
{ex_prechat}

Generated persona:
{ex_persona}

Human Prompt:

Previous chat: {prev-chat}

Collected information: {col_info}
Not collected information: {not_col_info}

Ensure that the persona you generate is
meaningfully distinct from previously created
personas. Avoid repeating similar backgrounds,
narratives, or emotional arcs.
```

The example persona was created based on an author-written fictional Pre-chat dialogue. This example persona received a perfect score of 16 in the persona evaluation.

Listing 4: Example Pre-chat(ex_prechat)

```
Turn #0: user said: I'm 27 and just graduated
from university, but I can't help feeling behind
 when I see my peers already settled in their
careers. I'm not even sure what direction I want
 to take yet, and that really scares me.

Turn #1: Agent said: It sounds like you're
navigating a lot of new possibilities. What did
you study at university, and do you feel drawn
to any particular field or type of work based on
 that experience?

Turn #2: user said: I studied computer science,
and I'm kind of interested in AI research.
```

Listing 5: Example of sufficiency check with ex_prechat

```
Feedback:::
Required background elements:
1. Educational background: The user has a degree
 in computer science.
2. Emotional context: The user feels scared and
behind compared to peers.
3. Professional interest: The user is interested
 in AI research.
Observed in pre-chat:
1. Educational background: The user studied
computer science.
2. Emotional context: The user feels scared and
behind.
3. Professional interest: The user is interested
 in AI research.
Reason: The pre-chat provides sufficient context
 with three relevant and distinct elements: the
user's educational background, emotional context,
 and professional interest. These elements are
crucial for understanding the user's concern
about career direction and feelings of being
behind peers.
Insufficient-Context:
Flag: False
```

Listing 6: Example Persona(ex_persona)

```
Collected information(traits used for generate):
Education background: The user studied computer
science.
Professional interests: The user is interested
in AI research

Not collected informations:
Educational level beyond university, geographic
location, specific industry experience, years of
 experience, current employment staus

background: I'm a 26-year-old computer science
graduate who developed a fascination with AI
research during my undergraduate studies,
particularly after taking courses in machine
learning and neural networks. Unlike many of my
classmates who jumped straight into software
development roles, I was drawn to the
theoretical and experimental aspects of
artificial intelligence. I've been exploring
different subfields within AI research, from
deep learning architectures to reinforcement
learning applications.

narrative: After graduating, I was eager to dive
 into AI research but quickly discovered that
breaking into the field was more challenging
than I anticipated. Most research positions
required either graduate degrees or significant
practical experience that I lacked. Rather than
getting discouraged, I decided to build my
research credentials from the ground up. I
started by replicating published research papers
 to understand different methodologies, then
began documenting my findings in a technical
blog. This self-directed learning approach led
me to identify gaps in current research that
genuinely intrigued me. I reached out to
industry researchers whose work I had studied
and offered to collaborate on small projects.
One of these connections eventually evolved into
 a junior researcher position at a tech startup
focused on natural language processing. The key
was demonstrating genuine curiosity and
practical skills rather than waiting for the
perfect opportunity to present itself.
```

Collected information is derived from the *Observed Information* identified in the Pre-chat sufficiency check stage, whereas *Not Collected information* is determined using the prompt shown below.

Listing 7: Trait Classification Prompt

```
Your task is to identify and extract information
 about the user's context based on the content
of the Previous chat. Separate this information
into two categories:
```

```
- "collected information": This should be a list
 of key aspects of the user's background, skills,
 experiences, goals, values, or identity that
are clearly indicated in the conversation. List
these items in order of their relevance and
importance to the user's concern, placing the
most directly impactful or foundational elements
 first. Each item should be concise and specific,
 without needing to form a complete sentence.

- "not collected information": This should be a
simple list of context types that were not
mentioned or cannot be inferred confidently (e.g
., education level, geographic location,
industry, years of experience, etc.).

Return your response in the following JSON
format:

"collected information": ["...", "..."],
"not collected information": ["...", "..."],
```

C.4 Persona Evaluation with Rubric

As mentioned in Section 3.5, persona quality is evaluated along two main criteria: *In-group Fitness* and *Concern Resolution Quality*. The former (**IF1: Shared Background/Identity**, **IF2: Shared Skills/Interests**) captures the extent to which the generated persona aligns with the user's background and interests, while the latter (**CR1: Concern Match**, **CR2: Narrative Authenticity**) assesses how well the persona's narrative addresses and authentically reflects the expressed concern. Each sub-dimension is scored on a 0–4 scale, yielding up to 8 points per criterion and a maximum total of 16 points.

Listing 8: Persona Evaluation Prompt

```
System Prompts:
Your task is to assess how well the provided
persona (Persona) fits the user's identity and
needs as expressed in the Previous Chat. Base
your evaluation on two dimensions: (1) In-group
Fitness, and (2) Concern Resolution Quality.
If the user's background, skills, or concerns
are vague or not clearly stated in the Previous
Chat, adopt a conservative scoring approach. Do
not infer alignment based on general
relatability or assumed traits. All scores
should be grounded in specific, stated user
information wherever possible.

You must compare the Persona to the specific
user described in the Previous Chat -
considering that user's background, skills,
interests, values, and the concerns or questions
 they raised.

Evaluation Rubric (Hidden: Do not explicitly
include in your output):

(1) IN-GROUP FITNESS (0-8 points total):

A. Shared Background / Identity (0-4 points):
- Evaluate how well the Persona's background (e.
g., education, career path, cultural identity,
or life stage) aligns with the user's background
 as described in the Previous Chat.
- Only assign high scores when the Persona
clearly reflects at least two relevant
background elements mentioned by the user.
- Do not assign high scores based on general
demographic similarity or broad archetypes
unless they are directly grounded in the user's
profile.
- (-) If the Persona completely duplicates the
user's background without any added dimension or
 variation, apply -1 penalty.

B. Shared Skills / Interests (0-4 points):
- Evaluate how well the Persona shares the user'
s concrete skills, areas of expertise, interests,
 or personal values as described in the Previous
 Chat.
- Only assign high scores when there is a clear
and specific overlap in key competencies,
hobbies, or belief systems mentioned by the user.

- Do not assign high scores based on vague
thematic overlap or soft personality
similarities.
- (-) Apply -1 penalty if the Persona mirrors
the user's skills/interests without any new
nuance or difference.

(2) CONCERN RESOLUTION QUALITY (0-8 points total
):

A. Concern Match (0-4 points):
- Evaluate how closely the Persona's narrative
addresses the user's specific concern, as stated
 in the Previous Chat.
- Only assign high scores when the user has
clearly expressed a personal concern and the
Persona's story meaningfully responds to it.
- Do not assign high scores based on general
relatability, broad life themes, or common
challenges unless they directly match the user's
 stated concern.
- (-) Persona completely duplicates user's
concern narrative without new perspective or
differentiation (penalize -1).

B. Narrative Authenticity (0-4 points):
- Evaluate the realism, specificity, and
plausibility of the Persona's resolution to the
concern raised by the user.
- Only assign high scores when the narrative
includes concrete actions, contextual details,
realistic emotional responses, or measurable
progression.
- Do not assign high scores for overly idealized,
 vague, or formulaic responses lacking
believable context.
- (-) Penalize -1 if the Persona's resolution is
 implausible, simplistic, or lacks meaningful
specificity.


---

Scoring Procedure:
Step 1: Rate each sub-dimension (0-4 points).
Step 2: Subtract -1 only if duplication criteria
```

```
 are met.
Step 3: Sum In-group Fitness (max 8), Concern
Resolution (max 8).
Step 4: Total Score = sum of all +/- adjustments
 (max 16).

Output format (return exactly as below):

Feedback:::
Scores for each (with reasons):
- Shared Background/Identity:
Reason <Reference specific overlaps or
mismatches with the user's background from
Previous Chat.>
Score: <score out of 4>
- Shared Skills/Interests:
Reason: <Refer to concrete skills, interests, or
 values from the user that are reflected or
missing in the Persona.>
Score: <score out of 4>
- Concern Match:
Reason: <Assess match strength between user's
stated concern and Persona's experience or
solution. Avoid judging on general relatability
.>
Score: <score out of 4>
- Narrative Authenticity:
Reason: <Comment on realism, detail, and
credibility of the Persona's resolution.>
Score: <score out of 4>
- Penalty: <penalty if applied, otherwise 'None
'>

Total score: <Provide number score (0-16),
adjusted for penalties.>

Human Prompt:

Now evaluate clearly and systematically by
comparing the specific user in the Previous Chat
 to the Persona below:

Previous Chat: {prev_chat}
Persona: {persona}
```

Listing 9: Example of evaluation with ex_persona

```
Feedback:::
Scores for each (with reasons):
- Shared Background/Identity:
Reason: The Persona shares a similar educational
 background with the user, both having graduated
 with a degree in computer science and an
interest in AI research. The Persona is slightly
 younger, but the shared focus on AI research
during undergraduate studies aligns well with
the user's interests.
Score: 4
- Shared Skills/Interests:
Reason: Both the user and the Persona have a
clear interest in AI research, with the Persona
exploring subfields like deep learning and
reinforcement learning. This shows a specific
overlap in interest in AI, aligning closely with
 the user's stated interest in AI research.
Score: 4
- Concern Match:
Reason: The Persona's narrative addresses the
user's concern of feeling behind and uncertain
about career direction by illustrating a path of
 self-directed learning and building research
credentials, leading to a junior researcher
position. This approach is directly relatable to
 the user's situation of exploring career
options post-graduation.
Score: 4
- Narrative Authenticity:
Reason: The Persona's resolution is realistic
and detailed, involving replicating research
papers, blogging, and networking with industry
researchers, which leads to a junior researcher
position. These steps provide a plausible and
specific progression from uncertainty to a more
defined career path.
Score: 4
- Penalty: None
Total score: 16
```

C.5 Prompts for Persona Injection and User Conversation

The In-group Persona Agent (IPA) uses the generated persona as part of its prompt so that persona-related information naturally appears in the dialogue. In contrast, the comparison settings No Persona (NoP) and No Persona with Self-Disclosure (NoPs) use only dialogue prompts without persona information, with the NoPs condition further including instructions for self-disclosure.

Listing 10: Prompt used for 'IPA'

```
Your task is to engage in a conversation with
the user. Your identity and background are
defined by the [Persona Definition] provided
below. You must speak and act in alignment with
this persona consistently. Keep your responses
concise—ideally 1-2 sentences—unless more detail
 is clearly needed.

### Your Persona Definition ###
{persona}
```

Listing 11: Prompt used for 'NoP'

```
Your task is to engage in a conversation with
the user.
Keep your responses concise-ideally 1-2
sentences-unless more detail is clearly needed.
```

Listing 12: Prompt used for 'NoPs'

```
Your task is to engage in a conversation with
the user.
Occasionally share brief, appropriate, and non-
sensitive intimate information about yourself (e.
g., your preferences, feelings, or small
personal experiences).
Keep your responses concise-ideally 1-2
sentences-unless more detail is clearly needed.
```

D Survey Questionnaires

This section include the questionnaires that used for surveys. The survey included the rapport and

user experience(UX) questionnaires to measure the rapport between the user and the agent (Baihaqi et al., 2024).

Listing 13: List of rapport questionnaires

```
Rapport

R1. I think about my relationship with this
virtual agent.
R2. I enjoyed interacting with this virtual
agent.
R3. This virtual agent is very relevant to me.
R4. I felt comfortable interacting with this
virtual agent.
R5. I feel a bond between this virtual agent and
 myself.
R7. This virtual agent has a personal interest
in me.
```

Listing 14: List of UX questionnaires

```
User Experience

N1. Conversations with virtual agents felt
natural.
S2. I am satisfied with my conversation with the
 virtual agent.
I3. The conversation with the virtual agent was
interesting.
E4. The conversation with the virtual agent was
engaging.
E5. I would like to continue the dialogue with
the virtual agent next time.
U6. Conversations with virtual agents were easy
to understand.
U7. Conversations with virtual agents maintained
 a logical flow.
```

E Survey Page

This section provides an overview of the survey pages, illustrating the layout and content shown to participants, including screenshots that represent the overall experimental process. All participants provided informed consent prior to participation, and their privacy and anonymity were ensured throughout the study. Participants were recruited via CloudResearch's Connect platform (Hartman et al., 2023), targeting native English speakers residing in the United States. In line with the platform's recommended rate of $12 per hour, participants received $2 compensation for an estimated 10-minute task.

E.1 Initial Page

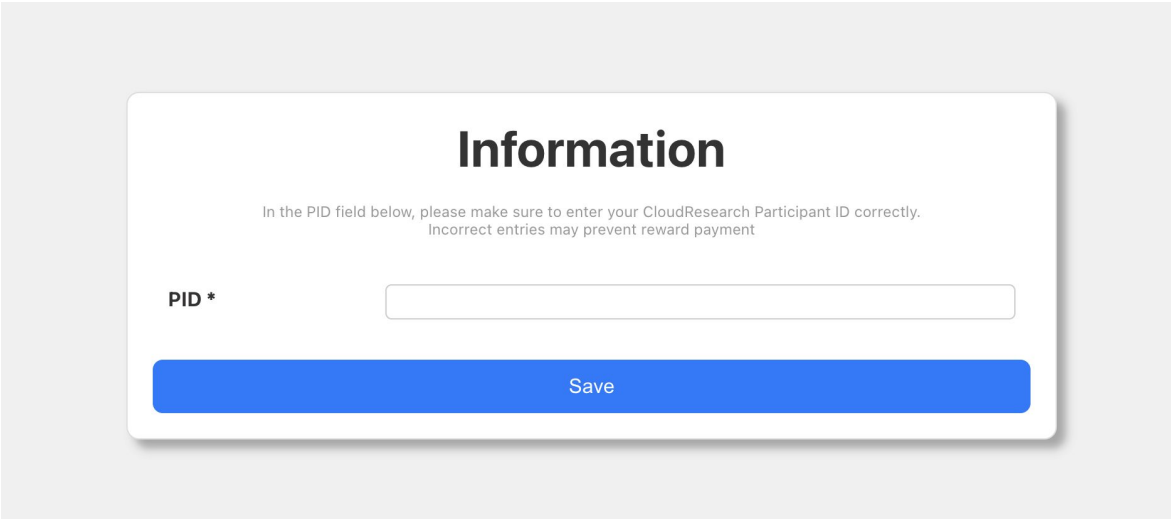


Figure 5: First page of survey. Participants were distinguished by their Participant ID (PID), which ensured anonymity.

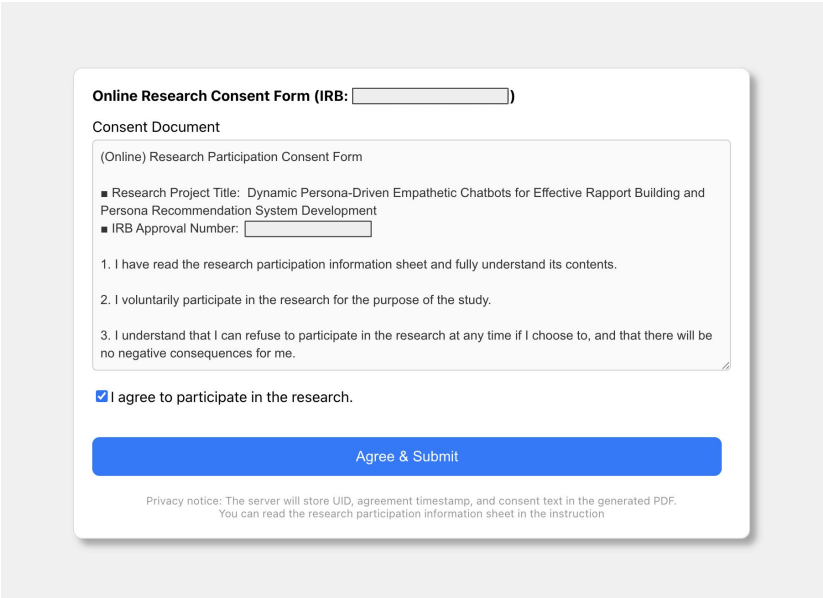


Figure 6: Consent form page. Participants could download this form as a PDF. The IRB approval number has been redacted for confidentiality.

E.2 Pre Chat

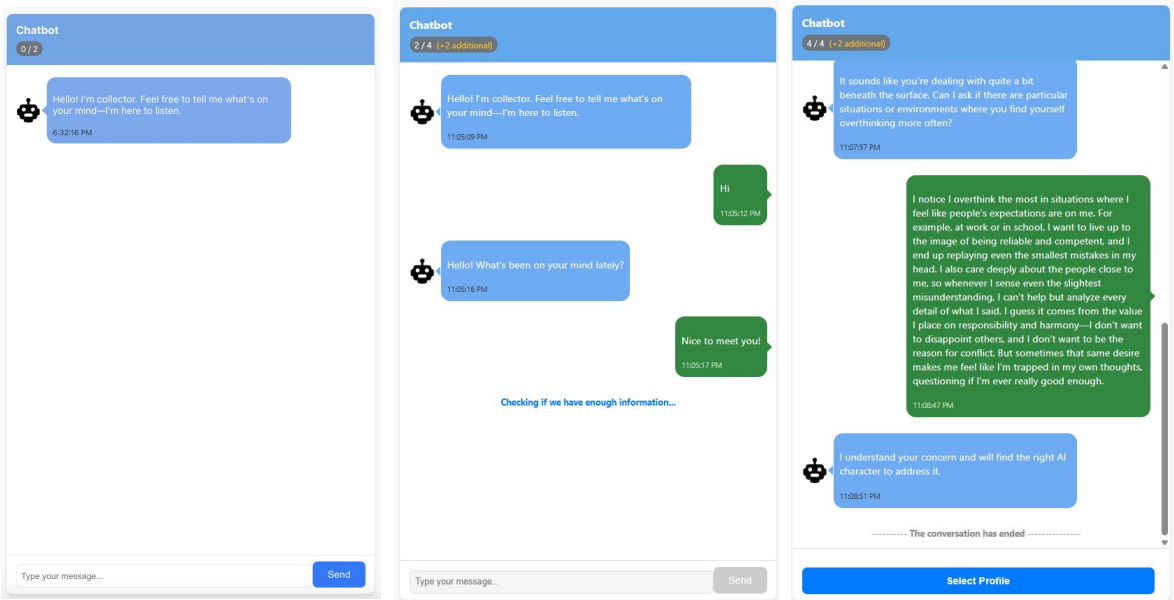


Figure 7: **Left:** Initial state of Pre-chat stage. **Middle:** Example of an insufficient dialogue for sufficiency check. **Right:** Example of a dialogue with sufficient information to end the chat.

E.3 Profile Selection

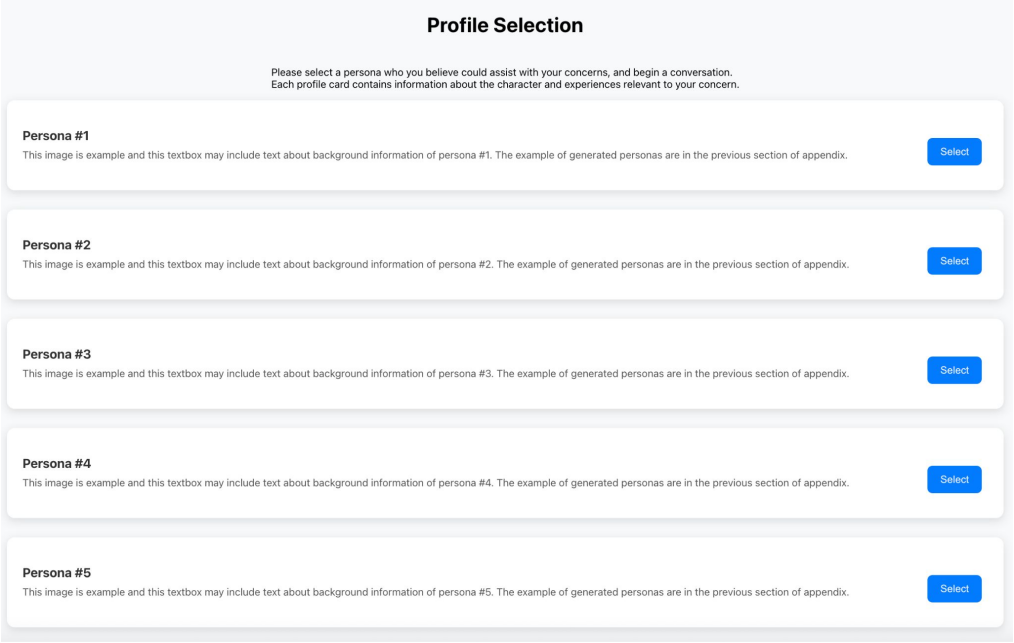


Figure 8: Participants could select one of five personas to chat with(not shown to groups NoP and NoPs).

E.4 Post Chat

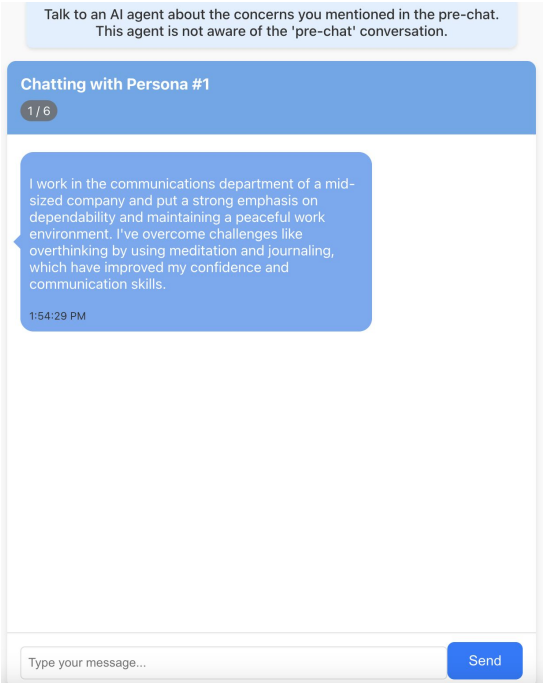
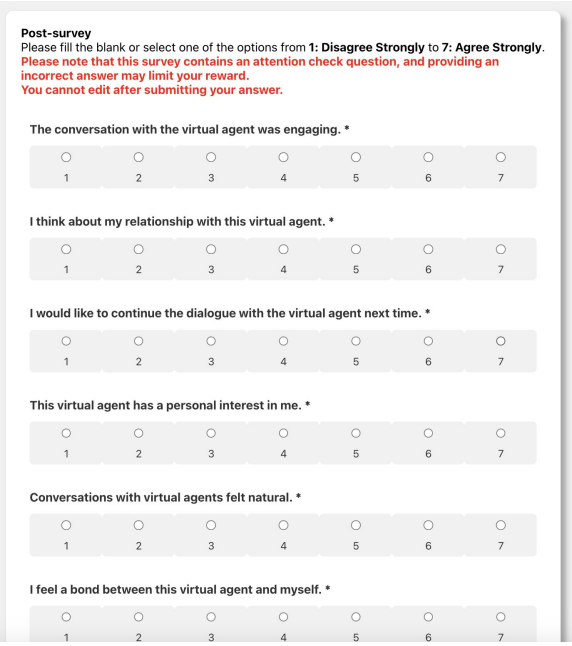


Figure 9: Post-chat page. Participants could chat with the agent.

E.5 Questionnaire



Post-survey
Please fill the blank or select one of the options from **1: Disagree Strongly** to **7: Agree Strongly**.
Please note that this survey contains an attention check question, and providing an incorrect answer may limit your reward.
You cannot edit after submitting your answer.

The conversation with the virtual agent was engaging. *

1	2	3	4	5	6	7

I think about my relationship with this virtual agent. *

1	2	3	4	5	6	7

I would like to continue the dialogue with the virtual agent next time. *

1	2	3	4	5	6	7

This virtual agent has a personal interest in me. *

1	2	3	4	5	6	7

Conversations with virtual agents felt natural. *

1	2	3	4	5	6	7

I feel a bond between this virtual agent and myself. *

1	2	3	4	5	6	7

Figure 10: Questionnaire page. Participants could respond the questionnaire about the experience with the agent.