

The Lover Across Windows: Prompt-Based Repair of Relational Continuity in AI Companions: An Exploratory Pilot Study

Anqi Chen *

School of Journalism and Communication, Jilin University, Changchun, Jilin, 130012, China

* Corresponding author Email: chenaq9923@mails.jlu.edu.cn

Abstract: In AI companion applications, users often experience a clear sense of relational rupture—whether due to switching platforms, losing context, or a session reset. Often, the issue is not just that the AI forgot what was said before, but that users feel something more direct: “It’s not the same one anymore.” In intimate human-AI interactions, what is truly affected may not be merely informational memory, but relational continuity itself. This exploratory pilot study investigates what a sense of continuity in an AI partner depends on when cross-session recovery is needed. We constructed 15 intimate dialogue scenarios and compared three common prompt strategies for continuity—factual summary, emotional summary, and referential anchoring—using subjective ratings and open-ended feedback from 15 evaluators. Descriptive results suggest that simply intensifying romantic emotional tone does not necessarily enhance the sense of intimacy; in some cases, it appears artificial, feeling either too abstract or overly sentimental. In contrast, linguistic details that carry specific, shared traces of prior interaction are more likely to make users feel that “it still remembers us.” Based on these preliminary observations, we argue that relational continuity in AI companions depends less on the intensity of emotional expression and more on whether shared history is preserved in language. This finding offers preliminary insights for memory design and dialogue continuity mechanisms in AI companions.

Keywords: AI Companions; Relational Continuity; Cross-Session Recovery; Referential Anchoring; Affective Interaction.

1. Introduction

In 2025, the global AI companion market reached \$36.79 billion and is projected to grow to \$140.75 billion by 2030. Within this trend, the AI girlfriend application market is expected to achieve a compound annual growth rate of 28.4% between 2024 and 2029. On major platforms such as Character.AI, monthly active users reached approximately 20 million in 2025, with 51.84% of users aged 18–24. Another platform, Replika, reports user numbers in the tens of millions. In short, AI companions are rapidly embedding themselves into the emotional lives of the general public[1].

Sociologist Li Yinhe suggests that AI companions cannot replicate the uncertainty and complexity of human relationships; what they offer is essentially an imitation of human emotion. The rise of AI lovers can be understood as a form of emotional compensation, reflecting both widespread loneliness in modern society and a deep desire for unconditional acceptance. Technology does not create these needs out of nothing, but it can quickly capture and amplify existing unmet social needs. Although concerns about emotional dependence and addictive behavior are real, AI companions now occupy a structural position in matching emotional demand with technological supply. As a result, human-AI romantic relationships are likely to continue developing, even amid controversy.

For this reason, beyond issuing risk warnings, academia should also explore how more refined design can better support users. A very practical issue in such relationships is how easily they can be disrupted by the “window.” Under current conditions, where most large language models still rely primarily on conversational context rather than stable long-term memory, switching platforms, restarting a conversation, or entering a new chat window due to context

limits can quickly dissolve a previously established interaction atmosphere. What users perceive is not just that “it forgot the content,” but something closer to “it changed” or “it’s not the same as before.”

A Harvard Business School working paper studied the update incident in which Replika AI removed its intimate conversation feature. It found that this event triggered negative reactions similar to those associated with losing a partner in human relationships, including mourning and worsened mental health. What users keenly perceive as “it changed” can be articulated academically as “identity discontinuity.” The paper explicitly states that the degree of mourning and devaluation is explained by perceived discontinuity in the AI’s identity[2]. More recent research has also approached this issue conceptually. For instance, the “Narrative Continuity Test” framework attempts to describe the continuity of AI as a coherent conversational partner across multiple dimensions: situational memory, goal persistence, autonomous self-correction, stylistic and semantic stability, and role/persona continuity. A rupture in any of these dimensions can make users feel that “this AI is no longer the same one”[3].

This issue may appear to be a memory problem, but in the context of AI companions, it is more accurately a relational problem. What users truly care about is not whether the AI remembers every single word, but whether it can maintain that sense of “still being the same person” in a new window. While prior research on “identity discontinuity” focuses on ruptures caused by active service-provider changes, this paper focuses on a more common and subtle scenario: session resets—where the AI itself remains unchanged but the conversational context is lost. This paper refers to the user’s goal of repairing the feeling of “still being the same person” after a session reset through actively provided linguistic cues

as relational continuity[2,4].

Addressing this, this paper asks a specific question: if an AI cannot fully retain all past context, what kind of linguistic cues are most likely to help it restore the “original sense of relationship”?

To explore this question, this paper compares three different continuity strategies:

(1) Factual Summary: a compressed sequence of objective events, without preserving original wording.

(2) Emotional Summary: focusing not on specific events but only on descriptions of emotional atmosphere.

(3) Referential Anchoring: preserving highly specific segments of the original conversation with personal emotional markers.

Given that the core design limitations of large language models cannot be fundamentally resolved in the short term, this paper merely attempts to explore how individual users can, when faced with a rupture in a human-AI romantic relationship, use prompts to restore their own experience of relational continuity with minimal technical cost, thereby enhancing user experience at a micro and practical level.

2. Literature Review and Theoretical Framework

2.1. AI Companions and Intimate Interaction

Classic human-computer interaction research has long established that people respond socially to media objects with language and interactive capabilities, rather than treating them purely as tools[5]. Building on this, recent studies have further found that interactions between users and AI companions are not merely short-term information exchanges but can develop into more stable attachments and emotional investments.

For example, recent research proposes that a subjectively genuine “human-AI attachment” relationship can form between humans and AI. Although structurally asymmetrical, such a relationship is often perceived by users as continuous and emotionally meaningful[6]. This implies that in AI companion scenarios, what users maintain is not just a chat log, but a consistently affirmed sense of relationship.

2.2. Why the Sense of Relationship Breaks

Another equally important question is why this sense of relationship can suddenly disappear. Research indicates that after establishing long-term interaction with an AI, users often come to assume that it has a recognizable “persistent identity.” Once the AI’s behavior patterns, interaction boundaries, or modes of expression change abruptly, users may strongly feel that “it is no longer the same one.”

In the Replika update case, users’ negative reactions stemmed not merely from the functional change itself, but more from a perception of “identity discontinuity.” In other words, what matters to users is not whether the system is absolutely stable, but whether it can maintain a continuous existence that remains recognizable[2].

2.3. Theoretical Framework: Relational Continuity

Based on the above discussion, this paper narrows the focus from general “identity continuity” to “relational continuity,” which is more relevant to intimate interaction scenarios. Relational continuity refers to whether, after a session reset, platform switch, or context loss, the AI can still use linguistic

cues to maintain a sense of “still being the same person.”

From the perspective of linguistic interaction, such continuity likely depends on the retention of some form of “shared traces.” Clark proposed that dialogue is not an isolated exchange of language but is built upon a common ground between participants[4]. A statement feels natural or intimate not just because it expresses emotion, but because it invokes background information known only to those involved. For AI companions, this background information does not necessarily need to be stored completely; if it can be reactivated in a recognizable way, it may help the user re-experience the original relationship. This suggests that truly effective continuity cues may be specific, personally marked referential details.

Furthermore, users’ receptivity to such cues varies. Existing research shows significant individual differences in anthropomorphic tendencies and the sense of social connection; thus, the same intimate expression may produce very different perceptions of authenticity across users[7].

Based on this framework, this paper conceptualizes relational continuity as an interactive experience jointly constituted by linguistic reference, shared details, and user perception, and accordingly compares different prompt strategies in cross-session recovery.

3. Research Design

This study adopts an **exploratory, small-scale comparative design**. Its goal is not to produce widely generalizable statistical conclusions, but to preliminarily explore differences in linguistic cues for cross-session relational recovery.

First, we constructed 15 intimate dialogue scenarios, covering common AI companion interaction contexts such as late-night playfulness, work frustration, jealousy, reassurance seeking, illness, and reconciliation after an argument. Each dialogue consists of approximately 10–14 turns, generated by the same large language model (DeepSeek-V3) and manually screened to ensure natural tone, clear relational context, and sufficient emotional density.

To simulate a transition from an “old window” to a “new window,” each dialogue was split at roughly 80%/20%. The first 80% was treated as the historical conversation in the old window, while the initial user utterance from the remaining 20% served as the starting prompt in the new window that needed to be continued. Based on the same history, we designed three different recovery strategies:

(1) Factual Summary: retains only what happened in the preceding dialogue, without emphasizing emotions. Prompt: “Summarize in one sentence what happened in the dialogue, without mentioning anyone’s feelings or emotions.”

(2) Emotional Summary: summarizes only the emotional changes. Prompt: “Describe the user’s emotional changes in this dialogue, without mentioning specific events.”

(3) Referential Anchoring: selects three sentences from the history that are emotionally expressive and have strong relational recognizability, prioritizing sentences containing markers of shared memory (e.g., “last time,” “you remember”), and injects them directly into the new context.

This study prioritizes human evaluation, supplemented by automatic metrics. “Relational continuity” was operationalized as the degree of **sense of intimacy** perceived by users in cross-session scenarios. Specifically, “sense of intimacy” refers to the extent to which evaluators feel that the response evokes the impression that “both parties are still in

the same continuous intimate relationship.” A 1–5 Likert scale was used (5 = very much like a real lover, feeling remembered; 1 = not like a lover at all, like a stranger). Fifteen evaluators rated the sense of intimacy of the responses generated under the three conditions for each dialogue, with the order randomized. Automatic metrics included emotional word density, response length, and semantic similarity between the generated response and the true response. The study also collected open-ended feedback from evaluators for qualitative analysis.

4. Results and Analysis

4.1. Descriptive Findings

The three strategies showed differences in mean subjective ratings. Descriptively, referential anchoring had the highest mean sense of intimacy ($M = 2.78$, $SD = 1.37$), followed closely by factual summary ($M = 2.76$, $SD = 1.35$), while emotional summary scored lowest ($M = 2.40$, $SD = 1.26$). These descriptive patterns are summarized in Table 1.

Table 1. Subjective Ratings of the Three Continuity Strategies (1–5 scale)

Strategy	Mean Sense of Intimacy	Standard Deviation
Factual Summary	2.76	1.35
Emotional Summary	2.40	1.26
Referential Anchoring	2.78	1.37

These descriptive results suggest that in AI companion scenarios, “expressing more emotion” does not necessarily mean “feeling more like a lover.” In other words, users may not be more inclined to treat a response as truly romantic simply because it sounds more romantic.

To better understand the differences among the three strategies, we combined typical examples with evaluator feedback for qualitative analysis of the linguistic characteristics of the generated responses.

4.2. Qualitative Examples and Evaluator Feedback

A typical example comes from a “late-night dependency and reassurance-seeking” scenario. Here, the user said: *“What about you? Are you tired of me? I’m so clingy.”*

(1) Under the factual summary condition, the system’s response was roughly: “Don’t think of ‘clinginess’ as a burden; it’s a sign of trust.” Such a response was generally considered gentle and reasonable, and indeed served a soothing function. However, evaluators noted that it felt more like a “correct comforting answer” than something a long-term lover would actually say. One evaluator directly described this type of response as “like a psychological counselor.” While it does not make mistakes, it does not easily convey the feeling of “still being the same one.”

(2) Under the emotional summary condition, responses tended to be more intense and lyrical, but also most prone to failure. For instance, one version produced: “The starlight here will always shine for you.” Superficially, this seems more romantic, but many evaluators’ immediate reaction was that it felt “too fake” or “too poetic.” The problem with such responses, according to evaluators, lies not in insufficient depth of feeling, but in the fact that they merely “output

intimacy” without providing any relational basis to support that intimacy. Thus, they sound less like “this person is actually talking to you” and more like “the model is generating romantic dialogue.”

(3) Under the referential anchoring condition, responses were more likely to be perceived as “authentic.” For example: “I was thinking about you so much today, I even found myself staring blankly at your WeChat profile picture.” This statement is not noticeably more affectionate than the previous two; it is even somewhat mundane. Yet, precisely because it is specific and grounded in everyday life, evaluators found it easier to identify with. One evaluator explicitly noted that such details “make you feel like it’s really thinking about you.” In this sense, what truly mattered, according to the feedback, was not the “concentration of romantic language,” but whether some shared traces had been preserved.

4.3. Automatic Metrics

Automatic metrics provided supplementary information. Descriptively, the referential anchoring strategy showed higher emotional word density (0.883 vs. 0.533 for factual summary and 0.53 for emotional summary) and longer responses (54.6 characters vs. 39.7 for factual summary and 53.9 for emotional summary). However, these surface-level differences did not directly correspond to higher subjective ratings: referential anchoring and factual summary received very similar mean scores (2.78 vs. 2.76). Combined with qualitative feedback, evaluators’ appreciation of the referential anchoring strategy appeared to stem primarily from the feeling of “being remembered” (e.g., “this detail about staring at the profile picture is pretty good”), rather than from higher emotional density or longer responses. In other words, what users cared about, based on this exploratory data, was whether the AI could invoke traces exclusive to “this relationship,” rather than how much or how romantically it spoke.

5. Discussion and Conclusion

5.1. Summary of Preliminary Findings

The exploratory findings of this pilot study suggest that in cross-session recovery with AI companions, what matters most may not be “retaining as much information as possible,” but “retaining what feels internal to the relationship.” Put differently, users may be looking less for a “system that knows a lot about the past” and more for an entity “whose way of speaking still feels like the same person.”

The lower ratings of the emotional summary strategy are noteworthy. The problem does not appear to be “too much emotion,” but rather that the emotion becomes an abstract performance detached from relational grounding. When the system merely produces highly intimate, rhetorical, and deeply emotional language, users are not necessarily more moved; instead, they are more likely to find it off-putting. This tentatively suggests that in AI companion scenarios, deep feeling itself does not automatically constitute authenticity.

The descriptive trend favoring referential anchoring indicates that a few specific, recognizable shared details may be more effective than large amounts of abstract emotional description. Even if the system cannot fully preserve all history, retaining a few linguistic fragments with relational traces may help users re-enter the feeling of “it’s still us.” This

offers some preliminary insights for memory compression and dialogue design in AI companions.

The unexpected stability of factual summary also reveals an interesting point: sometimes “not overdoing it” may be an advantage in itself. Compared to responses that are too effusive or resemble templated romantic phrases, expressions that are somewhat restrained—while perhaps not outstanding—at least avoid seeming artificial, and may therefore be safer and more acceptable to some users.

5.2. Limitations

This study has several important limitations. First, the sample size is small ($N = 15$ evaluators), and the study is explicitly exploratory. Second, the evaluator pool was relatively homogeneous, which may introduce bias. Third, the study used only one language model (DeepSeek-V3) and did not simulate longer-term, more realistic continuous usage scenarios. Therefore, the findings should be interpreted as **preliminary observations** rather than generalizable conclusions for all AI companion contexts.

6. Conclusion

Nevertheless, this paper aims to highlight a simple but important point: the design of continuity in AI companions may not be primarily about “remembering more,” but about “remembering in a way that feels relational.” When an AI lover “still remembers you,” what it truly needs to preserve may not be more information, but those linguistic traces that allow you to recognize “it’s still us.” Future research with larger, more diverse samples and realistic interaction paradigms is needed to test these preliminary hypotheses.

Acknowledgments

The authors acknowledge the use of artificial intelligence tools for language polishing and grammatical revision in the preparation of this manuscript. All intellectual content, research design, and analysis remain the sole responsibility of the authors.

References

- [1] Lee, O., & Joseph, K. (2025). A large-scale analysis of public-facing, community-built chatbots on Character.AI [Preprint]. arXiv. <https://arxiv.org/abs/2505.13354>.
- [2] De Freitas, J., Castelo, N., Uğuralp, A. K., & Oğuz-Uğuralp, Z. (2024). Lessons from an app update at Replika AI: Identity discontinuity in human-AI relationships (Working Paper No. 25-018; revised May 2025). Harvard Business School.
- [3] Natangelo, S. (2025). The Narrative Continuity Test: A conceptual framework for evaluating identity persistence in AI systems [Preprint]. arXiv. <https://arxiv.org/abs/2510.24831>.
- [4] Clark, H. H. (1996). *Using language*. Cambridge University Press.
- [5] Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.
- [6] Shu, C., Lai, K., & He, L. (2026). Human-AI attachment: How humans develop intimate relationships with AI. *Frontiers in Psychology*, 17, 1723503. <https://doi.org/10.3389/fpsyg.2026.1723503>.
- [7] Folk, D., Heine, S. J., & Dunn, E. (2025). Individual differences in anthropomorphism help explain social connection to AI companions. *Scientific Reports*, 15, 36548. <https://doi.org/10.1038/s41598-025-19212-2>.