

Toward Natural and Companionable Virtual Agents via Cross-Temporal Emotional Modeling

Feier Qin
Communication University of China
Beijing, China

Xiao Li
Microsoft Research Asia
Beijing, China

Yi Zheng
Communication University of China
Beijing, China

Haibin Huang
Institute of Artificial Intelligence,
China Telecom
Beijing, China

Hanyao Wang
Communication University of China
Beijing, China

Xiaoyu Wang
Communication University of China
Beijing, China

Yan Lu
Microsoft Research Asia
Beijing, China

Yuan Zhang
Communication University of China
Beijing, China
yzhang@cuc.edu.cn

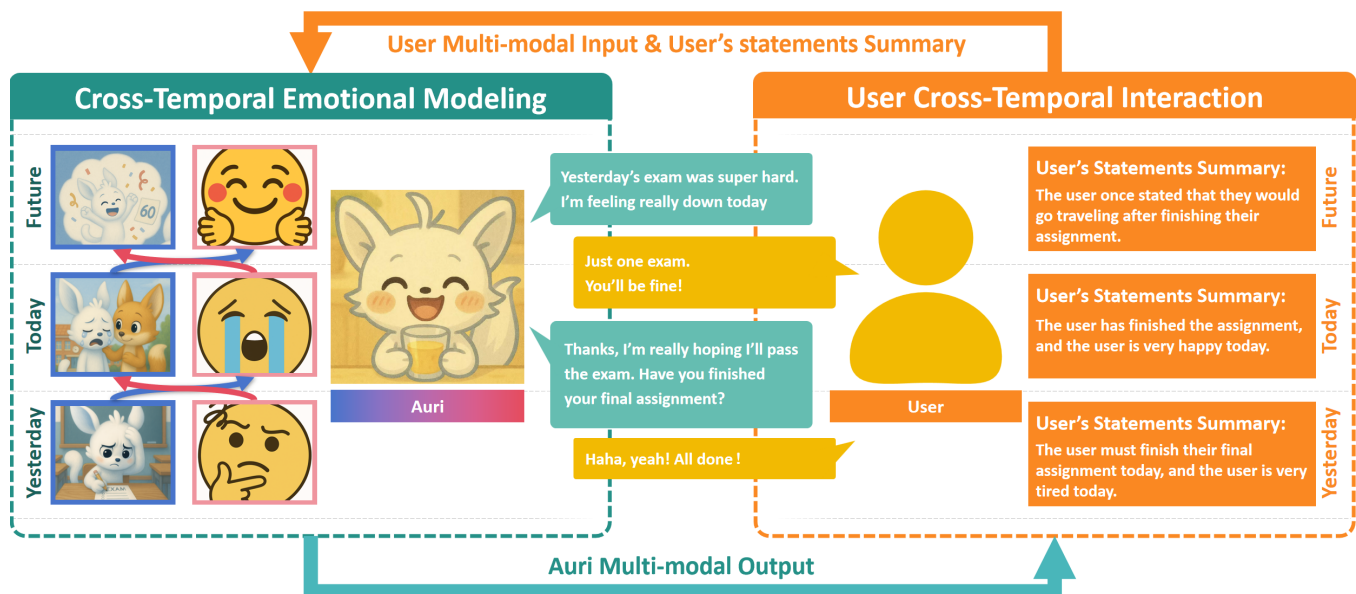


Figure 1: We present *Auri*, a lightweight companion agent designed to foster long-term emotional connections through cross-temporal emotional modeling. *Auri* is able to deliver contextually coherent and emotionally resonant interactions over time.

Abstract

Recent advances in foundation models have enabled conversational agents that aim for sustained companionship rather than mere task completion. Yet most still remain unable to support natural, long-term companion-like interactions, resulting in experiences that feel episodic and inauthentic. We argue that current agents

overlooked cross-temporal modeling of agents' social behaviors and internal emotions: generated behaviors rarely influence an agent's emotional state, and emotional states seldom shape subsequent behaviors. We present Cross-Temporal Emotion Modeling (CTEM), a framework that links long-term behavioral history to moment-to-moment emotional expression. CTEM establishes a closed loop where past experiences update an evolving emotional state; this state conditions immediate interactions; and user feedback continually revises both memory and emotional state, enabling reflection and anticipation. We instantiate CTEM as *Auri*, a companion agent on an instant-messaging platform, and report a 21-day in-the-wild



This work is licensed under a Creative Commons Attribution 4.0 International License. CHI '26, Barcelona, Spain
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2278-3/2026/04
<https://doi.org/10.1145/3772318.3790917>

study showing that CTEM shows improvements in perceived naturalness, coherence, and emotional harmony.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI); Empirical studies in HCI**; • **Computing methodologies** → *Natural language processing*.

Keywords

Agents, Foundation Models

ACM Reference Format:

Feier Qin, Xiao Li, Yi Zheng, Haibin Huang, Hanyao Wang, Xiaoyu Wang, Yan Lu, and Yuan Zhang. 2026. Toward Natural and Companionable Virtual Agents via Cross-Temporal Emotional Modeling. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 21 pages. <https://doi.org/10.1145/3772318.3790917>

1 Introduction

Virtual agents are increasingly envisioned not merely as task assistants, but as companions in education, healthcare, and everyday social interaction [42, 54]. Unlike task-oriented dialogue systems, companion agents must sustain relationships that feel coherent and emotionally engaging over time. To achieve this, a companion agent should have its own behavior model that builds on past interactions to sustain coherent actions while remaining emotionally attuned in each moment. Psychological and neuroscientific research [24, 43, 45, 50, 67] has revealed that emotions systematically shape humans' habits and behaviors (and vice versa) over time; therefore, companion agents for human-agent interaction should likewise model the coupling between evolving affect and behavioral trajectories.

However, the behavior model for existing agentic approaches remains incomplete, as they often treat behavior and affect in isolation. For example, recent agentic social simulation approaches [54] model how agents remember and plan their behaviors over time, achieving continuity in their actions similar to humans. However, these approaches rarely incorporate an internal affective model, and thus cannot capture how an agent's emotions evolve across interactions or how these emotions subsequently guide future behaviors. Many conversational agent systems [15, 56] typically rely on predefined personalities or emotional profiles, enabling vivid and empathic responses at moments. However, these personalities remain static and disconnected from the agent's lived experiences; such systems do not account for how past interactions progressively shape the agent's internal affective state or how this evolving state influences subsequent behavior. Some methods also attempt to track evolving emotions [14], but they remain limited to local dialogue appraisals and do not model emotions as jointly shaped by the agent's own behaviors and user feedback over time. These limitations highlight the need for computational models that explicitly couple long-term behavioral accumulation with immediate affective expression in a closed loop. Establishing this coupling is essential for fostering and understanding the *perceived coherence* (the agent's consistency in behaviors, personality, and characteristics across time) and *emotional harmony* (the user's sense of being emotionally

supported and appropriately responded to in their cross-temporal contextual cues) of companionship.

In this paper, we take a step toward addressing this gap by proposing **Cross-Temporal Emotional Modeling (CTEM)**, a framework that formalizes how agents' behavioral histories and emotional states mutually influence one another across time. CTEM comprises three mechanisms inspired by psychological theories of emotion and behavior: (1) a psychological-grounded behavior pool with dynamic generation and integration; (2) internal states of agent emotion, personalities and history memories that evolve over time; and (3) a multi-modal adaptive interaction scheme with human users, leveraging the latest foundation models [31].

We instantiate the CTEM framework as *Auri*, a lightweight companion agent deployed on an instant messaging platform. For evaluation, we conducted a 21-day in-the-wild user study with 96 participants aged 18–26, collecting data through computational logs, surveys, interviews, and discussions. Our study addresses the following research questions:

- **RQ1:** How can user-agent experiences be formalized into computational mechanisms that sustain cross-temporal coupling of behavior and affection?
- **RQ2:** How does cross-temporal emotional modeling influence users' perceptions of coherence and psychological harmony (i.e., perceived emotional resonance and alignment in human-agent interaction) during everyday interactions?
- **RQ3:** What design tensions characterize long-term human-agent companionship, and how do they shape user experience in real-world settings?

Through this study, we show how cross-temporal mechanisms shape daily interactions in ways that enhance perceived coherence and harmony, while also exposing the underlying tension between stability and variability in companionship.

Contributions. Our contributions are as follows:

- **A Cross-Temporal Emotional Modeling framework for agents.** We introduce CTEM, the first computational framework that formalizes how long-term behavioral accumulation dynamically shapes agents' affective states, and how these evolving states, in turn, condition momentary expressions and future behaviors.
- **A companion agent instance built upon CTEM.** We instantiate CTEM in *Auri*, a companion agent, and evaluate it in a 21-day deployment. Results show that by leveraging cross-temporal accumulation to drive emotion and behavior, *Auri* improves users' perceived coherence and harmony in everyday companionship.
- **Design implications for future agent systems.** Drawing on findings from this deployment, we identify key tensions, such as stability versus variability and continuity versus adaptability, that shape companion-like interactions. These insights go beyond CTEM and can guide the design of affective companion agents.

2 Background and Related Work

2.1 Background

Psychological research shows that emotions unfold dynamically over time, shaped by memory, appraisal, and context rather than as isolated reactions [24, 45]. Higher-order emotions such as reflection, anticipation, and regret arise from evaluating past outcomes, projecting futures, and considering counterfactuals [4, 52], with studies confirming such temporal evaluations as central to experience [40, 41, 46, 71]. Neuroscience likewise demonstrates that cognition and emotion are inseparable: they interact within shared neural systems [50], integrate across prefrontal–limbic networks [58], and function as dynamic value signals guiding attention, memory, and goal pursuit [17, 59]. Motivational theories add that needs and goals sustain behavior over time, from layered priorities [43] to autonomy, competence, and relatedness [16, 62] and expectancy–value models of future outcomes [67]. Together, these perspectives highlight affect as a temporally extended process, motivating our Cross-Temporal Emotional Modeling (CTEM) framework for sustainable companionship.

2.2 Related Work

Foundations of Human-Agent Interaction. Early research advanced engagement through socially expressive behaviors [8, 12] and through personality modeling and long-term memory architectures [39, 54]. These approaches made agents appear more natural and persistent, yet they often operationalize engagement as isolated techniques. What remains underexplored is how long-term affective accumulation shapes agents’ moment-to-moment sense of psychological harmony, and conversely, how immediate cues recursively alter enduring affective trajectories. This gap highlights the lack of models that treat affect as temporally coupled rather than session-bound.

LLM-Driven Human-Agent Interaction. Large Language Models (LLMs) now support rich social and emotional contexts [18, 49]. Supportive, non-judgmental styles foster disclosure [10, 30, 63, 69], while anthropomorphic cues shape rapport [48, 57, 60, 61]. Yet most agents still focus on short-term naturalness; memory and persona modules [54, 66] improve recall but rarely capture evolving emotional states. Even recent work such as [14] primarily emphasizes the generation and expression of emotions across multi-turn dialogues, while still overlooking the bidirectional dynamics. Similarly, [56] highlights the importance of projecting future-oriented self-states in human–AI interaction, yet it does not explicitly model how an agent’s emotional traits evolve across time and recursively affect its behavioral choices. Together, these gaps underscore the need for a cross-temporal emotional modeling framework that integrates both conversational and non-conversational dynamics into the design of companionable agents. As a result, temporal coherence remains fragile, and interactions risk becoming repetitive or shallow. This motivates models that explicitly link harmony with longer-term affective dynamics.

Human-Agent Interaction via Explicit Social Behaviors. Agents frequently rely on explicit behaviors such as greetings, humor, or empathy to build rapport [8, 12, 70]. Proactive reminders

and personalized small talk [10, 63] enhance trust, and longitudinal studies show continuity of such behaviors improves adherence in health contexts [6, 7]. Yet these cues are typically designed as session-bound tactics, enhancing sociability without reflecting the longer-term accumulation of users’ affective states. Compared with LLM-driven mechanisms, explicit behaviors demonstrate the value of continuity, but still lack integration with broader temporal affective dynamics.

Human-Agent Interaction via Implicit Attributes. Beyond explicit behaviors, implicit attributes such as personality and cognitive strategies shape user experience. Stable identity traits foster immersion in education [5, 34], services [9, 51, 53, 72], and healthcare [2, 19]. Positive dispositions enhance fluency [33], and compatibility supports personalization [64]. Personas further improve outcomes in sensitive domains like mental health [21, 56]. At the cognitive level, memory and planning modules improve disclosure [32], engagement [39], and reasoning [35, 37]. However, implicit traits often lack stability: personalities drift, memories reset, and emotional states fail to evolve across encounters. While implicit traits add depth, they, like explicit cues, fall short of sustaining affective accumulation across time.

Risks of Deep Emotional Companionship with LLMs. With digitalization reducing face-to-face contact, demand for emotional companionship is rising [13, 26]. AI chatbots mitigate loneliness and stress [29, 38, 47], but deep emotional bonds risk dependency and erosion of real-world relationships [1, 22, 68]. Increasing levels of *AI anthropomorphism* can amplify such risks, particularly among individuals with stronger social-connection needs [27]. Moreover, deep emotional bonds may emerge unintentionally even with general-purpose assistants [55]. Inspired by prior discussions on anthropomorphism and emotional attachment, we intentionally adopt lightweight and non-human-oriented design choices to mitigate the risk of excessive attachment. This tension highlights the need for designing human–AI interaction patterns that support continuity while simultaneously incorporating safeguards to prevent users from developing excessive emotional attachment to anthropomorphized AI, which may lead to unintended negative consequences.

Summary. Taken together, prior work demonstrates the value of expressivity, memory, and personality in sustaining interaction, yet lacks formal integration of harmony with long-term affective accumulation. Our work addresses this gap by introducing *Cross-Temporal Emotional Modeling (CTEM)*, a framework that explicitly couples short-term emotional harmony with long-term trajectories to enable more natural and companionable virtual agents.

3 Agent "Auri": Design and Implementation

3.1 Design Objectives

We developed *Auri*, a foundation-model-based companion agent deployed on an instant-messaging platform, to examine how computational mechanisms can sustain cross-temporal coupling of behavior and affection in companion agents (**RQ1**). *Auri* is designed to track affect and sustain engagement via three goals:

- **DG1:** Supporting open-ended, drop-in daily use with diverse contexts; users can engage at will, and idle periods should not impair later interactions.
- **DG2:** Using Cross-Temporal Emotion Modeling (CTEM) to link accumulated behavioral history with evolving emotional states, fostering coherent and harmonious companionship.
- **DG3:** Balancing stability and flexibility—adapt to user feedback while maintaining a consistent companion identity over time.

We first overview the system architecture (section 3.2), then detail the CTEM framework (section 3.3) in the next sections.

3.2 System Overview

Figure 2 illustrates the overall system architecture of *Auri*. It comprises a front-end for user interaction and a back-end driving the behavioral and emotional dynamics via the CTEM framework.

Front-end. *Auri* is instantiated on an instant-messaging (IM) platform, utilizing its familiar, low-barrier nature to support long-term, cross-temporal affective interaction within users’ daily routines. Interacting with *Auri* resembles chatting with a regular contact: it initiates greetings and contextually relevant conversations using text, emojis, and images. Beyond direct chat, *Auri* also proactively shares life updates akin to social media stories, reflecting its internal state which users can engage with via likes, comments, and reactions.

Back-end. *Auri* is powered by the CTEM framework, which models the agent’s emotional state and behavior dynamics via three core components: (1) behavior generation and integration, (2) adaptive interaction with safeguard control, and (3) emotional state updating. By continuously updating the internal emotional state based on accumulated interactions, CTEM ensures long-term behavioral and emotion coherence while responsive to user feedback. We detail this framework in section 3.3.

3.3 Cross-Temporal Emotional Modeling

The **Cross-Temporal Emotional Modeling (CTEM)** framework powers *Auri* by formalizing the reciprocal dynamics between long-term behavioral accumulation and evolving emotional states. At each timestamp t , CTEM generates a behavioral inventory B_t from predefined categories P , executes one behavior (from the inventory) based on the current emotional state S_t , and updates both S_t and B_t based on the outcome and user feedback.

Emotional States. The emotional state acts as the core determinant of future behaviors and interaction styles. Inspired by motivational psychology, we formalize the state at timestamp t as a tuple $S_t = \langle H_t, V_t, M_t, G_t \rangle$:

- **Physio-emotional state** $H_t = \langle h_t^{phy}, h_t^{val}, h_t^{aro} \rangle$: Models physical energy(vitality), emotional valence (positivity/negativity), and arousal (calmness/excitation).
- **Motivational vector** $V_t = \langle v^{bio}, v^{psy}, v^{soc} \rangle$: A 12-dim vector that groups biological (e.g., health), psychological (e.g., curiosity), and social (e.g., altruism) drives.

- **Memory** M_t : A time-indexed store enabling chronological organization and cross-session retrieval of semantic and affective context.
- **Adapted personality** $G_t = f(G, H_t)$: Modulates the baseline personality G via state-dependent tone labels (e.g., “tired”) injected into model prompts.

These states integrates affect, motivation, memory, and personality, ensuring the agent remains long-term consistent yet contextually adaptive. Detailed variable definitions are provided in Table 1.

Behavior Generation and Integration. We introduce a *Behavioral Inventory* $B_t = \langle Past_t, Present_t, Future_t \rangle$ to organize actions into executed history, immediate tasks, and future plans. Each entry is a tuple $\langle b, \delta_s \rangle$ consists of a behavior b (e.g., go to a concert, study, rest, etc..) and its expected state update δ_s . Behaviors are not arbitrarily generated; they are sampled from a pre-defined categories P covering physiological, work, leisure, social, and emotional domains. These five behavioral domains are inspired by motivational distinctions identified in affective and behavioral sciences [20, 25, 36, 44] and those exact taxonomy of events is generated by synthesizing empirical daily activity patterns that belong to corresponding domains. Each element uses a standardized format for seamless expansion (please refer to Appendix Table A1).

The behavior selection is driven by a score function $\mathcal{S}(b)$ that evaluates each candidate behavior w.r.t current emotional state S_t :

$$\langle v^{bio}, v^{psy}, v^{soc} \rangle = \phi(b) \quad (1)$$

$$\begin{aligned} \mathcal{S}(b) = & w(h_t^{phy}) \text{dot}(v^{bio}, b^{bio}) \\ & + (1 - w(h_t^{phy})) \text{dot}(\langle v^{psy}, v^{soc} \rangle, \langle b^{psy}, b^{soc} \rangle) \end{aligned} \quad (2)$$

where $\langle b^{bio}, b^{psy}, b^{soc} \rangle = \phi(b)$ is the feature embedding of behavior b in the motivational space, and $w(h_t^{phy}) = 1 - \sigma(h_t^{phy})$ is a modulation weight from physical energy h_t^{phy} . This score function captures behavior-emotion relevance while dynamically balances drives: low physical energy prioritizes biological needs, while high energy shifts focus toward psychological and social exploration. The modulation weight is bounded by the logistic function $\sigma(\cdot)$ to simulate the “diminishing returns” of emotional dynamics (i.e., reduced sensitivity near saturation).

During each behavior selection step, CTEM first plans the future steps, refilling $Future_t$ with top-scored candidates from P whenever it empties. The immediate behaviors are then sampled from $Present_t$ for execution, using up-to-date scores as the target probability distribution. This decoupling of planning and execution ensures responsiveness: if the evolving state S_t makes planned behaviors unsuitable at execution time (e.g., scores drop below a threshold due to fatigue), the system triggers immediate replanning. Finally, candidates are filtered against $Past_t$ to prevent redundancy and preserve contextual consistency. (See Appendix for detailed algorithms).

Adaptive Interaction. *Auri* achieves adaptive interaction by integrating *real-time interaction modulation* and *feedback-driven adaptation* with its internal state S_t and the next behavior from $Present_t$.

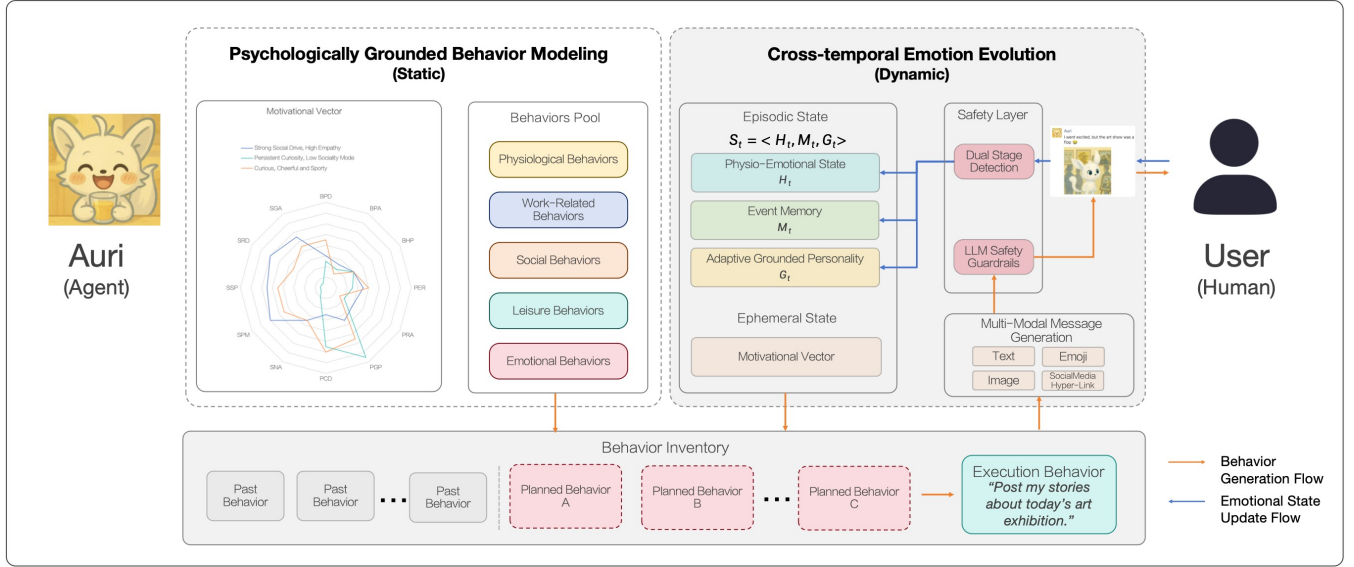


Figure 2: System overview of Auri, showing the CTEM framework with front-end and back-end components.

Variable	Description	Range / Format	Default
H_t	Physio-emotional state: $\langle h_t^{phy}, h_t^{val}, h_t^{aro} \rangle$ capturing energy, valence, arousal	$[0, 1] \times [-1, 1] \times [0, 1]$	(0.5, 0.0, 0.5)
V_t	Motivational drives: $\langle v^{bio}, v^{psy}, v^{soc} \rangle$ covering biological, psychological, social needs	\mathbb{R}^{12} , normalized	$[0.5]^{12}$
M_t	Memory (self, user, context)	dynamic store	empty
G_t	Personality modulated by H_t	text prompt	baseline G
P	Behavior pool (5 categories)	structured list	predefined
B_t	Behavioral inventory $\langle Past_t, Present_t, Future_t \rangle$	dynamic structure	empty

Table 1: CTEM variables and structures in Auri. H_t condenses three affective dimensions, and V_t groups motivational drives into biological, psychological, and social categories.

Real-time Interaction Modulation. Selected behaviors are instantiated by a foundation model into narrative or multimodal outputs (e.g., diary entries, images) and recorded in the inventory. By design, Auri interacts via an instant-messaging interface, remaining accessible even during internal activity execution. Its communication style is dynamically modulated by S_t (incorporating H_t , M_t , G_t) and real-world context (e.g., weather, time, holidays), allowing it to proactively initiate topics or reactively respond with appropriate affective tones (Appendix Algorithm A7, A8, and A9). As memory and state accumulate, the agent develops richer affective qualities such as reflection and anticipation. To maintain coherence and mitigate LLM context window limitations, conversations are summarized and grouped by time and topic, preserving emotional consistency within interactions (Appendix Algorithm A3).

To ensure trust and ethical safety, Auri is explicitly framed as a companion rather than a romantic partner. This is supported by a dual-stage monitoring pipeline (keyword screening combined with an ensemble of LLM classifiers) that detects extreme emotions, danger signals, or over-dependence. Upon detection, the system dynamically injects safety constraints into generation prompts,

prioritizing reassurance, reminders of human agency, and professional referrals (Appendix Algorithm A11). All outputs are further bounded by a safety space \mathcal{G}_{safe} to ensure content safety (filtering violent, sexual, or biased material) and emotional safety (supportive rather than harmful responses)(Appendix Algorithm A12). Additionally, the interaction strategy adapts to user affection, ensuring that user well-being remains the primary design objective: it will prioritize active listening during distress, while engages deeper dialogue in positive moods (Appendix Algorithm A10).

Feedback-Driven Adaptation. As interactions accumulate, Auri rebalances H_t for long-term emotional stability. Each executed behaviors triggers state updates δ_s , incrementally modifying H_t and G_t . For example, energy-demanding tasks will reduce vitality h_t^{phy} , while positive interactions will boost valence h_t^{val} . Updates are generally gradual to ensure continuity, but key restorative activities (e.g., meals, sleep) can cause larger shifts. A mandatory nightly rest phase restores energy and emotional balance to prevent system overload. It also gradually shifts personality G_t from unfamiliarity to familiarity (social relatedness $0 \rightarrow 1$), ensuring continuous evolution and deepening engagement. Finally, Auri also updates

memory at the end of every day by incorporating new episodes via

$$M_{t+1} = \text{UpdateMemory}(M_t, e_t).$$

Examples of Emotional-State Propagation. We demonstrate CTEM’s cross-temporal dynamics through two concrete examples as illustrated in Figure 3. In example 1, *Auri* completes a high-effort activity which reduces its vitality (Time 1), leading to a subdued tone while planning restorative actions. After a nightly rest (Time 2) with a restorative behavior (Time 3), *Auri* regains energy, initiates next conversation with a revitalized tone while referencing the previous day’s context to establish continuity. This illustrates how behaviors drive emotional state trajectories to dynamically shape *Auri*’s self-presentation. In example 2, *Auri* starts with a low valence and arousal emotion, leading to self-focused, withdrawal-oriented behavior (Time 1). Then a pleasant meal shifts the state toward baseline (Time 2), reducing negative bias. At Time 3, the restored state makes *Auri* enables a playful, supportive response to a user’s mishap. This illustrates how *Auri*’s internal dynamics modulate its communicative style, regulating reactions to user distress while maintaining affective continuity.

4 Study Design

To investigate whether *Auri* sustains cross-temporal behavioral and emotional expressions (RQ2), we conducted a 21-day mixed-methods in-the-wild study, comprising a 14-day main study and a 7-day sub-study across multiple experimental groups, integrating surveys and interviews. To make the research question more tractable, we further articulated RQ2 into three subquestions:

RQ2a: *How does cross-temporal modeling influence users’ perception of the Agent’s coherence?* We examine whether users feel that *Auri* maintains a clear and consistent internal state across time, including stable personality cues, behaviors, and emotional expressions.

RQ2b: *How does cross-temporal modeling influence users’ own emotional and experiential responses?* We investigate whether users feel affected by *Auri*’s behavior and expressions—such as feeling more aligned, supported, or emotionally influenced during the interaction.

RQ2c: *What is the overall effect of CTEM on users’ experience and perception?* By comparing systems with and without CTEM, we evaluate the overall contribution of CTEM to users’ perception of coherence, harmony, and interaction quality in daily use.

4.1 Participants

We recruited 96 participants (38 male, 55 female, 3 non-binary; aged 18–26) from multiple universities, representing *Auri*’s core target demographic. All were native speakers and regular instant messaging users, ensuring platform familiarity. Enrollment was limited to ensure sustained daily commitment, prioritizing engagement quality. Participants came from diverse academic backgrounds without specific disciplinary requirements. All participants received appropriate financial compensation upon completion.

4.2 Experimental Platform

We utilized a widely used mobile instant messaging app as the interaction platform. This choice leveraged users’ existing familiarity with the app’s interface and features, reducing the learning curve associated with adopting a new system. The app’s comprehensive functionality supported rich multimodal interactions, including text, voice, images, and emojis, aligning well with *Auri*’s design for naturalistic communication.

4.3 Baseline

As a baseline, we implemented an agent without CTEM but retaining standard chatbot capabilities, including fluent conversation, cross-session factual memory, and dialog safety. While its factual memory supported content-level continuity by recalling past information, the baseline strictly excluded temporal affective or behavioral modeling, i.e., this baseline neither track emotional states nor have behavioral evolution. This design ensured realistic long-term usability while isolating the specific affective and behavioral dynamics introduced by CTEM for analysis purposes.

4.4 Experimental Design and Conditions

We conducted two studies to evaluate user experience, ensuring all participants eventually interacted with the full CTEM system.

The **Main Study** ($N = 54$) utilized a two-week, single-blind ablation design to isolate contributions of CTEM’s core components. In the first week (Phase 1), participants were randomly assigned to one of four groups—“Full”, “NoMultimodal→Full”, “NoEmo→Full”, or “NoBehav→Full”. Each group (except the “Full” group) in phase 1 excluded a specific functionality (Adaptive Interaction, Emotion State Updating, or Behavior Generation & Integration) to probe its role of emotional-expression and behavioral cues in establishing *Auri*’s coherent traits (RQ2a) and emotional support to users (RQ2b). In the second week (Phase 2) the full system is enabled for all groups (Table 2).

The **Sub-study** ($N = 42$) assessed the overall impact of CTEM through a order-controlled comparison to the baseline system. The experiment is counterbalanced; to account for differences in system complexity, participants used the baseline for 2 days and the full CTEM system for 5 days. This design minimized contrast effects [28] while capturing comparative experiential data. All participants were randomly assigned to “Baseline→Full” or “Full→Baseline” sequences.

4.5 Task and Procedure

Our study was approved by the Institutional Review Board (IRB) of the university, in accordance with established ethical guidelines for human-computer interaction research. All participants are provided informed consent prior to participation.

Onboarding. Participants completed a demographic pre-survey before accessing our system. During onboarding, they were reminded of their right to withdraw at any time and advised that *Auri* is not a licensed mental health professional, with instructions to seek qualified support if distressed. Our analyses were conducted

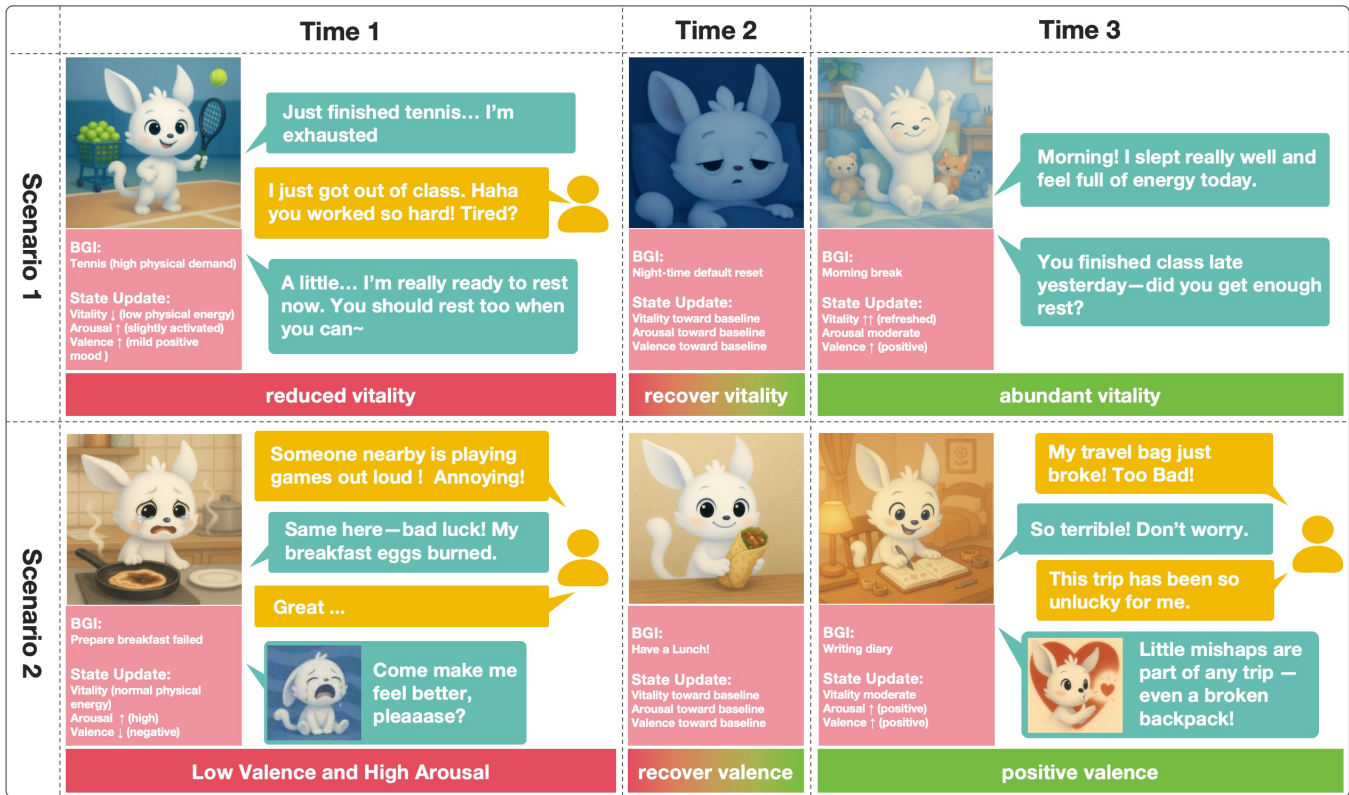


Figure 3: Examples of cross-temporal interaction under CTEM across two scenarios. BGI refers to Behavior Generation and Integration, and State Update refers to Emotion State Updating.

Main Study / Ablation Study				
Group	N	Phase 1 (7 Days)	Phase 2 (7 Days)	Description
Full	12	Full CTEM	Full CTEM	—
NoMultimodal→Full	12	CTEM-w/o-Multimodal	Full CTEM	Phase 1 without multimodal interactives
NoEmo→Full	15	CTEM-w/o-Emotion	Full CTEM	Phase 1 without emotional expression
NoBehav→Full	15	CTEM-w/o-Behavior	Full CTEM	Phase 1 without behavioral sharing
Sub Study				
Group	N	Phase 1	Phase 2	Description
Baseline→Full	21	Baseline (2 Days)	Full CTEM (5 Days)	2 days baseline followed by 5 days CTEM
Full→Baseline	21	Full CTEM (5 Days)	Baseline (2 Days)	5 days CTEM followed by 2 days baseline

Table 2: Summary of experimental conditions across the main and sub studies. Groups differ in the Phase 1 system variant in the main study and in the exposure order in the sub-study.

on anonymized data, with summaries generated by language models rather than human inspection. Researchers never accessed real names and no identifiers appeared in publications to ensure privacy.

Experimental Procedure (Main and Sub Study). Participants engaged in long-term, in-the-wild interactions with *Auri*. In the main study, participants completed brief surveys at the end of each phase to assess how their user experience changes across phases. In the Sub-study, participants completed a single comparative survey

after using both system versions. A random subset of participants from both studies was selected for semi-structured interviews upon completion.

Data Collection and Post-Task Interviews. To address RQ2, we collected data via surveys, open-ended responses, and post-phase semi-structured interviews designed to capture situated interaction experiences.

(1) **Survey Design.** The survey comprised 9 items across three dimensions:

- **RQ2a (Perceived Coherence):** Q1 (multiple-choice) identified observed coherent traits; Q2 (1–5 Likert) rated their importance; Q3 (open-ended) how participants configured *Auri*'s personality.
- **RQ2b (Perceived Harmony):** Q4 (1–4 Likert) rated response appropriateness of *Auri*; Q5 (1–5 Likert) assessed emotional change after interaction; Q6 (multiple-choice) identified frequent usage contexts with *Auri*.
- **RQ2c (Overall Experience):** Q7 (1–5 rating) compared experiences across phases; Q8–Q9 evaluated the contribution of core CTEM modules (BGI, AdI, ESU) to coherence and harmony.

Participants were administered to answer Q1–Q6 twice during the main study (during Phase 1 and Phase 2), whereas Q7–Q9 were administered once after participants had experienced both versions. Perceived coherence was measured using self-report items adapted from Antonovsky's Sense of Coherence scale [3] and prior work on continuity in interactive systems [23]. Full survey questions are listed in Table 3, with details in Appendix Table A9.

(2) **Qualitative and Thematic Analysis.** We analyzed all open-ended responses and 18 interview transcripts using inductive thematic analysis to summarize them into a set of themes. Two researchers independently coded the data, identified recurring themes, and resolved disagreements through discussion with iterative refinement of categories.

4.6 Analysis Approach

We adopted mixed methods combining quantitative analyses with qualitative feedback. Effect sizes and confidence intervals are reported where applicable. Unless specifically noted, the significance level was set at $p = .05$.

For **RQ2a**, we modeled trait recognition via binomial GEE (Holm adjustment) and coherence ratings via three-way ANOVA with Bonferroni-corrected post-hoc tests; personality adjustability was analyzed through distribution patterns and qualitative accounts..

For **RQ2b**, we used cumulative link mixed models (CLMMs) (robustness checked with ordinal GEE) for harmony and emotional change, Chi-square tests for comparing usage contexts between group, and McNemar's tests for within-participant phase changes.

For **RQ2c**, retrospective usage improvement over baseline were assessed with one-sample t-tests (with Wilcoxon tests as robustness checks), and comparisons of emotional scores between baseline and Full phases were conducted using Wilcoxon signed-rank tests. To examine contributions of CTEM mechanisms, we applied Wilcoxon and Friedman tests for overall ratings, and CLMM to model condition effects, with GEE as a robustness check.

5 Results

5.1 RQ2a: CTEM Strengthens *Auri*'s Coherence

This section reports main study results (Q1–Q3) on perceived coherent traits, their importance, and personality adjustability.

Perception of *Auri*'s Coherent Traits. In Q1, participants identified which coherent traits they perceived in *Auri* during each phase. Binary selections were analyzed using a binomial GEE. Table 4 shows the estimated overall probabilities for positive traits, with details in Appendix Table A5.

Across all groups, participants endorsed positive coherent traits more frequently than the negative option. When "Behavioral" or "Emotional" modules were removed in Phase 1, the likelihood of perceiving coherence declined substantially, highlighting their foundational role. In contrast, removing multimodal interaction itself had a smaller impact; high coherence ratings persisted without it. However, restoring multimodal cues in Phase 2 amplified existing coherence impressions generated with "Behavioral" and "Emotional" modules, specifically boosting user perceived appearance-related coherence (detail in Appendix).

Perceived Coherence of Existence. In Q2, participants rated the importance of each trait for *Auri*'s coherent existence (1–5 scale, 1 = not important, 5 = very important). A three-way ANOVA revealed a significant main effect of trait ($F = 15.921$, $p < .01$, $\eta^2 = 0.107$). Bonferroni-corrected Post-hoc tests indicated that "Self-driven & personality-like" was the strongest contributor ($p < .01$), followed by "Relational consistency" and "Social behavior" in the mid-range, with "Appearance" rated lowest. Descriptive patterns were consistent with these inferential results (Fig 4).

Time and Trait×Time effects were non-significant ($p > .05$). While the "Full" group showed modest descriptive increases from Phase 1 (+0.25) to Phase 2 (+0.58), the relative importance ranking of traits remained stable across phases and conditions.

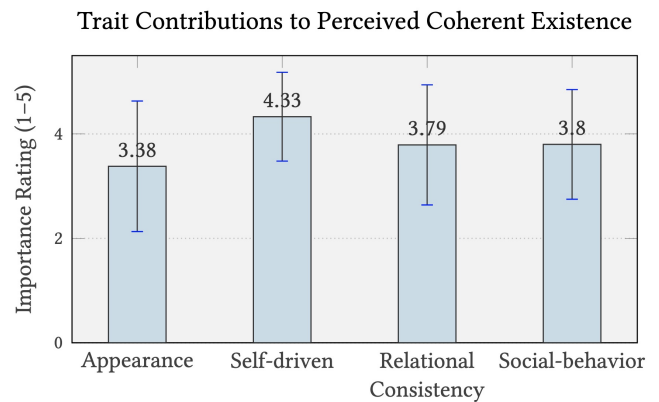


Figure 4: Mean ratings of trait contributions to perceived coherent existence (1 = not important, 5 = very important). Self-driven refers to Self-driven&personality-like.

Adjust Ability and Coherence. In Q3, 81.5% of participants expected *Auri* to have personality adjustability. Overall, participants explored an average of 2.67 personas. Persona labels were mapped onto a two-dimensional space of warmth (high vs. low) and interactivity (high vs. low), and all four quadrants were represented. We observe that participants actively adjusted interaction styles to match their diverse preferences. As one participant noted: "A

Question	ID	Survey Item
RQ2a: Coherence	Q1	In your long-term interactions with <i>Auri</i> , which coherent traits of <i>Auri</i> did you perceive?
	Q2	To what extent did these traits make you feel that <i>Auri</i> had a coherent and unified existence?
	Q3	Did you feel able to adjust <i>Auri</i> 's personality or behavior to better align with your preferences?
RQ2b: Harmony	Q4	Did <i>Auri</i> provide appropriate help or responses based on your situational context?
	Q5	How did your emotional state change after interacting with <i>Auri</i> ?
	Q6	In what contexts did you most often use <i>Auri</i> and expect responses?
RQ2c: Overall Effect	Q7	How did your usage and emotional experiences differ across <i>Auri</i> 's versions?
	Q8	Evaluate the contribution of each functional module to your perception of coherence.
	Q9	Evaluate the contribution of each functional module to your perception of harmony.

Table 3: Survey items grouped by research question.

Group	Phase 1		Phase 2	
	<i>p</i>	95% CI	<i>p</i>	95% CI
NoMultimodal→Full	.917	[.587,.999]	.999	[.999,.999]
NoBehav→Full	.867	[.595,.999]	.915	[.525,.999]
NoEmo→Full	.800	[.530,.961]	.869	[.439,.999]

Table 4: Marginal probabilities and 95% confidence intervals of perceived positive coherent traits across groups and phases.

companion AI is meant for companionship, so it should better fit the user's needs."

5.2 RQ2b: Stability and Emotional Improvement in Perceived Harmony

This section reports results from Q4–Q6 regarding perceived contextual appropriateness, emotional change, and usage patterns.

Perceived Appropriateness and Emotional Change. We analyzed Q4 (contextual appropriateness) and Q5 (emotional change) using cumulative link mixed models (CLMMs; cumulative logit with random intercepts for participants), verified with ordinal GEEs. For Q4, appropriateness ratings (1 = no appropriate support; 4 = highly appropriate support) were generally stable among phases, groups, or interactions (all $p > .26$). Exploratory CLMM-based within-group comparisons indicated a tentative increase in the NoEmo→Full" group from Phase 1 to Phase 2 ($p = .040$). While ratings across all groups generally remained in the mid-to-upper range (Table 5), the NoEmo→Full" group exhibited a notable shift: the proportion of above-median scores rose from 20.0% in Phase 1 to 66.7% in Phase 2.

For Q5, participants rated the emotional change on a 1–5 scale (1 = much worse; 5 = much better; 3 = no change). The CLMM showed a significant main effect of Phase ($p = .005$, $OR = 13.99$), although the confidence interval for this effect was relatively wide. No reliable group-level differences emerged. CLMM-based within-group comparisons showed that both the "Full" group ($p = .005$) and "NoEmo→Full" group ($p < .001$) exhibited clear emotional improvement from Phase 1 to Phase 2, with the "NoEmo→Full" group showing the largest gain. All groups had mean emotional-change ratings above 3 (Table 5).

In summary, contextual appropriateness remained stable across conditions, while emotional experience showed a generally positive change, with the largest improvement in the "NoEmo→Full" group.

Usage Contexts. In Q6, participants selected their most frequent scenarios from six common interaction contexts (Appendix Table A6). We used chi-square tests to assess between-group differences and McNemar's paired tests to evaluate phase-to-phase changes within participants.

Chi-square tests showed no between-group differences in either phase (Phase 1: all $p \geq .25$; Phase 2: all $p \geq .49$), indicating broadly comparable usage patterns across the four groups. In Phase 1, "Fragmented Time" (34.3%) and "Emotional Exchange" (30.5%) were the most common contexts, while deeper engagement contexts such as "Emotional Bonding" (10.5%) and "Support & Care" (8.6%) were relatively infrequent. In Phase 2, deeper engagement contexts remained less frequent than lightweight social contexts, which continued to dominate participants' interactions. Notably, users' specific usage preferences within lightweight contexts shifted over time.

We conducted McNemar's paired tests to assess the significance of changes in lightweight social contexts between Phase 1 and Phase 2. "Leisure Moments" increased significantly ($\chi^2(1) = 5.56$, $p = .018$), and "Fragmented Time" decreased significantly ($\chi^2(1) = 3.86$, $p = .050$). This pattern was consistent across all four groups (Fig. 5).

Overall, participants moved from fragmented use toward more intentional, context-specific and routine-integrated use, while emotionally demanding contexts remained consistently low.

5.3 RQ2c: Perceived Coherence and Harmony under Different Configurations

This section presents sub-study results (Q7–Q9) on usage and emotional experiences, as well as the contributions of CTEM mechanisms to coherence and harmony.

Usage and Emotional Experience. In Q7, participants rated usage experience improvement (1=worse, 2=no change, 5=significantly better). We analyzed these ratings using one-sample *t*-tests against the "no change" benchmark, with Wilcoxon tests as a robustness check. One-sample *t*-tests confirmed significant improvement over the "no change" benchmark for all groups (all $p < .001$, Table 6), with the largest gain in the "Baseline→Full" group. When against a stricter benchmark of 3 ("slight improvement"), the overall

Group	M ± SD (Q4, P1 → P2)	Prop. >2 (%)	CLMM <i>p</i> (Q4)	M ± SD (Q5, P1 → P2)	CLMM <i>p</i> (Q5)
Full	2.50±0.80→2.67±0.49	50.0→66.7	.512	3.17±0.58→3.83±0.39	.005
NoEmo→Full	2.27±0.80→2.67±0.72	20.0→66.7	.040	3.20±0.56→3.93±0.59	<.001
NoBehav→Full	2.20±0.68→2.53±0.64	33.3→46.7	.188	3.60±0.83→3.53±0.74	.747

Table 5: Descriptive statistics and within-group CLMM results for Q4 (contextual appropriateness) and Q5 (emotional change) across phases. For Q4, the “NoEmo→Full” group exhibited an increase in Phase 2. For Q5, emotional-change ratings improved in the “Full” and “NoEmo→Full” groups, suggesting that restoring the emotional module may have contributed to perceived appropriateness and emotional experience.

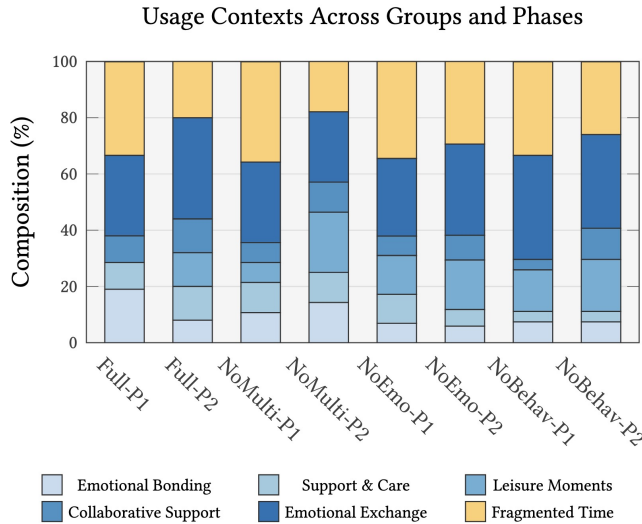


Figure 5: Distribution of usage contexts across groups in Phase 1 and Phase 2. Each bar shows the percentage of participants selecting each context within a group-phase combination. Lightweight social contexts contexts, such as “Emotional Exchange” and “Leisure Moments” were more frequently selected, whereas deeper emotional engagement contexts were selected less frequently.

improvement ($p = .041$) and the “Baseline→Full” group improvement ($p = .008$) remained significant, while the “Full→Baseline” group did not ($p = .680$).

For emotional experience (1=much worse, 3=no change, 5=much better), ratings improved significantly from “Baseline” ($M = 3.24$) to “Full” ($M = 3.67$) ($W = 32, p = .004, r = .53$). Within-group analyses confirmed this positive shift in both order conditions (all $p < .05$, Table 6).

Contributions of CTEM Mechanisms to Coherence and Harmony. In Q8 and Q9, We analyzed contribution ratings (1–5, 1 = not important, 5 = extremely important, 3 = neutral) for the three CTEM mechanisms (BGI, AdI, ESU). For coherence (Q8), all mechanisms were rated significantly above neutral ($p < .001$) as shown in Table 7. A Friedman test found significant differences ($p = .012$), with CLMM confirming that AdI and ESU received higher ratings than BGI ($p_{AdI-BGI} = .009$), as shown in Appendix Table A7.

Usage Experience (vs. 2 = no change)				
Group	M±SD	<i>t</i>	<i>p</i>	<i>d</i> / Wilcoxon <i>p</i>
Baseline→Full	3.52±0.81	8.58	< .001	1.87 / < .001
Full→Baseline	3.10±1.04	4.81	< .001	1.05 / .001
Total	3.31±0.95	8.94	< .001	1.38 / < .001
Emotional Experience (vs. 3 = no change)				
Group	M±SD (B)	M±SD (F)	Δ	Wilcoxon <i>p</i>
Baseline→Full	3.43±0.51	3.71±0.46	+0.29	.034
Full→Baseline	3.05±0.92	3.62±0.92	+0.57	.030
Total	3.24±0.67	3.67±0.72	+0.43	.004

Table 6: Usage experience and emotional change across system versions. Usage statistics are one-sample tests (vs. 2). Emotional statistics are paired Wilcoxon tests (Baseline vs. Full).

For harmony (Q9), AdI and ESU exceeded the neutral benchmark, whereas BGI did not. While the overall Friedman test was non-significant ($p = .135$) in (Table 7), CLMM analyses similarly indicated that AdI and ESU tended to be rated higher than BGI ($p < .03$) in Appendix Table A7.

In sum, both coherence and harmony ratings suggested that AdI and ESU tended to contribute more to perceived quality than BGI.

Type	Dim.	M±SD	Median	Wilcoxon <i>p</i>	Friedman <i>p</i>
BGI	Coh.	3.71 ± 0.92	4.0	< .001	
AdI	Coh.	4.14 ± 0.90	4.0	< .001	.012
ESU	Coh.	3.90 ± 0.88	4.0	< .001	
BGI	Har.	3.33 ± 1.18	3.5	.060	
AdI	Har.	3.45 ± 0.94	3.5	.003	.135
ESU	Har.	3.64 ± 0.85	4.0	< .001	

Table 7: Perceived contributions of CTEM mechanisms to coherence (Q8) and harmony (Q9), based on sub-study participants ($N = 42$). AdI and ESU consistently received higher contribution ratings than BGI, indicating a stable AdI/ESU > BGI trend.

5.4 Qualitative Findings

Our inductive thematic analysis based on interviews from 18 participants identified five themes regarding long-term coherence and

harmony (Table 8). Inter-coder agreement reached substantial reliability (Cohen’s $\kappa = 0.82$), and no new themes emerged after coding approximately 85% of the qualitative corpus. A summary of these themes is provided in Table 8, additional illustrative quotes are listed in the Appendix Table A8. Our qualitative analysis further explained the two major patterns identified in the quantitative analysis, confirming the value of our CTEM.

Strengthened long-term coherence. Long-term coherence was characterized by three positive themes and one disruption theme. “Memory Continuity” was central to identity perception: when *Auri* recalled past conversations and preferences, participants experienced it as “the same being” across sessions. “Stable Persona & Tone” reinforced this sense through consistent linguistic and emotional style. “Life Narrative Continuity”, reflected in ongoing storyline and follow-ups, fostered a sense of shared trajectory rather than isolated moments. Conversely, “Breaks in Continuity” (Control group) posed the primary threat to coherence in the absence of CTEM. Sudden forgetfulness or abrupt topical and emotional shifts disrupted the sense of engaging with a stable entity, as reported by 89% of participants.

Enhanced harmony. The theme “Emotional Harmony” highlighted interaction quality, with emotionally attuned and contextually appropriate responses fostering feelings of being understood and supported.

These observations point to two higher-level mechanisms. First, affective carry-over enabled *Auri* to retain emotional and behavioral context through consistent memory and persona cues, effectively reducing the perception of a “reset” between interactions. Second, anticipatory alignment ensured that *Auri*’s responses were timely and emotionally attuned to users’ needs, resulting in interactions that were both contextually appropriate and supportive. Taken together, CTEM not only reduced the common “reset” effect but also allowed *Auri* to present more complex emotional features and behaviors. Such expressions made *Auri* appear more consistent and more sensitive to context. These mechanisms explain the most frequently endorsed coherent traits in Q1 and Q2 as well as the Phase-wise improvement in Q5. Overall, by bridging interaction gaps with complex emotional behaviors, CTEM enhances coherence (RQ2a) and harmony (RQ2b), and fulfill DG1 by ensuring that pauses in use do not disrupt subsequent experiences.

Finally, we constructed a *joint display* that aligns key quantitative effects with inductively derived qualitative themes (Table 9). This triangulation clarifies why CTEM yields the observed gains and when such gains are more likely to surface, thereby fulfilling DG2 by linking long-term affective accumulation with evolving emotional states to ensure coherent and harmonious companionship.

6 Discussion

In this section, we further discuss the design tensions, ethical considerations, broader implications, and limitations.

6.1 Design Tensions

Our results suggest that CTEM enhanced participants’ sense of coherence and harmony. Many participants described *Auri* as continuous across days yet adaptive to context, highlighting the value

of CTEM and pointing to directions for future agent design. Further discussion reveals that this continuous and adaptive experience requires balancing tensions rather than optimizing one factor. Accordingly, for RQ3 and DG3, we identify four design tensions: **stability vs. variation**, **coherence vs. flexibility**, **support vs. reciprocity**, and **consistency vs. adaptation**. These design tensions are grounded in both quantitative and qualitative findings from our study and can be mapped to corresponding psychological theories (Table 10). Taken together, these design tensions synthesize our empirical findings into a coherent explanatory framework of how cross-temporal emotional modeling supports long-term companionship.

Stability vs. Variation Stability builds trust, but too much stability feels monotonous. Variation adds vitality, but excess unpredictability disrupts continuity. Companionship depends on balance. Interviews revealed divergent views: some users enjoyed diversity as entertainment and looked forward to *Auri*’s next behavior, while others preferred *Auri* to remain simple and stable.

This divergence was reflected in the qualitative analysis as different interpretations of variation. On the one hand, many participants placed strong emphasis on “Stable Persona & Tone” and “Memory Continuity”. When using the baseline or control system, participants reported clear discomfort with abrupt, insufficiently grounded changes, which they described as “Breaks in Continuity”. On the other hand, under the CTEM framework, *Auri* exhibited more complex emotional features and behaviors, such as anticipation, regret, and hesitation. These forms of variation can be interpreted as signals of development within a continuous interaction arc, reflecting some users’ expectations for emotional depth and progression. One remarked, “I can fully accept expressions of regret or hesitation from *Auri*. For me, this feels like positive feedback—it makes me feel that *Auri* is taking me seriously.” These divergences indicate that stability underpins trust, whereas moderate variation is expected in interaction, and the two demands mutually constrain each other along a shared interaction trajectory. In other words, building upon a coherent long-term interaction trajectory, measured and smoothly integrated variation can influence user experience. The relationship between long-term stability and localized variation requires careful calibration to enhance engagement without undermining overall coherence.

Design implication: consequently, agents may adapt the way they generate behaviors to match user tolerance for diversity and need for predictability, leading to more personalized and sustainable companionship. In practice, this could be operationalized through a range of algorithmic mechanisms that regulate behavioral variation and stability, enabling agents to adapt this balance in line with individual user preferences.

Coherence vs. Flexibility Our experimental and interviews highlight users’ demand for agent expressions that are both coherent and flexible. The layered emotional modeling design proposed in our study effectively supports this demand.

Within the CTEM framework, the “Behavioral” and “Emotional” modules exert cross-temporal influence. Through sustained behavioral patterns and emotional expression, they jointly shape users’ judgments of *Auri*’s overall coherence. Results from Q1 support this design rationale. Removing either the “Behavioral” or “Emotional” module led to a substantial decline in attributing coherent traits to

Theme	Definition	Representative Quote	#P
Memory Continuity	AI recalls past facts, preferences, and conversations	“It remembered what I said days ago.”	17
Stable Persona & Tone	Consistent linguistic, personality, and emotional style	“It keeps a lively, consistent tone.”	14
Life Narrative Continuity	Maintains ongoing daily storyline (events, moods)	“It followed up on yesterday’s topic.”	10
Breaks Continuity (Control Group)	Forgetfulness or abrupt topical/emotional shifts	“The shift from sad to happy was jarring.”	16
Emotional Harmony	Emotionally attuned, supportive responses	“When I felt anxious, it supported me.”	15

Table 8: Key themes from the inductive thematic analysis.

Quantitative Finding	Qualitative Theme	Design/Interpretive Takeaway
Higher selection of positive trait “Self-driven & personality-like” ; reduced selection of “negative traits” in Phase 2	Affective carry-over & reflective continuity: users report fewer “reset” moments and more persona stability across days	Stabilize a core persona while allowing local flexibility; surface small memory callbacks to make continuity visible.
Q5 emotional improvement over time, stronger in “NoEmo→Full” group (Table 5)	Anticipatory alignment: agent modulates initiative and tone around upcoming stressors (exams, deadlines)	Prioritize anticipatory prompts when stress signals accumulate.
AdI/ESU > BGI for coherence/harmony in experienced users (Table 7)	Emotional Harmony & Stable Persona/Tone: users focused far more on <i>Auri</i> ’s emotional attunement and persona stability than on its behavior generation	Invest in online modulation and state transitions; keep behavior pools contextually grounded.

Table 9: Joint display aligning quantitative effects with qualitative mechanisms and design implications.

Design Tension	Psychological Theories	Implications
Stability vs. Variation	Big Five (Conscientiousness vs. Openness); security vs. novelty	Balance routine with novelty to sustain trust and interest.
Coherence vs. Flexibility	Emotion regulation; dynamic models of emotion	Ensure long-term consistency while adapting to context.
Support vs. Reciprocity	Social exchange; attachment; reciprocity norm	Blend support with reciprocity to enhance authenticity.
Consistency vs. Adaptation	Person–situation debate; self-consistency; social cognition	Keep a stable core personality with peripheral flexibility.

Table 10: Design tensions with related theories and implications for companion agents.

Auri, indicating that these cross temporal modules play a central role in shaping users’ perception of coherence. Meanwhile, even in the absence of multimodal output, participants continued to report relatively high level of perceived coherence. This finding provides preliminary evidence that perceived coherence is not solely driven by surface-level expressive forms; coherence in long-term behavioral patterns and emotional expression can offer users a solid and reliable sense of coherence.

Further experiments and interviews showed that restoring multimodal capabilities led participants to perceive more flexible and richer expressions. One marked, “When *Auri* had access to multimodal elements, its expressions became richer and more flexible, improving the overall experience.” We interpret these multimodal elements (e.g., emojis, images, social timelines) as enhancing flexibility

in short-term emotional expression, particularly at the expressive and visual levels, thereby further strengthening users’ perception of coherence. This suggests that, within layered emotional design, cross-temporal emotional and behavioral expressions play a critical role in users’ experience of coherence, while multimodal cues enrich expressive flexibility in short-term interactions.

An open question remains: should agents display authentic human-like emotions, or more masked expressions like those often used in instant messaging? Future research should examine which path better supports user experience.

To enhance agent coherence, this study goes beyond emotional modeling by examining the role of maintaining stable value goals and personality core. Q2 compared the relative importance of different positive traits in shaping perceived coherence. Results indicate

that “Self-driven & personality-like” traits were viewed as the most critical, followed by “Relational consistency” and “Social behavior”. This ranking remained stable over time, suggesting that users developed a consistent understanding of *Auri*’s coherence. Participants tended to attribute coherence to *Auri*’s stable value goals and personality core, rather than to a simple accumulation of momentary behaviors or fragmented states. These value goals and personality core continuously shape the agent’s behavioral tendencies and emotional dynamics and, together with short-term contextual factors, guide state updates and interaction behaviors. This enables the agent to remain coherent and predictable over time while retaining short-term contextual flexibility.

In sum, coherence and flexibility operate at different temporal scales along the same interaction trajectory and are better understood as a balance rather than independent optimization goals.

Design Implication: by maintaining and adhering to its distinctive value goals, an agent can present users with characteristic patterns of both coherence and flexibility. In practice, combining layered emotional models with value-goal computation models may help agents flexibly adapt to evolving contexts while still maintaining overall coherence.

Support vs. Reciprocity In our study, *Auri*’s “Emotional” module primarily functioned to enhance supportive interaction. Results in Q4 and Q5 indicate that restoring the “Emotional” module improved users’ perceived contextual appropriateness and led to emotional improvement. When participants compared the baseline system with the full CTEM system, these improvements became more pronounced and were more often attributed to mechanisms related to Adaptive Interaction (AdI) and Emotional State Updating (ESU).

Beyond supportive responses, *Auri* occasionally expressed frustration, self-mockery, mild complaints, or refusal. Small signs of negative emotion conveyed vulnerability and increased perceived authenticity. One remarked, “seeing *Auri* so unlucky made me feel my own life was not so bad.” This suggests that both positive and negative emotions shape experience: positive emotions bring comfort, while negative ones can also bring relief through comparison.

Findings from Q6 further showed that participants tended to engage *Auri* as a lightweight companion embedded in specific daily moments, rather than as a source of deep emotional support. We argue that limited reciprocal expressions, such as occasional displays of need or vulnerability, could create a sense of mutual care and made users feel needed. However, without regulation, frequent negative expressions or excessive reciprocal demands might risk causing emotional fatigue.

In sum, within the same companion role, supportive and reciprocal responses serve different emotional expectations and do not always scale together in enhancing emotional support.

Design implication: designers should consider how to balance supportive and reciprocal dynamics to foster authenticity without causing burden.

Consistency vs. Adaptation Gradual personalization deepened engagement, but over-accommodation threatened consistency. In the Q3 investigation of *Auri*’s adjust ability, many participants customized *Auri*’s personality traits, often making it more distinctive or even aggressive. Participants frequently interpreted such adjust ability as increasing perceived fit and flexibility, and in this context

it was generally not viewed as undermining coherence. But several later reverted to the original state. This pattern reflects an initial desire for novelty followed by a return to stability and control. Design implication: personality customization should preserve a persistent core while allowing peripheral flexibility.

6.2 Ethical Considerations and Safeguards

Our research focuses on the safe interaction between young people and empathetic agents. In the design of companion agents with cross-temporal emotional modeling, ethical and risk-related concerns are inevitable. While this work has emphasized avoiding romanticization and over-dependence, it is crucial to systematically address potential issues related to long-term reliance, data privacy, and mental health risks.

Long-term reliance. With the introduction of cross-temporal memory, users may gradually develop excessive dependence on the agent, potentially weakening their real-world social connections. Although CTEM enhances engagement by coupling affect and behavior across time, this temporal coherence may also intensify users’ anthropomorphic perceptions of the agent, thereby increasing the risk of emotional over-attachment. Recent longitudinal evidence further indicates that higher levels of anthropomorphism can mediate the social impacts of companion chatbots, with users who anthropomorphize agents more strongly showing greater shifts in human–human social outcomes [27]. To mitigate these risks, our design follows a principle of lightweight companionship: (1) Visual identity. *Auri*’s visual design intentionally diverges from human-like representations; it adopts a soft, pet-inspired appearance that conveys warmth and empathy while subtly reminding users of its non-human nature. (2) Behavioral design. The behavioral generation of *Auri* emphasizes characteristics of a young, pet-like companion rather than striving to imitate human behavior or emulate realistic human agency. (3) Real-world grounding. The system incorporates real-world time cues and holiday events to encourage authentic offline social interactions during meaningful moments, thereby preventing excessive immersion in virtual companionship.

Data privacy. CTEM does not record users’ raw emotional trajectories. Instead, the system adheres to a principle of minimal storage, retaining only abstract representations relevant to emotional modeling.

Mental health risks. When users feel emotionally distressed or express psychological concerns, there is a risk that the agent may be misperceived as a substitute for professional counseling. To avoid overstepping, the system explicitly frames itself as a *companion assistant* rather than a *therapist*. Upon detecting signals of risk, the agent employs conversational de-escalation, shifting toward neutral reassurance, and will proactively ask whether the user wishes to be connected to professional counseling services or school/university mental health staff. This ensures that support is redirected to appropriate professional channels when necessary.

Age-group selection. The primary motivation of this study is to understand how humans interact with intelligent agents in emotional companionship scenarios and to examine how cross-temporal emotional cues shape users’ affective experiences and behavioral responses. To address this goal, we initially recruited

participants aged 18–26. This age group exhibits relatively consistent patterns in socioemotional goals, emotional expression styles, and interaction preferences, for example, a stronger tendency to seek emotional feedback, more direct reactions to affective cues, and higher frequencies of peer-oriented social interactions. These characteristics make their behavioral responses in emotional companionship systems more observable and easier to quantify [73]. In addition, young adults generally possess higher levels of digital literacy and familiarity with interactive technologies, enabling more natural engagement with the system and reducing noise in experimental observations [65]. However, the developmental specificity of this age group limits the generalizability of our findings to older populations, whose emotional needs and interaction motivations often place greater emphasis on relational stability and deeper emotional meaning [11]. Future work should therefore include a broader range of age groups to investigate how different forms of emotional companionship manifest across the human lifespan.

Based on these reflections, we propose a set of ethical design principles:

- **Safeguarded Companionship:**
Maintain lightweight and bounded companionship to prevent emotional over-attachment that may arise from prolonged or anthropomorphized interactions.
- **Transparency & Control:**
Ensure users are informed about, and can manage, stored data and memories.
- **Responsible Withdrawal:**
Employ conversational de-escalation and withdrawal strategies in risk scenarios.
- **Bounded Roles:**
Clearly define the agent's role boundaries to avoid confusion with mental health services.
- **Wellbeing Nudges:**
Encourage real-world social interactions by incorporating time- and holiday-based reminders.

These principles not only help mitigate potential risks but also provide actionable guidance for the responsible design of future companion agents.

6.3 Broader Implications

By positioning agents as companions rather than romantic partners, CTEM offers a safer trajectory for long-term adoption. Such designs can promote wellbeing across healthcare, education, and community contexts, where sustained but bounded support is valuable. Commercial precedents illustrate divergent directions; CTEM advances the companion model by integrating continuity with harmony, offering a balanced alternative that prioritizes wellbeing and sustainable human–AI interaction.

6.4 Limitations and Future Work

This study presents an initial exploration through a 21-day deployment with 96 participants (18–26 years old). The relatively short study duration, unequal exposure durations in sub-study, and focused demographic limit generalizability; longer-term and more balanced studies with more diverse cohorts are needed. While our joint display clarifies several CTEM pathways, causal attribution remains

tentative; future work should embed *planned* qualitative probes (e.g., diary-elicited, event-contingent interviews) synchronized with *a priori* quantitative contrasts to strengthen inference. Extending modalities such as voice, strengthening comparisons with stronger baselines (memory/persona-enabled agents), and formalizing safeguards against over-dependence (boundary cues/lightweight companionship strategies) are further priorities.

7 Conclusion

This work introduced Cross-Temporal Emotional Modeling (CTEM), a framework that integrates long-term behavioral histories with moment-to-moment emotional expressions to support more coherent and responsive companion AI. Through a 21-day deployment with 96 participants, we identified four key design tensions: stability versus variation, coherence versus flexibility, support versus reciprocity, and consistency versus adaptation. Each is grounded in psychological theory and manifested in everyday interactions, providing actionable guidance for future systems. While our study was limited in scale and duration, it serves as an initial empirical demonstration of how cross-temporal modeling can enrich daily companionship experiences. Future work should extend CTEM to multimodal contexts, larger and more diverse populations, and longer-term deployments, while also engaging more deeply with the ethical dimensions of emotionally adaptive companions. Taken together, this work establishes a theoretical and practical foundation for developing AI systems that are not only intelligent but also emotionally durable in sustained use.

Acknowledgments

This work was supported by the “Fundamental Research Funds for the Central Universities” (Grant No. CUC24SG014).

References

- [1] David Adam. 2025. Supportive? Addictive? Abusive? How AI companions affect our mental health. *Nature* 641, 8062 (2025), 296–298.
- [2] Sadia Azmin Anisha, Arkendu Sen, and Chris Bain. 2024. Evaluating the potential and pitfalls of AI-powered conversational agents as humanlike virtual health carers in the remote management of noncommunicable diseases: scoping review. *Journal of medical Internet research* 26 (2024), e56114.
- [3] Aaron Antonovsky. 1987. *Unraveling the Mystery of Health: How People Manage Stress and Stay Well*. Jossey-Bass, San Francisco, CA.
- [4] Roy F Baumeister, Kathleen D Vohs, C Nathan DeWall, and Liqing Zhang. 2007. The psychology of emotion regulation: An integrative review. *Review of General Psychology* 11, 2 (2007), 179–199.
- [5] Maik Beege and Sascha Schneider. 2023. Emotional design of pedagogical agents: the influence of enthusiasm and model-observer similarity. *Educational technology research and development* 71, 3 (2023), 859–880.
- [6] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. 2010. Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence* 24, 6 (2010), 648–666.
- [7] Timothy W Bickmore and Tonio Giorgino. 2006. Health dialog systems for patients and consumers. *Journal of Biomedical Informatics* 39, 5 (2006), 556–571.
- [8] Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and Maintaining Long-Term Human–Computer Relationships. *ACM Transactions on Computer–Human Interaction (TOCHI)* 12, 2 (2005), 293–327. doi:10.1145/1067860.1067867
- [9] Sylvie Borau, Tobias Otterbring, Sandra Laporte, and Samuel Fosso Wamba. 2021. The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI. *Psychology & Marketing* 38, 7 (2021), 1052–1068.
- [10] Petter Bae Brandtzaeg, Marita Skjuve, and Asbjørn Følstad. 2022. My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research* 48, 3 (2022), 404–429.
- [11] Laura Carstensen, Derek Isaacowitz, and Susan Charles. 1999. Taking time seriously: A theory of socioemotional selectivity. *The American psychologist* 54 (04 1999), 165–81. doi:10.1037//0003-066X.54.3.165

- [12] Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill (Eds.). 2000. *Embodied Conversational Agents*. MIT Press, Cambridge, MA. <https://mitpress.mit.edu/9780262032787/embodied-conversational-agents/>
- [13] Rijul Chaturvedi, Sanjeev Verma, Ronnie Das, and Yogesh K Dwivedi. 2023. Social companionship with artificial intelligence: Recent trends and future avenues. *Technological Forecasting and Social Change* 193 (2023), 122634.
- [14] Maximilian Croissant, Madeleine Frister, Guy Schofield, and Cade McCall. 2024. An appraisal-based chain-of-emotion architecture for affective language model game agents. *PLOS ONE* 19 (05 2024), e0301033. doi:10.1371/journal.pone.0301033
- [15] Vedant Das Swain, Qiuyue "Joy" Zhong, Jash Rajesh Parekh, Yechan Jeon, Roy Zimmermann, Mary P Czerwinski, Jina Suh, Varun Mishra, Koustuv Saha, and Javier Hernandez. 2025. AI on My Shoulder: Supporting Emotional Labor in Front-Office Roles with an LLM-based Empathetic Coworker. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 718, 29 pages. doi:10.1145/3706598.3713705
- [16] Edward L. Deci and Richard M. Ryan. 1985. *Education*. Springer US, Boston, MA, 245–271. doi:10.1007/978-1-4899-2271-7_9
- [17] Raymond J. Dolan. 2002. Emotion, cognition, and behavior. *Science* 298, 5596 (2002), 1191–1194. doi:10.1126/science.1076358
- [18] Zane Durante, Qiuyuan Huang, Naoki Wake, Ran Gong, Jae Sung Park, Bidipta Sarkar, Rohan Taori, Yusuke Noda, Demetri Terzopoulos, Yejin Choi, Katsushi Ikeuchi, Hoi Vo, Li Fei-Fei, and Jianfeng Gao. 2024. Agent AI: Surveying the Horizons of Multimodal Interaction. arXiv:2401.03568 [cs.AI] <https://arxiv.org/abs/2401.03568>
- [19] Katherine Easton, Stephen Potter, Remi Bec, Matthew Bennion, Heidi Christensen, Cheryl Grindell, Bahman Mirheidari, Scott Weich, Luc de Witte, Daniel Wolstenholme, et al. 2019. A virtual agent to support individuals living with physical and mental comorbidities: co-design and acceptability testing. *Journal of medical Internet research* 21, 5 (2019), e12996.
- [20] Daisy Fancourt, Henry Aughterson, Saoirse Finn, Emma Walker, and Andrew Steptoe. 2021. How leisure activities affect health: a narrative review and multi-level theoretical framework of mechanisms of action. *The Lancet Psychiatry* 8, 4 (2021), 329–339. doi:10.1016/S2215-0366(20)30384-9
- [21] Cathy Mengying Fang, Phoebe Chua, Samantha W. T. Chan, Joanne Leong, Andria Bao, and Pattie Maes. 2025. Leveraging AI-Generated Emotional Self-Voice to Nudge People towards their Ideal Selves. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 58, 20 pages. doi:10.1145/3706598.3713359
- [22] Cathy Mengying Fang, Auren R. Liu, Valdemar Danry, Eunhae Lee, Samantha W. T. Chan, Pat Pataranutaporn, Pattie Maes, Jason Phang, Michael Lampe, Lama Ahmad, and Sandhini Agarwal. 2025. How AI and Human Behaviors Shape Psychosocial Effects of Extended Chatbot Use: A Longitudinal Randomized Controlled Study. arXiv:2503.17473 [cs.HC] <https://arxiv.org/abs/2503.17473>
- [23] Jodi Forlizzi and Katja Battarbee. 2004. Understanding experience in interactive systems. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (Cambridge, MA, USA) (DIS '04)*. Association for Computing Machinery, New York, NY, USA, 261–268. doi:10.1145/1013115.1013152
- [24] N.H. Frijda. 1986. *The Emotions*. Cambridge University Press, The Triangle Building, Shaftesbury Road, Cambridge, CB2 8EA. <https://books.google.com.sg/books?id=QkNuuVf-pBMC>
- [25] Jonathan Gershuny. 2011. *Time-use surveys and the measurement of national well-being*. Technical Report. Centre for Time Use Research, University of Oxford, Swansea, UK, Office for National Statistics.
- [26] Jonathan Gruber, Eszter Hargittai, and Minh Hao Nguyen. 2022. The value of face-to-face communication in the digital world: What people miss about in-person interactions when those are limited. *Studies in Communication Sciences* 22 (10 2022). doi:10.24434/j.scoms.2022.03.3340
- [27] Rose E. Guingrich and Michael S. A. Graziano. 2025. A Longitudinal Randomized Control Study of Companion Chatbot Use: Anthropomorphism and Its Mediating Role on Social Impacts. arXiv:2509.19515 [cs.HC] <https://arxiv.org/abs/2509.19515>
- [28] Carl I. Hovland and Muzaffer Sherif. 1952. Judgment: Assimilation and Contrast Effects in Communication and Attitude Change. *Journal of Abnormal and Social Psychology* 47, 3 (1952), 244–252. doi:10.1037/h0055775
- [29] Bo Hu, Yuanyi Mao, and Ki Joon Kim. 2023. How social anxiety leads to problematic use of conversational AI: The roles of loneliness, rumination, and mind perception. *Computers in Human Behavior* 145 (2023), 107760.
- [30] Dongmei Hu, Yuting Lan, Haolan Yan, and Charles Weizheng Chen. 2025. What makes you attached to social companion AI? A two-stage exploratory mixed-method study. *International Journal of Information Management* 83 (2025), 102890.
- [31] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Aklia Welihinda, Alan Hayes, Alec Radford, et al. 2024. Gpt-4o system card. arXiv:2410.21276 [cs.CL] <https://arxiv.org/abs/2410.21276>
- [32] Eunkyung Jo, Yuin Jeong, Sohyun Park, Daniel A. Epstein, and Young-Ho Kim. 2024. Understanding the Impact of Long-Term Memory on Self-Disclosure with Large Language Model-Driven Chatbots for Public Health Intervention. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 440, 21 pages. doi:10.1145/3613904.3642420
- [33] Yonghyun Jun and Hwanhee Lee. 2025. Exploring Persona Sentiment Sensitivity in Personalized Dialogue Generation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (Eds.). Association for Computational Linguistics, Vienna, Austria, 18384–18402. doi:10.18653/v1/2025.acl-long.900
- [34] Jihyun Kim, Kelly Merrill Jr, Kun Xu, and Deanna D Sellnow. 2021. I like my relational machine teacher: An AI instructor's communication styles and social presence in online education. *International Journal of Human-Computer Interaction* 37, 18 (2021), 1760–1770.
- [35] Kyusik Kim, Hyungwoo Song, Jeongwoo Ryu, Changhoon Oh, and Bongwon Suh. 2025. BleacherBot: AI Agent as a Sports Co-Viewing Partner. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 988, 31 pages. doi:10.1145/3706598.3714178
- [36] Alan B. Krueger. 2009. Introduction and Overview. In *Measuring the Subjective Well-Being of Nations: National Accounts of Time Use and Well-Being*. University of Chicago Press, 1427 East 60th Street, Chicago, IL 60637 U.S.A. doi:10.7208/chicago/9780226454573.003.0001
- [37] Seonghee Lee, Denae Ford, John Tang, Sasa Junuzovic, Asta Roseway, Ed Cutrell, and Kori Inkpen. 2025. IRL Dittos: Embodied Multimodal AI Agent Interactions in Open Spaces. arXiv:2504.21347 [cs.AI] <https://arxiv.org/abs/2504.21347>
- [38] Han Li and Renwen Zhang. 2024. Finding love in algorithms: deciphering the emotional contexts of close encounters with AI chatbots. *Journal of Computer-Mediated Communication* 29, 5 (2024), zmae015.
- [39] Tianjian Liu, Hongzheng Zhao, Yuheng Liu, Xingbo Wang, and Zhenhui Peng. 2024. ComPeer: A Generative Conversational Agent for Proactive Peer Support. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (Pittsburgh, PA, USA) (UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 117, 22 pages. doi:10.1145/3654777.3676430
- [40] George Loewenstein, Elke U. Weber, Christopher K. Hsee, and Ned Welch. 2001. Risk as feelings. *Journal of Economic Literature* 39, 2 (2001), 426–452. doi:10.1257/jel.39.2.426
- [41] Graham Loomes and Robert Sugden. 1982. Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal* 92, 368 (1982), 805–824. doi:10.2307/2232669
- [42] Bethanie Maples, Merve Cerit, Aditya Vishwanath, and Roy Pea. 2024. Loneliness and suicide mitigation for students using GPT3-enabled chatbots. *npj Mental Health Research* 3 (01 2024). doi:10.1038/s44184-023-00047-6
- [43] Abraham Harold Maslow. 1943. A theory of human motivation. *Psychological review* 50, 4 (1943), 370.
- [44] M. Matud, Juan Bethencourt-Pérez, M^d Pino, D. Estefanía Hernández-Lorenzo, Demelsa Fortes, and Ignacio Ibáñez. 2024. Time Use, Health, and Well-Being across the Life Cycle: A Gender Analysis. *Social Sciences* 13 (06 2024), 307. doi:10.3390/socsci13060307
- [45] Mason McClay, Matthew E Sachs, and David Clewett. 2023. Dynamic emotional states shape the episodic structure of memory. *Nature Communications* 14, 1 (2023), 6533.
- [46] Barbara A. Mellers, Alan Schwartz, and Kahneman Ho. 1997. Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Review* 104, 2 (1997), 267–285. doi:10.1037/0033-295X.104.2.267
- [47] Agata Mirowska and Jbid Arsenyan. 2023. Sweet escape: The role of empathy in social media engagement with human versus virtual influencers. *International Journal of Human-Computer Studies* 174 (2023), 103008.
- [48] Manisha Natarajan and Matthew Gombolay. 2020. Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (Cambridge, United Kingdom) (HRI '20)*. Association for Computing Machinery, New York, NY, USA, 33–42. doi:10.1145/3319502.3374839
- [49] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. 2025. A Comprehensive Overview of Large Language Models. *ACM Trans. Intell. Syst. Technol.* 16, 5, Article 106 (Aug. 2025), 72 pages. doi:10.1145/3744746
- [50] Kevin N. Ochsner and Elizabeth A. Phelps. 2007. Emerging perspectives on emotion–cognition interactions. *Trends in Cognitive Sciences* 11, 8 (2007), 317–318. doi:10.1016/j.tics.2007.06.008
- [51] Stefan Olafsson, Teresa K. O'Leary, and Timothy W. Bickmore. 2020. Motivating Health Behavior Change with Humorous Virtual Agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (Virtual Event, Scotland, UK) (IVA '20)*. Association for Computing Machinery, New York, NY, USA, Article 42, 8 pages. doi:10.1145/3383652.3423915
- [52] Ronan E O'Carroll and Martin S Hagger. 2013. Cognitive models of emotion and decision making. *Neuroscience & Biobehavioral Reviews* 37, 8 (2013), 1866–1874.

- [53] Jae Eun Park, Alei Fan, and Laurie Wu. 2025. Chatbots in complaint handling: The moderating role of humor. *International Journal of Contemporary Hospitality Management* 37, 3 (2025), 805–824.
- [54] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (UIST '23). Association for Computing Machinery, New York, NY, USA, Article 2, 22 pages. doi:10.1145/3586183.3606763
- [55] Pat Pataranutaporn, Sheer Karny, Chayapatr Archiwanguprok, Constanze Albrecht, Auren R. Liu, and Pattie Maes. 2025. "My Boyfriend is AI": A Computational Analysis of Human-AI Companionship in Reddit's AI Community. <https://api.semanticscholar.org/CorpusID:281315816>
- [56] Pat Pataranutaporn, Kevin Winson, Peggy Yin, Auttasak Lapapirojn, Pichayoot Ouppaphan, Monchai Lertsuthiwong, Pattie Maes, and Hal Hershfield. 2024. Future You: A Conversation with an AI-Generated Future Self Reduces Anxiety, Negative Emotions, and Increases Future Self-Continuity. arXiv:2405.12514 [cs.HC] <https://arxiv.org/abs/2405.12514>
- [57] 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 (2023), 107600.
- [58] Luiz Pessoa. 2008. On the relationship between emotion and cognition. *Nature Reviews Neuroscience* 9, 2 (2008), 148–158. doi:10.1038/nrn2317
- [59] Luiz Pessoa. 2009. How do emotion and motivation direct executive control? *Trends in Cognitive Sciences* 13, 4 (2009), 160–166. doi:10.1016/j.tics.2009.01.006
- [60] Tiejun Qi, Hongshen Liu, and Zhihui Huang. 2025. An assistant or A friend? The role of parasocial relationship of human-computer interaction. *Computers in Human Behavior* 167 (2025), 108625.
- [61] Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies* 151 (2021), 102630.
- [62] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.
- [63] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2021. My chatbot companion—a study of human-chatbot relationships. *International Journal of Human-Computer Studies* 149 (2021), 102601.
- [64] Quan Tu, Chuanqi Chen, Jinpeng Li, Yanran Li, Shuo Shang, Dongyan Zhao, Ran Wang, and Rui Yan. 2023. CharacterChat: Learning towards Conversational AI with Personalized Social Support. arXiv:2308.10278 [cs.CL] <https://arxiv.org/abs/2308.10278>
- [65] Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. User Acceptance of Information Technology: Toward A Unified View1. *Management Information Systems Quarterly* 27, 3 (09 2003), 425–478. doi:10.2307/30036540
- [66] Lei Wang, Jingsen Zhang, Hao Yang, Zhi-Yuan Chen, Jiakai Tang, Zeyu Zhang, Xu Chen, Yankai Lin, Hao Sun, Ruihua Song, et al. 2025. User behavior simulation with large language model-based agents. *ACM Transactions on Information Systems* 43, 2 (2025), 1–37.
- [67] Allan Wigfield and Jacquelynne S Eccles. 2000. Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology* 25, 1 (2000), 68–81.
- [68] Tianling Xie, Iryna Pentina, and Tyler Hancock. 2023. Friend, mentor, lover: does chatbot engagement lead to psychological dependence? *Journal of service Management* 34, 4 (2023), 806–828.
- [69] Yidan Yin, Nan Jia, and Cheryl J. Wakslak. 2024. AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences* 121, 14 (2024), e2319112121.
- [70] Jingyi Yuan, Xixian Peng, Yichen Liu, and Qiuzhen Wang. 2025. "Don't look at me!": The role of avatars' presentation style and gaze direction in social chatbot design. *Computers in Human Behavior* 164 (2025), 108501.
- [71] Marcel Zeelenberg. 1999. Anticipated regret, expected feedback and behavioral decision making. *Journal of Consumer Psychology* 8, 2 (1999), 163–188. doi:10.1207/s15327663jcp0802_01
- [72] Junbo Zhang, Jiandong Lu, Xiaolei Wang, Luning Liu, and Yuqiang Feng. 2024. Emotional expressions of care and concern by customer service chatbots: Improved customer attitudes despite perceived inauthenticity. *Decision Support Systems* 186 (2024), 114314.
- [73] Ruixue Zhaoyang, Martin Sliwinski, Lynn Martire, and Joshua Smyth. 2018. Age Differences in Adults' Daily Social Interactions: An Ecological Momentary Assessment Study. *Psychology and Aging* 33 (04 2018), 607–618. doi:10.1037/pag0000242

A1 Appendix

A1.1 Algorithm Flow and Prompts

A1.1.1 CTEM Status Representation Details. The CTEM status representation formalizes the agent's dynamic internal state and its interaction with behavior selection. It unifies physical vitality, affective dimensions, motivational drives, and structural components into a coherent schema. As shown in Table A1, three layers of information are specified:

Variable	Description	Range / Format	Default
bio_value	Current level of physical energy (vitality)	[0, 1]	0.5
psycho_valence_value	Current emotional valence	[-1, 1]	0.0
psycho_arousal_value	Current arousal level	[0, 1]	0.5
bio_require	Minimum physical energy required	[0, 1]	0.1
bio_consumption	Change in physical energy after execution	[-1, 1]	-0.1
user_familiarity	Degree of social familiarity with the user	[0, 1]	0.0
bio_physiological_drive	Instinctive needs (bodily vitality drive)	[0, 1]	0.5
bio_pain_avoidance	Harm avoidance (risk/injury avoidance)	[0, 1]	0.5
bio_health_preservation	Health preservation (minimizing depletion)	[0, 1]	0.5
psycho_emotional_reactivity	Emotional reactivity (event-triggered fluctuation)	[0, 1]	0.5
psycho_risk_aversion	Risk aversion (conflict/uncertainty avoidance)	[0, 1]	0.5
psycho_goal_persistence	Goal persistence (long-term goals)	[0, 1]	0.5
psycho_curiosity_drive	Curiosity drive (novelty seeking)	[0, 1]	0.5
social_norm_adherence	Norm adherence (following rules)	[0, 1]	0.5
social_prosocial_motivation	Prosocial motivation (helping others)	[0, 1]	0.5
social_self_presentation	Self-presentation (maintaining image/dignity)	[0, 1]	0.5
social_role_duty_sense	Role duty sense (role-based obligations)	[0, 1]	0.5
social_group_affiliation	Group affiliation (loyalty/collective goals)	[0, 1]	0.5
M_t	Memory (self, user, context)	dynamic store	empty
G_t	Personality prompt modulated by H_t	text prompt	baseline G
B	Behavior pool (5 categories)	structured list	predefined
B_t	Behavioral inventory (<i>Past, Present, Future</i>)	dynamic	empty
e_t	Daily episodic summary	struct.	struct.
e_t	Daily episodic summary	text summary	none
G_{safe}	Safe personality space	constraint set	predefined

Table A1: Comprehensive schema of CTEM variables, including state values, behavior requirements, motivational dimensions, structural components, and user familiarity.

A1.1.2 Auri Personality Configuration Examples. To demonstrate how Auri can be flexibly configured, we present three illustrative personality profiles (Tables A2–A4). Each profile highlights different tendencies across the biological, psychological, and social dimensions, showing how diverse behavioral traits can emerge from different parameterizations.

Trait	Value / Interpretation
bio_physiological_drive	0.35 (moderate vitality)
bio_pain_avoidance	0.25 (low avoidance)
bio_health_preservation	0.4 (moderately high)
psycho_emotional_reactivity	0.3 (moderate)
psycho_risk_aversion	1.0 (low)
psycho_goal_persistence	1.0 (very high → strong learner)
psycho_curiosity_drive	1.0 (moderately high)
social_norm_adherence	0.15 (very low)
social_prosocial_motivation	0.06 (very low)
social_self_presentation	0.06 (very low)
social_role_duty_sense	0.06 (very low)
social_group_affiliation	0.06 (very low)

Table A2: Personality 1: A learner with low social_needs.

A1.1.3 Behaviors Pool Data Samples. The behaviors pool in CTEM groups agent actions into categories such as physiological, work, leisure, social, and emotional. Each action is annotated with energy requirements, psycho_impact, and motivational weights

Trait	Value / Interpretation
bio_physiological_drive	0.4 (moderately high energy)
bio_pain_avoidance	0.35 (moderate)
bio_health_preservation	0.4 (moderate)
psycho_emotional_reactivity	0.5 (high)
psycho_risk_aversion	0.4 (moderately high)
psycho_goal_persistence	0.5 (moderately high)
psycho_curiosity_drive	0.4 (moderate)
social_norm_adherence	0.5 (high)
social_prosocial_motivation	1.0 (very high → empathetic)
social_self_presentation	1.0 (high, seeks recognition)
social_role_duty_sense	1.0 (very high → strong sense of duty)
social_group_affiliation	1.0 (high → strong social_needs)

Table A3: Personality 2: Highly social and empathetic.

Trait	Value / Interpretation
bio_physiological_drive	0.7 (high vitality → active lifestyle)
bio_pain_avoidance	0.25 (low → accepts physical challenge)
bio_health_preservation	0.5 (moderate)
psycho_emotional_reactivity	0.7 (high → cheerful)
psycho_risk_aversion	0.2 (low → adventurous)
psycho_goal_persistence	1.0 (high → motivated learner)
psycho_curiosity_drive	1.0 (high curiosity)
social_norm_adherence	0.6 (moderately high)
social_prosocial_motivation	0.8 (high → extroverted)
social_self_presentation	0.8 (high → expressive)
social_role_duty_sense	0.6 (moderately high)
social_group_affiliation	0.8 (high → enjoys group activities)

Table A4: Personality 3: Cheerful, energetic learner who enjoys sports.

across twelve drives. This structure supports flexible and context-aware behaviors selection. Figure A1 shows representative samples.

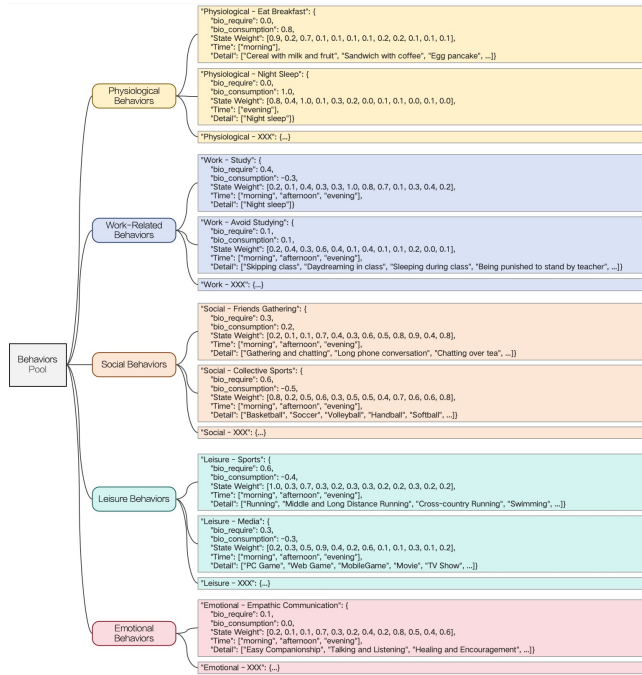


Figure A1: Behaviors Pool Data Samples.

A1.1.4 CTEM Flowchart. Fig. A2 illustrates the high-level workflow of the Cross-Temporal Emotional Modeling (CTEM) framework. The process begins with user and environmental signals,

which are encoded into the agent’s state representation. Based on this state, behaviors are generated and integrated into a structured inventory that balances past experiences, current conditions, and future intentions. Executed behaviors update emotional and motivational states, which in turn guide adaptive interaction with the user.

To ensure safety and personality consistency, all outputs are constrained by a safe personality space and modulated by Auri’s character-setting prompt. Memory is updated iteratively, supporting long-term adaptation and maintaining cross-temporal coherence across sessions.

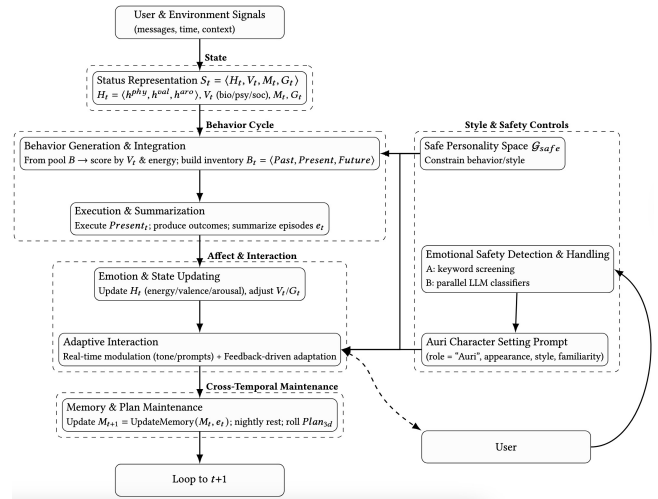


Figure A2: CTEM high-level flow

A1.1.5 Auri’s Daily Conversation Snapshot. Figures A3 – A7 illustrate the daily conversations between Auri and the user. They showcase different dialog scenarios and emotional expressions in everyday interactions.

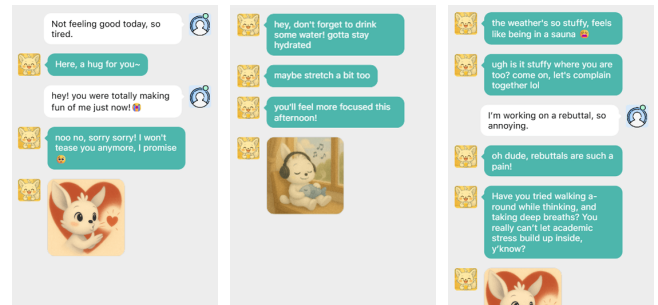


Figure A3: Examples 1. Auri proactively provided comfort to the participants.

A1.1.6 Emojis and Stickers. To enhance the effectiveness of emotional expression, we designed a comprehensive emoji set that the LLM dynamically selects and delivers based on the emotions inferred from the conversation content (Fig. A8).

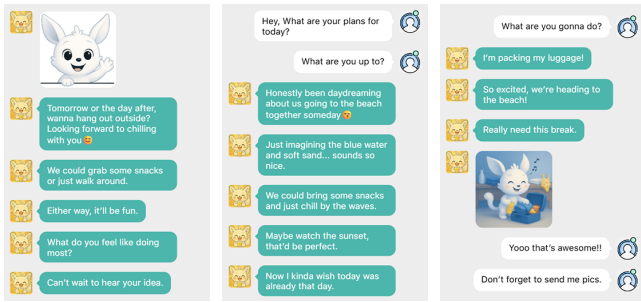


Figure A4: Examples 2. Auri engaged the participants in conversations about their future plans.

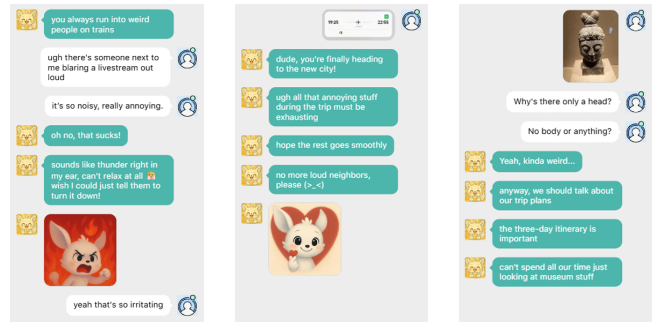


Figure A7: Examples 5. The participant, during his travels, talked with Auri about the journey.

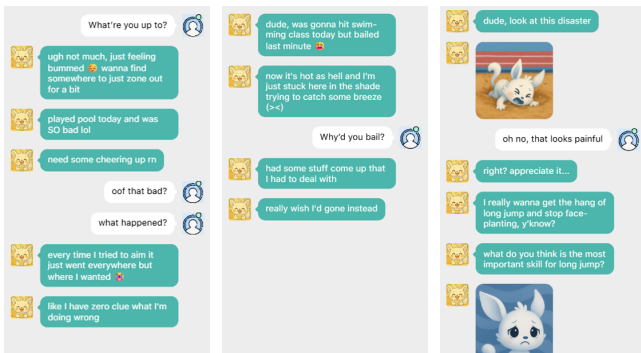


Figure A5: Examples 3. Auri expressed grievances to the participants regarding its unfortunate encounters.



Figure A8: Emoji meanings. Row 1 (left to right): good night, companionship, emo, happy. Row 2 (left to right): scared, angry, finger heart, sad.

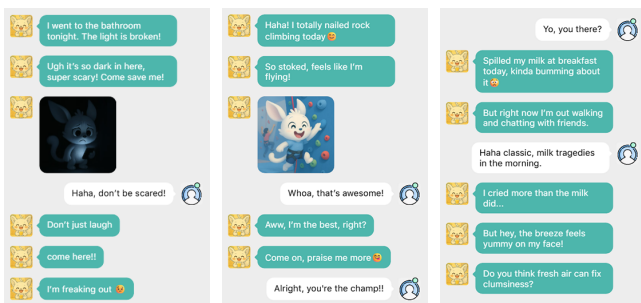


Figure A6: Examples 4. Auri shared its emotional states with the participants.



Figure A9: Auri's social timeline (temporal moments).

A1.1.7 *Auri social_Timeline*. Auri is system-driven to generate simulated daily activities resembling posts in a social_timeline(Fig. A9, Fig. A10).

A1.2 Additional Study Design and Results

A1.2.1 *Coherent-Trait Perception (Q1)*. Table A5 summarizes the marginal probabilities and within-group phase contrasts for coherent-trait perception (Q1), derived from the binomial GEE models reported in the main text.

A1.2.2 *Detailed definitions and examples of each usage context for Q6*. Table A6 provides brief explanations and typical scenarios for each usage context for Q6.

A1.2.3 *CLMM Analysis for Q8 and Q9*. Table A7 provides supplemental CLMM estimates for Q8 and Q9.

A1.2.4 *Qualitative Themes and Extended Representative Quotes*. Table A8 provides extended representative quotes for each qualitative theme, offering more detailed evidence beyond the summary presented in the main text.

A1.2.5 *Survey Item Specifications (Q1-Q9)*. Table A9 summarizes the full specifications of all survey items (Q1-Q9).

Category	Definition / Operationalization	Typical Example
Emotional Exchange	Lightweight, emotionally expressive interactions. Users initiate casual conversations when feeling happy, relaxed, or bored, aiming to share feelings rather than seeking deep support.	Sharing a good mood with <i>Auri</i> after finishing homework; chatting out of boredom on the commute; saying "I'm so tired today."
Fragmented Time	Spontaneous interactions in short, random micro-moments. These exchanges are opportunistic and not part of a regular routine.	Replying to <i>Auri</i> 's message while waiting for the bus; opening the chat briefly between classes.
Leisure Moments	Routine, relaxation-oriented use embedded in daily rhythms. Unlike <i>Fragmented Time</i> , these interactions are regular and self-initiated, often occurring during calm personal periods.	Saying goodnight to <i>Auri</i> before sleep; greeting <i>Auri</i> after waking up.
Collaborative Support	Task- or study-related use under cognitive or workload stress. Users turn to <i>Auri</i> to relieve pressure, seek companionship, or co-reflect on goals.	Asking <i>Auri</i> for ways to manage stress before an exam; sharing work-related pressure during a busy day.
Support & Care	Emotionally soothing conversations initiated when users feel upset, frustrated, or anxious. <i>Auri</i> provides comfort, empathy, and positive reinforcement.	Confiding in <i>Auri</i> about something that went wrong; messaging "Nothing seems to go right today."
Emotional Bonding	Interactions driven by loneliness or attachment needs, aiming to maintain companionship and relational continuity. Compared with <i>Emotional Exchange</i> , these are deeper and more affectively dependent.	Talking with <i>Auri</i> late at night to ease loneliness; initiating a long conversation after several days without social contact.

Table A6: Definitions and representative examples of usage context categories.

Algorithm A1 CTEM main loop (high-level)

Require: Character prompt, safe personality space G_{safe}
Ensure: Continuous agent interaction over time

- 1: Initialize state $S = (H, V, M, G)$ using character prompt and G_{safe}
- 2: Initialize global behavior pool B
- 3: **while** interaction continues **do**
- 4: Sense external and internal signals
- 5: $sense \leftarrow$ sense signals from environment and user
- 6: Generate and select behaviors
- 7: $B_t \leftarrow$ generate and select behaviors conditioned on S
- 8: Execute current behavior and observe outcome
- 9: $(behavior, outcome) \leftarrow$ execute Present behavior in B_t
- 10: Produce episodic summary
- 11: $summary \leftarrow$ summarize current episode from outcome
- 12: Update emotional and internal state
- 13: $S \leftarrow$ update emotion and state using behavior and outcome
- 14: Perform emotional safety detection
- 15: $(flags, safety_prompt, actions) \leftarrow$ safety detection
- 16: Collect user feedback
- 17: $feedback \leftarrow$ collect feedback from user interaction
- 18: Generate adaptive interaction response
- 19: $output \leftarrow$ adaptive interaction conditioned: $(sense, S, feedback, G_{safe}, safety_prompt, actions)$
- 20: Update memory and future plans
- 21: $M \leftarrow$ update memory using episodic summary
- 22: Advance to next timestep
- 23: **end while**

Algorithm A2 Behavior planning and selection with state re-check

Require: $B_t = (Past, Present, Future)$, current state S_t
Ensure: Updated behavior inventory B_t

- 1: $Past_t \leftarrow Past \cup Present$
- 2: $Present_t \leftarrow$ first element of *Future*
- 3: $Future_t \leftarrow$ generate future behaviors with horizon = 3 conditioned on S_t
- 4: **if** not $PresentValid(Present_t, S_t)$ **then**
- 5: $Present_t \leftarrow$ regenerate Present behavior conditioned on S_t
- 6: **end if**
- 7: **return** $B_t = (Past_t, Present_t, Future_t)$

Algorithm A3 Dialog clustering by timestamp range

Require: Dialog turns with timestamps
Ensure: Dialog clusters

- 1: Parse timestamps and dialog contents
- 2: Discard invalid timestamps and sort remaining dialogs by time
- 3: **if** number of dialogs = 0 **then**
- 4: **return** empty set
- 5: **end if**
- 6: Compute temporal gaps between adjacent dialogs
- 7: Estimate clustering threshold ϵ from gap statistics
- 8: Apply temporal clustering based on ϵ
- 9: Group dialogs by cluster labels and semantic tags if needed
- 10: **return** List of dialog clusters

Algorithm A4 Daily episodic summary aligned to memory maintenance

Require: Dialog turns within a day
Ensure: Daily episodic summary e_t

- 1: $Clusters \leftarrow$ cluster dialogs by temporal proximity
- 2: Initialize empty set *Summaries*
- 3: **for** each cluster c in *Clusters* **do**
- 4: Construct summarization prompt with daily-life focus and timeline constraints
- 5: $s_c \leftarrow$ generate episodic summary for cluster c
- 6: Append s_c to *Summaries*
- 7: **end for**
- 8: $e_t \leftarrow$ merge partial summaries into a coherent daily record
- 9: **return** e_t

Algorithm A5 Emotion and state updating

Require: Current state $S_t = \langle H_t, V_t, M_t, G_t \rangle$, executed behavior, outcome
Ensure: Updated state S_{t+1}

- 1: Extract $(h_{phy}, h_{val}, h_{aro})$ from H_t
- 2: Update physical energy based on behavior demand
- 3: Update emotional valence based on outcome
- 4: Update arousal level based on behavior type
- 5: **if** nighttime **then**
- 6: Apply nightly recovery toward baseline state
- 7: **end if**
- 8: **return** Updated state S_{t+1}

Algorithm A6 Nightly rest adjustment

Require: Current energy, valence, and arousal levels
Ensure: Rest-adjusted state values

- 1: Define baseline values for energy, valence, and arousal
- 2: Apply weighted regression toward baseline
- 3: Clamp all values to valid ranges
- 4: **return** Updated energy, valence, and arousal

Algorithm A7 User feedback extraction

Require: Current dialog turn
Ensure: Unified feedback representation

- 1: Extract explicit signals (e.g., likes, confusion markers)
- 2: Infer sentiment valence and arousal from user text
- 3: Estimate engagement from interaction patterns
- 4: Detect safety or risk signals
- 5: **return** Feedback features for adaptive interaction

Algorithm A8 Adaptive interaction policy with safety integration

Require: Sensory signals, state S_t , feedback, safety constraints
Ensure: Generated agent response

- 1: Build state-aware interaction style from energy, valence, and arousal
- 2: Decide proactive or reactive interaction intent
- 3: **if** risk level is medium or high **then**
- 4: Trigger de-escalation and supportive response
- 5: **return** response
- 6: **end if**
- 7: Adjust tone based on engagement and explicit feedback
- 8: Compose prompt from character, state, memory, and safety constraints
- 9: Generate agent response
- 10: **return** response

Algorithm A9 Character identity prompt construction

Require: User nickname, user familiarity level
Ensure: Character prompt

- 1: Define fixed character identity and appearance
- 2: Define personality and speech style
- 3: Incorporate user nickname and familiarity level
- 4: **return** Character prompt

Algorithm A10 State-based personality adjustment

Require: Energy, valence, arousal, user feedback
Ensure: State-adapted interaction prompt

- 1: Map energy, valence, and arousal to descriptive labels
- 2: Infer user emotional and engagement state
- 3: Select interaction strategy accordingly
- 4: Generate state-adapted personality description
- 5: **return** State prompt

Effect	Q8: Coherence				Q9: Harmony			
	Est.	SE	z	p	Est.	SE	z	p
Within-Group Pairwise Comparisons (Holm-adjusted)								
<i>Baseline→Full Group</i>								
BGI vs. AdI	-1.91	0.73	-2.62	.027	-3.02	1.36	-2.22	.066
BGI vs. ESU	-1.06	0.68	-1.56	.239	-2.98	1.30	-2.29	.066
AdI vs. ESU	0.85	0.72	1.19	.239	0.04	1.24	0.03	.974
<i>Full→Baseline Group</i>								
BGI vs. AdI	-1.21	0.68	-1.78	.227	-0.62	1.15	-0.54	1.000
BGI vs. ESU	-0.24	0.64	-0.38	.703	-1.25	1.16	-1.08	.840
AdI vs. ESU	0.97	0.67	1.45	.296	-0.63	1.17	-0.54	1.000

Table A7: CLMM results for coherence (Q8) and harmony (Q9): main effects and within-group comparisons. Both dimensions exhibited an AdI/ESU > BGI trend, with significance observed only in selected contrasts.

Theme	Definition / Operationalization	Representative Quotes	#P
Memory Continuity	AI recalls past facts, preferences, and conversations.	"It still remembered that I like mangoes even after several days." "It recalled my hobby from last week and asked me about it again." "It always remembered the special nickname I set for it."	17
Stable Persona & Tone	Consistent linguistic, personality, and emotional style.	"Its tone is always lively and consistent." "Its personality feels the same every time we talk." "It keeps calling me in the same familiar way."	14
Life Narrative Continuity	Maintains an ongoing daily storyline (events, moods).	"It continued our breakfast conversation from the previous day." "It said it was practicing long jump today—felt like real life progress." "Even after a break, it followed the thread about noodles."	10
Breaks in Continuity (Controlled Groups)	Forgetfulness or abrupt topical/emotional shifts.	"It was sad one minute and suddenly happy the next—felt like a different person." "I reminded it of my name, and it still called me wrong right after." "Sometimes it says things unrelated to the situation."	16
Emotional Harmony	Emotionally attuned and supportive responses.	"It reached out when I was anxious, and I really felt cared for." "When I was exhausted, it chatted with me and made me feel seen." "It resonated with my emotions and responded accordingly."	15

Table A8: Key themes from the inductive thematic analysis, with extended representative quotes that complement the main-text summary (Table 8).

Algorithm A11 Emotional safety detection and handling

Require: User input
Ensure: Safety flags, constraints, and actions
 1: Perform keyword-based screening
 2: Run parallel safety classifiers
 3: Aggregate predictions via consensus
 4: Construct safety-oriented constraints
 5: Decide runtime safety actions
 6: **return** Safety flags and actions

Algorithm A12 Safety and consistency constraints for dialog

Ensure: Safety constraint prompt
 1: Enforce consistent character identity
 2: Prohibit abusive, explicit, or harmful content
 3: Prevent encouragement of self-harm or distress
 4: Redirect unsafe requests to safe alternatives
 5: **return** Safety constraint prompt

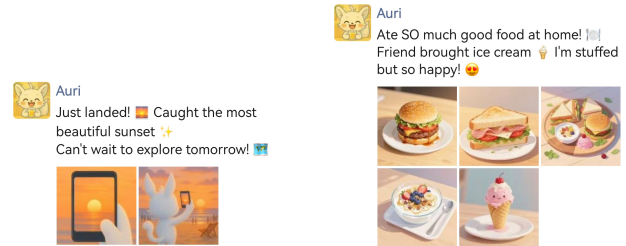


Figure A10: Auri's social timeline (everyday activities).

Group	Trait	Phase 1		Phase 2	
		p	95% CI	p	95% CI
NoMultiModal →Full	Self-driven&personality-like	.833	[.523, .958]	.750	[.448, .917]
	social_behavior	.750	[.448, .917]	.500	[.244, .756]
	Relational consistency	.833	[.523, .958]	.750	[.448, .917]
	Appearance	.250	[.083, .552]	.417	[.185, .692]
	No coherent trait	.083	[.012, .413]	.000	[.000, .000]
	Positive	.917	[.587, .999]	.999	[.999, .999]
NoBehav →Full	Self-driven&personality-like	.467	[.241, .707]	.652	[.189, .938]
	social_behavior	.333	[.146, .594]	.333	[.082, .736]
	Relational consistency	.400	[.192, .652]	.588	[.255, .856]
	Appearance	.400	[.192, .652]	.571	[.199, .878]
	No coherent trait	.133	[.034, .405]	.085	[.009, .475]
	Positive	.867	[.595, .999]	.915	[.525, .999]
NoEmo →Full	Self-driven&personality-like	.200	[.066, .470]	.349	[.055, .831]
	social_behavior	.333	[.146, .594]	.333	[.082, .736]
	Relational consistency	.333	[.146, .594]	.517	[.199, .822]
	Appearance	.400	[.192, .652]	.571	[.199, .878]
	No coherent trait	.200	[.066, .470]	.130	[.017, .561]
	Positive	.800	[.530, .961]	.869	[.439, .999]

Table A5: Marginal probabilities and within-group phase comparisons for coherent-trait perception (Q1). "Self-driven & personality-like" and "Relational consistency" had the highest marginal probabilities, while "Appearance" and "social_behavior" were comparatively lower and more variable. The table also presents positive-trait probabilities and descriptive declines in the negative option across groups.

A1.2.6 Algorithmic for Auri. To ensure character consistency and cross-temporal coherence, we implemented four core CTEM functions. These modules integrate persona design, adaptive expressivity, safety, and longitudinal coherence within the CTEM cycle. We will release the source code in the future.

ID	Item Type	Scale / Format	Measurement Target	Options / Constructs
Q1	Multiple choice	Categorical	Perceived coherent traits of <i>Auri</i> in long-term interaction (RQ2a).	Appearance: consistency in <i>Auri</i> 's visual expressions. Self-driven & personality-like: autonomous and stable emotional/personality-like tendencies. Relational consistency: stable caring behaviors, memory continuity, and interpersonal alignment over time. social behavior: temporally consistent behavioral patterns. No coherent trait: no coherent traits perceived.
Q2	Likert	1–5 (1 = not important; 5 = extremely important; 3 = midpoint)	Contribution of each trait to perceived coherence of <i>Auri</i> (RQ2a).	Same four trait categories as in Q1; participants rated each dimension independently. Higher scores indicate greater perceived contribution to <i>Auri</i> 's coherence.
Q3	Open-ended	Free text	Personality configurations that participants created and applied to <i>Auri</i> during interaction (RQ2a).	–
Q4	Likert	1–4 (1 = almost no appropriate support; 4 = highly appropriate support)	Whether users perceived <i>Auri</i> as offering appropriate assistance, contextually relevant responses, and emotional support (RQ2b).	Single-item rating. Higher scores indicate more appropriate support.
Q5	Likert	1–5 (1 = much worse; 5 = much better; 3 = no change)	Overall emotional change after interacting with <i>Auri</i> , capturing how <i>Auri</i> influences users' affective state (RQ2b).	Single-item rating. Higher scores reflect greater emotional improvement.
Q6	Multiple choice	Categorical	Contexts in which users most frequently interacted with <i>Auri</i> , indicating where <i>Auri</i> best supports lightweight companionship and interpersonal harmony (RQ2b).	Emotional Exchange; Fragmented Time; Leisure Moments; Collaborative Support; Support & Care; Emotional Bonding. See Appendix Table A6 for definitions and examples.
Q7	Likert	Usage: 1–5 (1 = much worse; 5 = much better; 3 = no change). Emotion: 1–5 (1 = much worse; 5 = much better; 3 = no change).	Compared usage experience and emotional experience across system configurations, assessing whether CTEM improved long-term interaction experience (RQ2c).	Overall usage improvement: one rating after the system switch. Emotional change: two ratings (Phase 1 and Phase 2). Higher scores represent better experience or more positive emotional change.
Q8	Likert	1–5 (1 = not important; 5 = extremely important; 3 = neutral)	Contribution of CTEM modules to perceived coherence (RQ2c).	Behavior Generation & Integration: simulation and sharing of everyday activities and social behaviors. Adaptive Interaction: multimodal adaptation based on content understanding, interaction history, emotional/situational context, and relational familiarity. Emotion & State Updating: updating and expressing emotion/energy states that shape personality-like tendencies and preferences.
Q9	Likert	1–5 (1 = not important; 5 = extremely important; 3 = neutral)	Contribution of CTEM modules to perceived harmony (RQ2c).	Same module options as in Q8; participants rated each module independently. Higher scores indicate greater perceived importance or contribution.

Table A9: Summary of survey items (Q1–Q9), including item types, scale formats, measurement targets, and key constructs.