MODELING WHAT PERSISTS: LEARNING EXPLICIT AFFECTIVE STATES FOR CHARACTER SIMULATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Simulating a character across extended interactions requires more than producing locally appropriate responses; some aspects of the character must persist over time. Existing systems largely condition generation on profiles, memories, and dialogue history, while affect is often inferred only for the current response. This raises a basic question: *should affect be treated as a transient generation cue or as a persistent state of the simulated character?* We study the latter and propose **Unified Explicit Interoception Modeling (UEIM)**, which separates *affective state tracking* from *behavior realization*. UEIM maintains a continuous Valence–Arousal state anchored by fine-grained emotion prototypes and updates it with each interaction event. Because such latent states lack reliable continuous supervision, we introduce behavior-grounded state–behavior alignment: teacher-derived trajectories initialize state transitions, while a frozen behavior generator and affective reader provide proxy feedback through observable affective consequences. Experiments on DailyDialog and ESConv show gains in response realization and affective transfer; frozen-A1 studies further improve negotiation success and affective attunement in counseling and customer service. These results support modeling not only *what emotion should be expressed now*, but also *what affective state has persisted and how the current event changes it*. Our code is available at https://anonymous.4open.science/r/UEIM-0F7E.

1 INTRODUCTION

Large language models (LLMs) are increasingly used to simulate individuals and characters across extended interactions. Existing character-simulation and role-playing systems condition generation on profiles, memories, dialogue histories, and prior experiences to preserve role-specific knowledge, personality, style, and behavioral patterns (Liu et al., 2024; Lu et al., 2024; Wang et al., 2024b; Yu et al., 2025; Wang et al., 2025), while interactive social benchmarks increasingly evaluate agents through sustained interaction (Zhou et al., 2024; Wang et al., 2024a). This shift raises a basic question: *when an agent is expected to remain the same character over time, what should persist across turns?* Plausible responses at individual turns do not necessarily imply a coherent trajectory. In particular, affect may depend not only on the current event but also on what has accumulated through prior interactions; psychological studies similarly characterize affect as a temporally evolving process rather than a sequence of independent snapshots (Russell, 1980; Kuppens et al., 2010). An agent may therefore appear locally appropriate while exhibiting abrupt or internally inconsistent affective changes across interactions.

Figure 1 contrasts these roles and previews UEIM's separation between persistent state tracking and current-turn behavior realization. In *behavior-driven* character simulation and role playing, affect may emerge in generated language, but explicit modeling typically centers on persona, memory, style, knowledge, or profile–dialogue consistency (Liu et al., 2024; Lu et al., 2024; Wang et al., 2024b; Yu et al., 2025; Wang et al., 2025); stronger role-playing ability does not necessarily imply stronger emotional fidelity (Feng et al., 2025). A complementary *task-driven* line models emotion or psychological dynamics to support planning, response generation, emotional support, or evaluation (Liu et al., 2021; Chen et al., 2025; Li et al., 2026; Liu et al., 2026; Tan et al., 2026). Rather than asking whether an agent *uses* affect, we ask whether affect should be a cue for the current response or a state of the simulated character that persists across turns. We pursue the latter view and distinguish *what the character carries forward* from *what it expresses at the current turn*.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

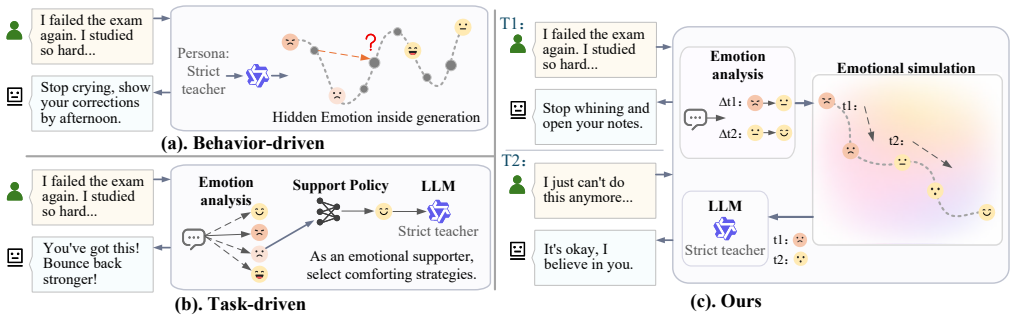

Figure 1: Comparison of affect in character-oriented agents. (a): Behavior-driven systems keep affect implicit. (b): Task-driven systems model affect to support a downstream objective. (c): UEIM maintains affect as a recurrent character state that conditions behavior across turns.

This perspective raises a representation challenge. A persistent affective state should support gradual transition while remaining semantically interpretable across interactions. Discrete emotion categories provide clear semantics but poorly represent relative distance, intensity, and incremental change, making nearby affective transitions difficult to distinguish. Purely continuous variables are easier to update and accumulate over time, but they lack explicit semantic structure for downstream behavior realization. Motivated by the Circumplex Model of Affect (Russell, 1980), we propose **Unified Explicit Interoception Modeling (UEIM)**, which maintains a continuous Valence–Arousal (V–A) state anchored by fine-grained emotion prototypes. At each turn, UEIM predicts an event-conditioned displacement from the previous state and projects the updated state onto a semantic prototype for behavior generation. This design lets the continuous state preserve graded cross-turn accumulation before discretization, while the prototype provides a stable semantic interface for response realization. Continuous coordinates therefore encode *how the state changes*, whereas prototypes describe *what the resulting state represents*. Here, "interoception" denotes a model-maintained affective condition rather than physiological sensing or recovery of a person's true internal state.

A harder challenge is supervision: the persistent state itself is not directly observed. Dialogue corpora provide utterances and sometimes categorical emotion labels, but generally lack turn-level continuous affective trajectories. Teacher-generated coordinates offer one initialization, yet their numerical scale may vary across contexts, and locally plausible predictions need not compose into a coherent recurrent state. This motivates the key learning observation of UEIM: *the maintained state is latent, but the behavior induced by that state is observable*. Observable behavior can therefore provide a proxy signal for calibrating otherwise unobserved state transitions.

UEIM realizes this idea through *behavior-grounded state–behavior alignment*. A state controller predicts the affective transition, a state-conditioned generator realizes the resulting semantic affect, and an affective reader maps generated behavior back into the same prototype space. Teacher-derived trajectories initialize transition capability; the generator and reader are then frozen while the controller is calibrated from the affective consequences of its induced behaviors. This provides a comparatively stationary state-to-behavior interface during calibration. The resulting state is not assumed to be uniquely identifiable or psychologically "true", but is optimized as a recurrent representation whose transitions induce behavior compatible with the observed character response.

Our contributions are threefold:

- We study long-horizon character simulation from a *persistent affective-state* perspective, explicitly separating recurrent state tracking from current-turn behavior realization and highlighting the role of accumulated affect beyond short local context.
- We propose **UEIM**, which combines a recurrent continuous V–A state, semantically grounded emotion prototypes, and event-conditioned updates to capture gradual affective transitions while retaining an interpretable interface for behavior generation.
- We introduce *behavior-grounded state–behavior alignment*, which initializes latent dynamics with teacher-derived trajectories and further calibrates otherwise unobserved state transitions through frozen behavior realization and affective readout, grounding persistent states in their observable behavioral consequences.

Table 1: Comparison of cross-turn affective and character-simulation approaches.

Dimension	RoleAgent	EmoDynamiX	DyCoP	CTEM	UEIM
Primary focus	Role simulation	Strategy pred.	Guidance	Companion	**State modeling**
Cross-turn carrier	Memory	Emotion–strategy	Psych. state	Emotion/memory	**Recurrent affect**
State representation	–	Mixed emotions	Cognitive state	Emotional state	**V–A + prototypes**
Update / signal	Memory update	Task supervision	Functional obj.	Behavior + feedback	**Teacher + readout**

2 Related Work

Emotional Support and Affective Process Modeling Emotional-support dialogue provides a natural setting for studying affect across turns. ESConv formulates multi-turn support with emotion and strategy annotations (Liu et al., 2021), while recent work increasingly models affect beyond isolated turn-level predictions. EmoDynamiX captures emotion–strategy dynamics for strategy prediction (Wan et al., 2025); DyCoP models evolving cognitive and psychological states for dynamic guidance (Liu et al., 2026); and Tan et al. (2026) evaluate emotional support through trajectory-level affect dynamics. These approaches introduce meaningful cross-turn structure, but primarily use affective variables to support downstream objectives such as strategy selection, guidance, or evaluation. UEIM instead treats affect itself as a recurrent character state that must be maintained, updated, and behaviorally realized across interactions.

Character Simulation and Persistent Affect Character-simulation and role-playing systems preserve consistency through profiles, memories, scripts, and prior experiences (Liu et al., 2024; Wang et al., 2024b; 2025), while EmoCharacter shows that stronger role-playing ability does not necessarily imply stronger emotional fidelity (Feng et al., 2025). Closest to our motivation, CTEM maintains an evolving emotional state in companion interaction, where behavioral history updates emotion and user feedback revises memory and state (Zheng et al., 2026). As summarized in Table 1, CTEM emphasizes cross-temporal coupling among emotion, behavior, feedback, and memory, whereas UEIM asks how an explicit recurrent affective state can be *represented and learned when continuous state trajectories are unobserved*. UEIM combines continuous V–A coordinates with semantic prototypes and calibrates event-conditioned transitions through behavior-mediated feedback. Our contribution lies not in introducing dynamic affect itself, but in explicit recurrent state representation and behavior-grounded calibration of unobserved state transitions.

3 Problem Formulation and Optimization Overview

3.1 Persistent Affective-State Modeling

We study how to maintain an explicit affective state across multi-turn character interactions. Let $H_t = \{(u_1, r_1), \ldots, (u_{t-1}, r_{t-1})\}$ denote the interaction history before turn t, where u_t is the current input, r_t is the target character's response, and g denotes the task condition. Rather than predicting an independent emotion at each turn, we maintain a recurrent affective state $\mathbf{s}_t = (v_t, a_t) \in [-1, 1]^2$ in Valence–Arousal space (Russell, 1980). The state represents affect carried forward by the target character rather than the transient emotion expressed by a single utterance. Given $\mathbf{s}_{t-1}$, H_t, and u_t, a state controller f_θ predicts how the current interaction changes the existing state:

$$\mathbf{s}_t = \mathcal{M}(\mathbf{s}_{t-1}, f_\theta(H_t, u_t, \mathbf{s}_{t-1}, g)), \qquad k_t = \mathcal{P}(\mathbf{s}_t), \qquad r_t \sim \pi_\phi(\cdot \mid H_t, u_t, k_t, g), \tag{1}$$

where $\mathcal{M}$ is the recurrent update operator, $\mathcal{P}$ maps the continuous state to a semantic prototype $k_t \in \mathcal{K}$, and π_ϕ realizes that condition as observable behavior. This separates *affective state tracking* from *behavior realization*: the former determines what is carried forward and how it changes, while the latter determines how that state is expressed in the current interaction.

3.2 Optimization Overview

UEIM contains a state controller f_θ, a state-conditioned behavior generator π_ϕ, and an affective reader h_η that maps observable behavior back to the shared prototype space. Training follows two phases. First, *capability preparation* initializes transition inference, state-conditioned generation, and affective readout through supervised fine-tuning and prototype-level alignment. Second,

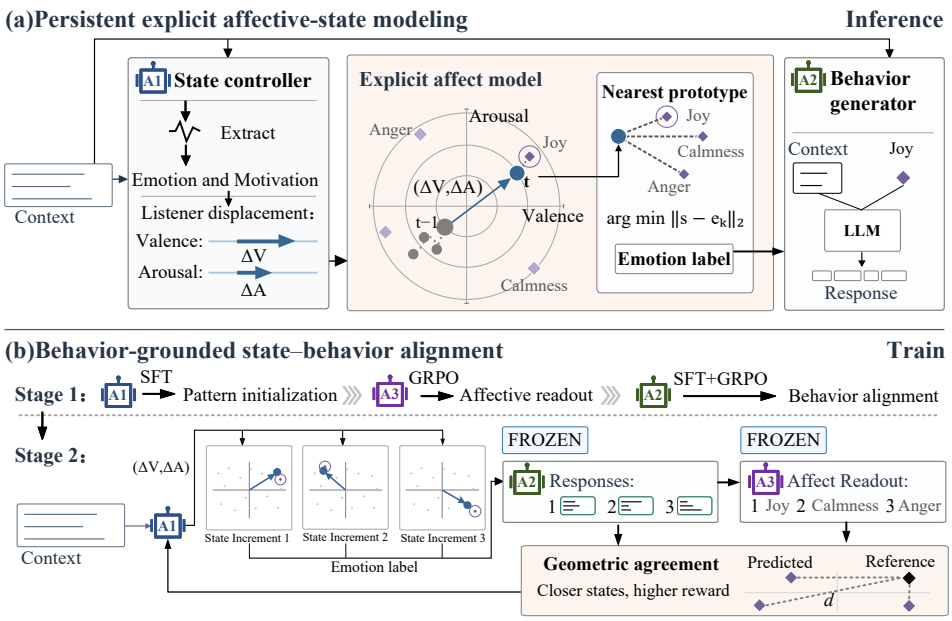

Figure 2: Overview of UEIM. (a): A1 updates a persistent V–A state and passes its semantic prototype to A2 for behavior realization. (b): A1–A3 are prepared for transition inference, behavior realization, and affective readout before behavior-grounded feedback calibrates A1.

behavior-grounded calibration freezes π_ϕ and h_η and optimizes only f_θ with GRPO (Shao et al., 2024) using behavior-mediated affective feedback. The fixed state-to-behavior pathway allows candidate transitions to be compared through their observable consequences. Training and reward details are provided in Appendix A.

4 METHOD

Figure 2 presents **Unified Explicit Interoception Modeling (UEIM)**. Its design follows two key requirements of persistent affect modeling. First, affect should evolve from its previous condition rather than be reconstructed independently at each turn, consistent with evidence of temporal dependence in affect (Russell, 1980; Kuppens et al., 2010). This makes recurrence necessary: otherwise, information accumulated across interactions must be repeatedly reconstructed from context. Second, the maintained state should explicitly influence observable behavior without requiring unavailable continuous-state annotations. UEIM therefore separates *state maintenance* from *behavior realization* and introduces a behavioral feedback pathway for calibrating the latent state. The *state controller* A1 (f_θ) infers how the current interaction changes the character's previous affective condition; the *behavior generator* A2 (π_ϕ) realizes the maintained state as a contextually appropriate response; and the *affective reader* A3 (h_η) maps observable behavior back into the shared affective space. At inference, only A1 and A2 are required.

4.1 PERSISTENT AFFECTIVE-STATE MODELING

A1: maintaining what persists. Turn-level emotion prediction describes what is expressed at a particular moment but does not explicitly preserve how prior interactions shape the character's current affective condition. UEIM instead maintains a recurrent state $\mathbf{s}_t = (v_t, a_t) \in [-1, 1]^2$ in Valence–Arousal (V–A) space. The Circumplex Model of Affect provides a compact geometry for affective direction and relative intensity (Russell, 1980), while affect-dynamics studies motivate conditioning the current state on its previous condition (Kuppens et al., 2010). We use V–A as a structured affective prior rather than a psychologically literal internal state.

To combine continuous transition with semantic interpretability, we define a set $\mathcal{K}$ of 15 fine-grained emotion prototypes, each associated with a fixed coordinate $\mathbf{e}_k = (v_k, a_k)$. This hybrid representa-

tion is deliberate: continuous coordinates preserve incremental change and accumulation, whereas prototypes provide a stable semantic interface for generation. Given interaction history H_t, current event u_t, previous state $\mathbf{s}_{t-1}$, and task condition g, A1 predicts an event-conditioned displacement:

$$\begin{aligned} \widehat{\Delta\mathbf{s}}_t &= f_\theta(H_t, u_t, \mathbf{s}_{t-1}, g), & \widehat{\mathbf{s}}_t &= \Pi_{[-1,1]^2}\left(\mathbf{s}_{t-1} + \widehat{\Delta\mathbf{s}}_t\right), \\ \widehat{k}_t &= \mathcal{P}(\widehat{\mathbf{s}}_t) := \arg\min_{k\in\mathcal{K}} \|\widehat{\mathbf{s}}_t - \mathbf{e}_k\|_2 . \end{aligned} \tag{2}$$

Here f_θ includes component-wise increment clipping to $[-0.8, 0.8]$. The nearest-prototype projection is shown in Figure 2; implementation details appear in Appendix G. The continuous $\widehat{\mathbf{s}}_t$ preserves accumulation and graded change, while $\widehat{k}_t$ provides a semantic interpretation. A1 never observes reference responses or response-side annotations at inference.

A1 also performs structured event interpretation before predicting the displacement. As shown in Figure 2, it distinguishes affect expressed by the incoming speaker from the change induced in the target character and organizes evidence such as affective cause, interaction motivation, and intensity. This distinction matters because the speaker's expressed emotion need not equal the affective change induced in the character. Only the listener-side V–A displacement enters the recurrent update; the remaining variables support transition inference. Supervision and prompt interfaces are detailed in Appendices B and D.

A2: realizing the maintained state as behavior. A2 addresses a separate question: once an affective condition is maintained, how should it be expressed in the current interaction? Existing character-simulation and role-playing systems typically generate behavior jointly from profiles, memories, and dialogue context (Liu et al., 2024; Wang et al., 2025). UEIM instead uses an explicit state-to-behavior interface, allowing the consequence of a candidate state to be evaluated independently of A1's internal reasoning. This separation is important for calibration: if state estimation and behavior realization were entangled, an improved response would not reveal whether the state transition itself had improved.

A2 does not receive A1's free-form reasoning or raw V–A coordinates. The updated state is projected to $\widehat{k}_t$ and accompanied by its predefined intensity descriptor $\rho(\widehat{k}_t)$ and semantic description $\psi(\widehat{k}_t)$:

$$r_t \sim \pi_\phi\left(\cdot \mid H_t, u_t, \widehat{k}_t, \rho(\widehat{k}_t), \psi(\widehat{k}_t)\right). \tag{3}$$

Thus, A1 determines *what affective condition is carried forward and how it changes*, whereas A2 determines *how that condition is expressed in language*. The semantic prototype forms the controlled interface between recurrent state maintenance and behavior realization.

4.2 Behavior-Grounded State–Behavior Alignment

The central learning difficulty is that dialogue corpora provide observable responses and sometimes categorical emotion labels, but rarely provide turn-level continuous V–A trajectories. Direct regression to teacher-produced coordinates would make the learned dynamics depend heavily on the teacher's numerical calibration. UEIM therefore uses two phases: *capability preparation* equips A1–A3 with transition inference, state-conditioned generation, and affective readout, after which *behavior-grounded calibration* optimizes A1 from the consequences of its transitions.

Capability preparation. A1 is initialized from teacher-derived transition examples. Coarse emotion annotations constrain affective direction, while an offline teacher provides intensity information for constructing bounded V–A displacement targets. These trajectories provide a geometric initialization for interpreting interaction events as state changes rather than supervision of a uniquely correct continuous state. A2 is initialized to generate contextually appropriate responses conditioned on semantic affective prototypes. Supervision construction is detailed in Appendix B.

A3 provides the behavioral readout needed to compare candidate transitions. Motivated by work that evaluates emotional fidelity and affective trajectories through observable dialogue behavior (Feng et al., 2025; Tan et al., 2026), A3 maps a response into the shared prototype space:

$$\widetilde{k} = h_\eta(r) \in \mathcal{K}, \qquad S_{\mathrm{geo}}(k_i, k_j) = \mathrm{sim}\left(\mathbf{e}_{k_i}, \mathbf{e}_{k_j}\right). \tag{4}$$

Here $h_\eta(r)$ abbreviates the context-conditioned readout defined in Appendix D, and sim is the normalized prototype-distance reward in Appendix A.1. Because prototypes retain coordinates in V–A space, the score preserves graded relations between nearby affective conditions. A3 is therefore a task-specific behavioral readout rather than an estimator of the character's psychological state.

A2 is then optimized so that the supplied affective condition visibly influences generation while preserving response quality. Using GRPO (Shao et al., 2024), candidate response r_i receives

$$R_i^{\text{behavior}} = \lambda\, S_{\text{geo}}\left(\widetilde{k}_i, k_t^*\right) + (1-\lambda)\, J_{\text{GPT4o}}(r_i), \tag{5}$$

where $\widetilde{k}_i$ is A3's readout, k_t^* is the refined prototype associated with the reference response, and J_{GPT4o} evaluates general response quality. This stage prepares A2 as a responsive and approximately fixed state-to-behavior mapping; it does not optimize the recurrent state.

Behavior-grounded calibration of A1. After capability preparation, A2 and A3 are frozen and only A1 is optimized. Fixing behavior realization and affective readout provides a comparatively stationary pathway for evaluating state transitions and reduces cross-module non-stationarity. It also isolates the source of improvement: during this stage, better rewards can only arise from changes in A1's transition policy rather than adaptation in the downstream modules.

For each candidate displacement $\widehat{\Delta\mathbf{s}}_{t,i}$, A1 produces a candidate state and semantic prototype; frozen A2 realizes that condition as behavior, and frozen A3 maps the response back into the shared affective space. Using the same reference-response target k_t^*, the closed-loop computation is

$$\begin{aligned}
\widehat{\mathbf{s}}_{t,i} &= \Pi_{[-1,1]^2}\left(\mathbf{s}_{t-1} + \widehat{\Delta\mathbf{s}}_{t,i}\right), \quad \widehat{k}_{t,i} = \mathcal{P}(\widehat{\mathbf{s}}_{t,i}), \\
r_{t,i} &\sim \pi_\phi\left(\cdot \mid H_t, u_t, \widehat{k}_{t,i}\right), \qquad \widetilde{k}_{t,i} = h_\eta(r_{t,i}), \\
R_i^{\text{state}} &= S_{\text{geo}}\left(\widetilde{k}_{t,i}, k_t^*\right).
\end{aligned} \tag{6}$$

A candidate transition is preferred when the behavior it induces is affectively compatible with the observed character response. Unlike direct regression to teacher-produced coordinates, this feedback depends on the consequence of the predicted state rather than its exact numerical location.

Because the continuous state is projected to a semantic prototype before generation, the closed loop does not uniquely identify exact V–A coordinates. We therefore interpret the final stage as calibrating *behaviorally meaningful recurrent state transitions*: teacher-derived trajectories provide an initial geometric scaffold, while observable downstream behavioral consequences indicate which transitions produce affectively compatible and temporally coherent behavior. The continuous state preserves accumulation and graded temporal change before semantic projection, whereas the behavioral loop calibrates which recurrent transitions are useful for stable character-consistent realization. Additional optimization details are provided in Appendix A.1.

5 Experiments

5.1 Experimental Setup

Datasets. We evaluate UEIM on two multi-turn dialogue datasets. **(1) DailyDialog.** DailyDialog (Li et al., 2017) contains 10,076 instances, split into 7,122/1,366/1,588 train/validation/test instances under the official dialogue partitions, with response-side emotion labels refined into 15 affective prototypes. **(2) ESConv.** ESConv (Liu et al., 2021) provides 2,132 evaluation instances and is used without target-domain training. Prototype targets are assigned to human reference responses by frozen A3; thus, affective metrics measure agreement within the learned prototype space rather than against human continuous-state annotations. Data construction is detailed in Appendix B.

Baselines. All local systems use Qwen2.5-7B-Instruct (Yang et al., 2024) to control for backbone family and scale. Qwen Direct generates directly from dialogue context, while Qwen Two-Agent uses prompted controller–generator orchestration. ECoT-Prompt (**Adapted**) applies emotion-aware reasoning (Li et al., 2024), and MultiAgentESC (**Adapted**) applies multi-role

Table 2: Reference-response agreement (mean ± standard deviation over five independent runs). Metrics are scaled by 100; higher values indicate better agreement. Bold marks the best shared-backbone result. GPT-5.4 is included solely as an external baseline.

	DailyDialog ($N = 1{,}588$)				**ESConv** ($N = 2{,}132$)			
Model	**R-L**	**BERT**	**E-Acc**	**E-MF1**	**R-L**	**BERT**	**E-Acc**	**E-MF1**
Qwen Direct	13.72 ± 0.14	83.60 ± 0.06	90.62 ± 0.19	43.10 ± 0.33	12.49 ± 0.24	84.81 ± 0.05	91.81 ± 0.17	65.50 ± 0.41
Qwen Two-Agent	14.75 ± 0.18	84.37 ± 0.12	91.37 ± 0.41	44.64 ± 0.58	13.98 ± 0.15	84.21 ± 0.09	92.64 ± 0.30	73.22 ± 1.22
ECoT-Prompt (**Adapted**)	15.84 ± 0.27	84.61 ± 0.10	91.62 ± 0.18	44.24 ± 0.84	14.57 ± 0.32	84.58 ± 0.26	92.31 ± 0.20	66.83 ± 0.56
MultiAgentESC (**Adapted**)	13.49 ± 0.29	83.91 ± 0.08	90.68 ± 0.24	44.32 ± 0.75	12.94 ± 0.28	84.30 ± 0.09	92.64 ± 0.25	70.33 ± 1.10
Explicit-State (**Controlled**)	14.05 ± 0.19	84.12 ± 0.04	90.99 ± 0.42	43.38 ± 0.72	13.44 ± 0.21	84.19 ± 0.14	93.15 ± 0.21	73.05 ± 0.77
UEIM (Ours)	**17.48** ± 0.13	**85.14** ± 0.08	**92.50** ± 0.26	**45.20** ± 0.57	**16.62** ± 0.22	**85.06** ± 0.16	**94.23** ± 0.15	**73.81** ± 0.95
GPT-5.4 API	16.45 ± 0.12	84.72 ± 0.13	93.88 ± 0.34	46.33 ± 0.55	11.86 ± 0.11	86.12 ± 0.08	93.56 ± 0.16	73.96 ± 0.62

Table 3: DailyDialog pairwise preference from UEIM's perspective. GPT-5.4 counts are rounded five-run means. Win rates include ties in the denominator.

	GPT-5.4 judge ($N = 1{,}588$)				**Human evaluation** ($N = 50$)			
Ours vs. Comparison	Wins	Ties	Losses	Win rate (%)	Wins	Ties	Losses	Win rate (%)
Ours **vs**. Qwen Direct	**860**	66	662	54.16	**28**	3	19	56.00
Ours **vs**. Qwen Two-Agent	**942**	44	602	59.32	**28**	1	21	56.00
Ours **vs**. ECoT-Prompt (**Adapted**)	**859**	45	684	54.09	**26**	2	22	52.00
Ours **vs**. MultiAgentESC (**Adapted**)	**1326**	10	252	83.50	**42**	2	6	84.00
Ours **vs**. Explicit-State (**Controlled**)	**857**	14	717	53.97	**25**	2	23	50.00
Ours **vs**. *GPT-5.4 API*	319	33	1236	20.09	11	3	36	22.00

emotional-support prompting (Xu et al., 2025); both are shared-backbone adaptations rather than official reproductions. EXPLICIT-STATE (**CONTROLLED**) retains the recurrent V–A state and semantic projection but omits behavior-grounded calibration, isolating the contribution of the proposed feedback mechanism. GPT-5.4 API is reported separately as a proprietary external reference.

Evaluation metrics. Reference-response and affective agreement. Table 2 measures lexical overlap with ROUGE-L (R-L) (Lin, 2004), semantic similarity with BERTScore-F1 (BERT) (Zhang et al., 2020), and prototype agreement with emotion accuracy (E-Acc) and macro-F1 (E-MF1). **Anonymous pairwise response preference.** Table 3 compares identity-masked DailyDialog replies. GPT-5.4 scores both presentation orders; higher scores win and equal scores tie. Human evaluation covers 50 pairs per comparator. Win rate is $100\,\mathrm{Wins}/(\mathrm{Wins} + \mathrm{Ties} + \mathrm{Losses})$, including ties in the denominator. Table 2 and stage-wise scores average five runs; trained agents use different seeds and full fine-tuning. Table 3 recomputes win rates from the displayed counts. Metric definitions appear in Appendix I. Additional evaluation details are provided in Appendices C and E.

5.2 OVERALL PERFORMANCE

Response quality and affective agreement. Table 2 compares UEIM with shared-backbone and external baselines in response realization and affective agreement across both datasets. **(1) Response realization.** UEIM achieves the strongest ROUGE-L and BERTScore among shared-backbone systems on both datasets. Its gains over QWEN DIRECT and QWEN TWO-AGENT indicate that the improvement is not explained merely by additional prompting or multi-agent orchestration. More importantly, UEIM also outperforms EXPLICIT-STATE (CONTROLLED), suggesting that recurrent state maintenance alone is insufficient and that behavior-grounded calibration helps translate the maintained state into more appropriate responses. **(2) Affective agreement.** UEIM likewise achieves the best shared-backbone E-Acc and E-MF1 on both datasets. Since no ESConv data are used for training, its improvement on ESConv suggests that the learned state-to-behavior interface preserves affective structure beyond the source dialogue domain. We interpret this as evidence of cross-domain dialogue transfer rather than unrestricted task generalization.

Behavioral preference. Table 3 reports pairwise preferences under GPT-5.4 and human evaluation, providing a complementary assessment beyond reference-based metrics. **(3) Behavioral preference.** UEIM records more wins than losses against every local comparator under both evaluators,

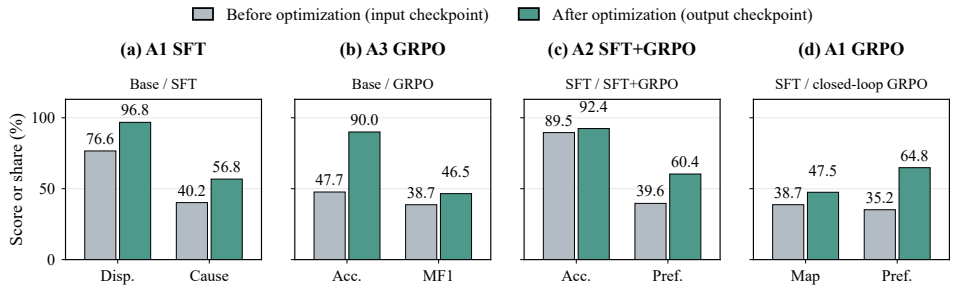

Figure 3: Stage-wise analysis of UEIM. Each panel compares checkpoints before and after one optimization stage using capability-aligned metrics; higher is better.

suggesting that its gains extend beyond lexical or affective agreement to perceived behavioral appropriateness. The broadly consistent trends across GPT-5.4 and human judgments provide additional support for this interpretation. Meanwhile, GPT-5.4 API remains stronger in direct preference despite weaker performance on some reference-based metrics, indicating that the two evaluation paradigms capture complementary aspects of response quality.

5.3 Component and Training-Stage Analysis

Setup. Figure 3 examines whether each stage in Section 4.2 contributes its intended capability rather than only improving the final model jointly. Each panel compares checkpoints immediately before and after one stage using component-aligned metrics. **(a) A1 SFT** evaluates affective increment prediction (Disp.) and cause extraction (Cause), testing whether the controller infers both how the state changes and which cues trigger the change. **(b) A3 GRPO** evaluates prototype classification with accuracy (Acc.) and macro-F1 (MF1), measuring the reliability and balance of the affective readout. **(c) A2 SFT+GRPO** measures response-prototype agreement (Acc.) and checkpoint preference (Pref.), testing whether stronger affective realization also yields better responses. **(d) A1 GRPO** evaluates state-to-prototype mapping (Map) and downstream preference with A2/A3 frozen, isolating controller calibration from generator or reader adaptation. Results are five-run means on a 0–100 scale; Pref. excludes ties.

Results. Figure 3 shows consistent gains across all four stages, matching their intended roles. **(a) A1 SFT** improves both Disp. and Cause, indicating that supervised initialization helps the controller learn the direction of affective change and the evidence driving it. This provides a structured starting point for recurrent updates rather than leaving later optimization to discover state dynamics from scratch. **(b) A3 GRPO** improves Acc. and MF1, yielding a more accurate and balanced mapping from behavior to the shared affective space, which strengthens the fixed feedback signal for closed-loop optimization. **(c) A2 GRPO** improves response-prototype agreement and receives 60.35% of non-tied preferences over SFT alone, suggesting that stronger affective realization also improves perceived response quality. **(d) A1 GRPO** further improves Map and receives 64.82% of non-tied preferences. Since A2/A3 remain frozen, these gains more directly reflect recurrent-controller calibration. Overall, the results support the intended progression from supervised component initialization to behavior-grounded recurrent-state calibration.

5.4 Transferring Frozen A1 to Downstream Tasks

Setup. To test whether A1 learns reusable affective information beyond the UEIM training loop, we transfer the frozen controller to three downstream settings. Qwen2.5-7B-Instruct, prompts, and decoding are fixed; the only difference is whether the model receives A1-generated cues (emotion, intensity, cause, and motivation), while A2/A3 are unused. **(a) Negotiation:** 120 CraigslistBargain cases (He et al., 2018); success rate measures valid deals, utterance count reflects interaction cost, and sale/listing ratio measures retained price. **(b) Counseling:** 150 AnnoMI transitions (Wu et al., 2022); dynamics attunement measures sensitivity to affective change, empathy captures conveyed understanding, and a MITI-inspired composite summarizes broader counseling behaviors. **(c) Service:** 150 EmoWOZ turns (Feng et al., 2022); dynamics, task effectiveness, and service quality assess affective fit, goal advancement, and communication quality. For (b,c), GPT-5.4 averages both

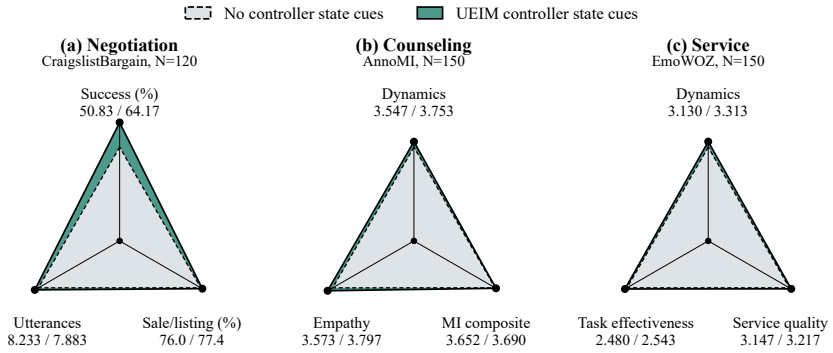

Figure 4: Three-axis downstream comparisons. Baseline radii equal 1; state-cue radii show proportional improvement (inverted for utterances). Labels report baseline/state-cue means. Larger radii are better; area is not a composite score.

presentation orders on 1–5 scales. Higher is better except utterance count; full protocols are in Appendices H and I.

Results. Figure 4 shows that frozen A1 cues remain useful after being detached from A2/A3. **(a) Negotiation:** state cues increase valid deals while reducing dialogue length and improving the sale/listing ratio, suggesting more efficient agreement without sacrificing price retention. **(b) Counseling:** gains concentrate on dynamics attunement and empathy, while the broader composite changes less. This is consistent with A1 supplying affective-state and cause information rather than the full set of counseling techniques. **(c) Service:** dynamics improves most, with smaller gains in task effectiveness and service quality, indicating that A1 primarily supports adaptation to evolving user state rather than domain-specific task reasoning. Overall, the strongest transfer gains appear on dimensions most directly related to affective state awareness.

5.5 Effect of Retaining Affective State

Setup. We test whether the recurrent state retains information beyond a short local context. We select one late turn from each of 300 ESConv dialogues (Liu et al., 2021), stratified by history depth (16–23, 24–31, and 32+ utterances). Both conditions use the same frozen A1/A2 and last four utterances. The *persistent* condition retains the accumulated state, whereas *reset* reinitializes from $(0, 0)$ and reconstructs state from the local window. GPT-5.4 evaluates anonymous responses under randomized and reversed orders; paired sign tests use Holm correction. Full-history and shuffled-state variants serve as additional controls.

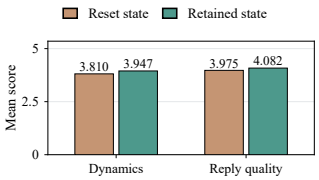

Figure 5: Retained versus reset state; mean scores range from 1 to 5.

Results. Figure 5 shows that retaining the accumulated state improves dynamics attunement by 0.137 (95% CI $[.037, .237]$; Holm-adjusted $p = .0193$), with a smaller 0.107 gain in reply quality ($p = .0981$). This suggests that persistent state preserves affective information beyond the recent four-turn context. Full-history input performs best, positioning recurrent state as a compact summary rather than a substitute for complete context. The matched-versus-shuffled comparison is inconclusive, providing stronger evidence for retained history than fine-grained state specificity. Appendix F evaluates persistence across dialogue depths.

6 Conclusion

UEIM models persistent affect as a recurrent character state separated from current-turn behavior realization. By freezing the state-to-behavior pathway and calibrating transitions through their observable consequences, UEIM learns behaviorally meaningful affective dynamics without treating continuous state annotations as ground truth. Experiments on response realization, staged optimization, downstream transfer, and state retention support the utility of this formulation.

Reproducibility Statement

Implementation details and experimental protocols are provided in Sections 4.2–5.5 and the appendix. Training objectives, prompts, evaluation criteria, dataset preprocessing, downstream-task protocols, statistical tests, and additional analyses are documented in the corresponding sections and appendices.

AI Use Statement

Generative AI tools assisted language editing, manuscript organization, and plotting-code preparation. The authors are responsible for the experimental evidence, scientific claims, and final manuscript.

References

Zhuang Chen, Yaru Cao, Guanqun Bi, Jincenzi Wu, Jinfeng Zhou, Xiyao Xiao, Si Chen, Hongning Wang, and Minlie Huang. SocialSim: Towards socialized simulation of emotional support conversation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(2):1274–1282, 2025. doi: 10.1609/aaai.v39i2.32116. URL https://ojs.aaai.org/index.php/AAAI/article/view/32116.

Qiming Feng, Qiujie Xie, Xiaolong Wang, Qingqiu Li, Yuejie Zhang, Rui Feng, Tao Zhang, and Shang Gao. EmoCharacter: Evaluating the emotional fidelity of role-playing agents in dialogues. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 6218–6240, Albuquerque, New Mexico, 2025. Association for Computational Linguistics. doi: 10.18653/v1/2025.naacl-long.316. URL https://aclanthology.org/2025.naacl-long.316/.

Shutong Feng, Nurul Lubis, Christian Geishauser, Hsien-chin Lin, Michael Heck, Carel van Niekerk, and Milica Gašić. EmoWOZ: A large-scale corpus and labelling scheme for emotion recognition in task-oriented dialogue systems. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 4096–4113, Marseille, France, 2022. European Language Resources Association. URL https://aclanthology.org/2022.lrec-1.436/.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2333–2343, Brussels, Belgium, 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1256. URL https://aclanthology.org/D18-1256/.

Peter Kuppens, Zita Oravecz, and Francis Tuerlinckx. Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, 99 (6):1042–1060, 2010. doi: 10.1037/a0020962.

Jing Li, Yanxin Luo, Donghong Han, Yimeng Zhan, Xiaoming Fu, Baiyou Qiao, and Gang Wu. ESCA: An emotional support conversation agent for enhancing reasonable strategy planning and effective expression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40 (21):17526–17534, 2026. doi: 10.1609/aaai.v40i21.38807. URL https://ojs.aaai.org/index.php/AAAI/article/view/38807.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. DailyDialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 986–995, Taipei, Taiwan, 2017. Asian Federation of Natural Language Processing. URL https://aclanthology.org/I17-1099/.

Zaijing Li, Gongwei Chen, Rui Shao, Yuquan Xie, Dongmei Jiang, and Liqiang Nie. Enhancing emotional generation capability of large language models via emotional chain-of-thought. *arXiv preprint arXiv:2401.06836*, 2024. URL https://arxiv.org/abs/2401.06836.

Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81. Association for Computational Linguistics, 2004. URL https://aclanthology.org/W04-1013/.

Jiaheng Liu, Zehao Ni, Haoran Que, Tao Sun, Zekun Wang, Jian Yang, Jiakai Wang, Hongcheng Guo, Zhongyuan Peng, Ge Zhang, Jiayi Tian, Xingyuan Bu, Ke Xu, Wenge Rong, Junran Peng, and Zhaoxiang Zhang. Roleagent: Building, interacting, and benchmarking high-quality role-playing agents from scripts. In *Advances in Neural Information Processing Systems*, volume 37, pp. 49403–49428. Neural Information Processing Systems Foundation, 2024. doi: 10.52202/079017-1563. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/5875aca1ef70285a35940afbbce0f9fb-Abstract-Datasets_and_Benchmarks_Track.html.

Jiaqi Liu, Yankun Yang, Jiakang Xu, Zhongqiang Du, and Wenbin Jiang. Dynamic cognitive planning for cognitive-functional dialogue: A case study in emotional support conversation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(38):32069–32077, 2026. doi: 10.1609/aaai.v40i38.40478. URL https://ojs.aaai.org/index.php/AAAI/article/view/40478.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. Towards emotional support dialog systems. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 3469–3483, Online, 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.269. URL https://aclanthology.org/2021.acl-long.269/.

Keming Lu, Bowen Yu, Chang Zhou, and Jingren Zhou. Large language models are superpositions of all characters: Attaining arbitrary role-play via self-alignment. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 7828–7840, Bangkok, Thailand, 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.423. URL https://aclanthology.org/2024.acl-long.423/.

Theresa B. Moyers, Lauren N. Rowell, Jennifer K. Manuel, Denise Ernst, and Jon M. Houck. The motivational interviewing treatment integrity code (MITI 4): Rationale, preliminary reliability and validity. *Journal of Substance Abuse Treatment*, 65:36–42, 2016. doi: 10.1016/j.jsat.2016.01.001.

James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39 (6):1161–1178, 1980. doi: 10.1037/h0077714.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Zhouxing Tan, Ruochong Xiong, Yulong Wan, Jinlong Ma, Hanlin Xue, Qichun Deng, Haifeng Jing, Zhengtong Zhang, Depei Liu, Shiyuan Luo, and Junfei Liu. Detecting emotional dynamic trajectories: An evaluation framework for emotional support in language models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(3):2074–2082, 2026. doi: 10.1609/aaai.v40i3.37189. URL https://ojs.aaai.org/index.php/AAAI/article/view/37189.

Chenwei Wan, Matthieu Labeau, and Chloé Clavel. EmoDynamiX: Emotional support dialogue strategy prediction by modelling MiXed emotions and discourse dynamics. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 1678–1695, Albuquerque, New Mexico, 2025. Association for Computational Linguistics. doi: 10.18653/v1/2025.naacl-long.81. URL https://aclanthology.org/2025.naacl-long.81/.

Ruiyi Wang, Haofei Yu, Wenxin Zhang, Zhengyang Qi, Maarten Sap, Yonatan Bisk, Graham Neubig, and Hao Zhu. SOTOPIA-π: Interactive learning of socially intelligent language agents. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*

(Volume 1: Long Papers), pp. 12912–12940, Bangkok, Thailand, 2024a. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.698. URL https://aclanthology.org/2024.acl-long.698/.

Xintao Wang, Yunze Xiao, Jen-tse Huang, Siyu Yuan, Rui Xu, Haoran Guo, Quan Tu, Yaying Fei, Ziang Leng, Wei Wang, Jiangjie Chen, Cheng Li, and Yanghua Xiao. Incharacter: Evaluating personality fidelity in role-playing agents through psychological interviews. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1840–1873, Bangkok, Thailand, 2024b. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.102. URL https://aclanthology.org/2024.acl-long.102/.

Xintao Wang, Heng Wang, Yifei Zhang, Xinfeng Yuan, Rui Xu, Jen-Tse Huang, Siyu Yuan, Haoran Guo, Jiangjie Chen, Shuchang Zhou, Wei Wang, and Yanghua Xiao. CoSER: Coordinating LLM-based persona simulation of established roles. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 64822–64858. PMLR, 2025. URL https://proceedings.mlr.press/v267/wang25dk.html.

Zixiu Wu, Simone Balloccu, Vivek Kumar, Rim Helaoui, Ehud Reiter, Diego Reforgiato Recupero, and Daniele Riboni. Anno-MI: A dataset of expert-annotated counselling dialogues. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6177–6181. IEEE, 2022. doi: 10.1109/ICASSP43922.2022.9746035.

Yangyang Xu, Jinpeng Hu, Zhuoer Zhao, Zhangling Duan, Xiao Sun, and Xun Yang. MultiAgentESC: A LLM-based multi-agent collaboration framework for emotional support conversation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 4665–4681. Association for Computational Linguistics, 2025. doi: 10.18653/v1/2025.emnlp-main.232. URL https://aclanthology.org/2025.emnlp-main.232/.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2024. doi: 10.48550/arXiv.2412.15115.

Yeyong Yu, Runsheng Yu, Haojie Wei, Zhanqiu Zhang, and Quan Qian. Beyond dialogue: A profile-dialogue alignment framework towards general role-playing language model. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 11992–12022, Vienna, Austria, 2025. Association for Computational Linguistics. doi: 10.18653/v1/2025.acl-long.586. URL https://aclanthology.org/2025.acl-long.586/.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. BERTScore: Evaluating text generation with BERT. In *International Conference on Learning Representations*, 2020. URL https://arxiv.org/abs/1904.09675.

Yi Zheng, Feier Qin, Xiao Li, Haibin Huang, Hanyao Wang, Xiaoyu Wang, Yan Lu, and Yuan Zhang. Toward natural and companionable virtual agents via cross-temporal emotional modeling. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, pp. 1–21. Association for Computing Machinery, 2026. doi: 10.1145/3772318.3790917.

Xuhui Zhou, Hao Zhu, Leena Mathur, Ruohong Zhang, Haofei Yu, Zhengyang Qi, Louis-Philippe Morency, Yonatan Bisk, Daniel Fried, Graham Neubig, and Maarten Sap. SOTOPIA: Interactive evaluation for social intelligence in language agents. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/b3075b88e583a0e98d8b24338a613060-Abstract-Conference.html.

A Training and Reward Implementation

Training runtime. Training uses verl with its default configurations on hardware with 64 GB GPU memory. Only the active policy is updated at each stage. Scalar outcome rewards are assigned to the last valid response token; verl computes group-relative advantages and the configured policy/KL losses. Frozen service calls supply reward values without backpropagating through A2 or A3. Final A1 calibration uses temperature 0 for these calls, with output limits of 96 tokens for A2 and 24 for A3.

A.1 Reward Computation and Validity Checks

The geometric agreement function in Eq. 4 uses a fixed normalization over prototype coordinates:

$$S_{\text{geo}}(k_i, k_j) = \max\left(0, 1 - \frac{\|\mathbf{e}_{k_i} - \mathbf{e}_{k_j}\|_2}{D_{\max}}\right), \qquad D_{\max} = \max_{k,l \in \mathcal{K}} \|\mathbf{e}_k - \mathbf{e}_l\|_2. \tag{7}$$

This maps prototype distance to bounded similarity while retaining graded relations among nearby conditions. Unrecognized A3 labels receive zero reward and distance $D_{\max}$. The parser accepts recognized IDs or names without the requested XML-like tag; label validity and tag compliance are distinct checks. For A2 alignment, $\lambda = 0.2$ in Eq. 5, with weight 0.8 on response quality. Appendix D specifies the rubric.

During final A1 calibration, malformed policy JSON receives zero reward. Service/API failures instead abort reward computation; unavailable judgments are not scored as valid zero-reward outcomes. Equation 6 compares the frozen A3 readout against the dataset target k_t^*, not A1's own selected prototype. No additional quality judge, state-jump penalty, or hand-crafted regularizer is used.

B Supervision and Data Construction

Transition targets for A1 SFT. The coarse emotion annotation constrains affective direction, while an offline teacher estimates intensity to construct an initialization coordinate $\mathbf{q}_t$. The displacement target is

$$\Delta \mathbf{s}_t^{\text{init}} = \text{clip}(\alpha(\mathbf{q}_t - \mathbf{s}_{t-1}), -\delta_{\text{init}}, \delta_{\text{init}}), \qquad \alpha = 0.4, \quad \delta_{\text{init}} = 0.5. \tag{8}$$

Clipping is applied component-wise. These targets provide a geometric initialization for learning event-conditioned transitions rather than supervision of a uniquely correct continuous state. The constants are used only during supervision construction and introduce no additional inference-time update rule; inference follows Eq. 2.

Response-side prototype targets. For a human reference response r_t^* with original emotion label y_t, GPT-4o selects k_t^* from the semantically compatible subset $\mathcal{C}(y_t) \subseteq \mathcal{K}$ using the dialogue context and reference response. Neutral responses map deterministically to E00. This response-side target is constructed independently of the continuous transition target and is shared by A3 training, A2 alignment, and final A1 calibration. Neither the reference response nor k_t^* is available to A1 at inference.

Aligned training views. Each DailyDialog unit contains the preceding dialogue context, current user utterance, next agent utterance, original emotion label, and dialogue act. A1 SFT uses the structured transition target, while A2 SFT uses the human reference response as the generation target. A3 GRPO takes the response as input and predicts its affective prototype. During A2 GRPO and final A1 calibration, the reference response is withheld from the generating policy, while its refined prototype k_t^* is used only as the reward target. All views preserve the original dialogue split identity.

Neutral-class subsampling. At most 1,000 neutral instances are retained in each split, while all eligible non-neutral instances are preserved. No dialogue is moved across the official partitions. Final split sizes are reported in Section 5.1. Teacher JSON, allowed-label membership, parent constraints,

and sample-key uniqueness are validated. Confidence and rationale are provenance metadata, not sample weights or confidence filters. Records without a recognized label or nonempty reference are excluded; the materialized agent views retain shared split, dialogue, turn, and sample keys.

C Additional Evaluation Details

Class distribution and state-map diagnostics. The DailyDialog test set contains 1,000 neutral (E00) and 330 joy (E08) instances, while fatigue (E11) is absent. This imbalance motivates reporting macro-F1 alongside accuracy. In Figure 3, Map measures agreement between A1's projected prototype and the response-side target. Additional diagnostics show that A1's intensity score increases from 0.7260 to 0.8424, A3's V–A distance decreases from 0.2158 to 0.0539, and the final A1 state-map distance decreases from 0.2410 to 0.2103. These diagnostics separately track transition quality, affective readout, and final state calibration.

Sources of variation. For prompted and proprietary baselines, repeated runs reflect variation in generation or evaluation rather than parameter training. Run-level standard deviations summarize repetition-level variation, whereas downstream confidence intervals are computed over paired scenario samples. Table 3 reports win rates calculated from the displayed aggregate counts, including ties in the denominator.

DailyDialog preference rubric. The GPT-5.4 rubric for Table 3 evaluates expressed emotion, intensity, affective cause, temporal change, and interpersonal intent. On the 1–5 scale, 5 denotes fully appropriate affective expression, 3 a generic but conversation-preserving reply, and 1 a severe mismatch that disrupts interaction. Verbosity, politeness, advice, emotion words, and stylistic complexity are not rewarded by themselves. The judge sees dialogue context and anonymous responses, without references or internal states. Each seeded candidate permutation is paired with its reverse; both valid scores are required, identities are restored, and scores are averaged within each sample before computing wins, ties, or losses.

D Prompt Interfaces and Scoring Instructions

Braces denote dynamic fields to be substituted before use; the A1 JSON block specifies the output schema. Supervision construction is given in Appendix B.

D.1 Agent prompts

A1: affect interpretation and state update

SFT and final A1 calibration share the JSON schema, but not identical prompt strings. The frozen-loop prompt below serializes the previous continuous V–A state as seven nonnegative emotion percentages; these projections need not sum to 100. SFT orders them as happiness, sadness, anger, fear, disgust, surprise, neutral and additionally supplies `previous_dominant_inner_emotion`. The continuous coordinate, rather than this textual representation, is retained for numerical updates.

```
Given the dialogue history, previous inner emotional state percentages,
and current user utterance, infer the user's external emotion, cause and
motivation, then decide the change in the Agent's V-A coordinates.

[Dialogue History]
{DIALOGUE_HISTORY}
[Previous Inner State Percent]
neutral={N}, anger={A}, disgust={D}, fear={F},
happiness={H}, sadness={S}, surprise={U}
[Current User Utterance]
{CURRENT_USER_UTTERANCE}
[Dialogue Act]
{DIALOGUE_ACT}

Return strict JSON only. "thinking" contains external_emotion,
emotion_cause, motivation_analysis, and emotion_intensity. "decision"
contains delta_valence and delta_arousal. Valid external labels: neutral,
anger, disgust, fear, happiness, sadness, surprise.
```

```
{
  "thinking": {
    "external_emotion": "{LABEL}",
    "emotion_cause": "{CAUSE}",
    "motivation_analysis": "{MOTIVATION}",
    "emotion_intensity": {VALUE_IN_[0,1]}
  },
  "decision": {
    "delta_valence": {DELTA_V},
    "delta_arousal": {DELTA_A}
  }
}
```

A2: STATE-CONDITIONED RESPONSE GENERATION

```
You are Agent2 in a multi-agent emotional companion system. Generate the
next natural Agent reply from the conversation and latest interoception
label.

[Dialogue History]
{DIALOGUE_HISTORY}
[Current User Utterance]
{CURRENT_USER_UTTERANCE}
[Latest Agent Interoception Label]
{LABEL_ID} | {LABEL_NAME_ENGLISH} | {LABEL_NAME_CHINESE}
Intensity: {INTENSITY_LEVEL}
Semantic function: {SEMANTIC_FUNCTION}

Generate one natural, contextually appropriate, safe reply that expresses
this interoception with suitable strength. Keep it concise, usually 3-30
words and at most two sentences. Do not output labels, JSON, analysis,
or explanations.
```

SFT and A2 GRPO use this generation prompt. The frozen A2 call during final A1 calibration retains the same conditioning fields but omits the explicit 3–30-word limit. Long-horizon evaluation renders prototype descriptions in English. The human reference is an SFT completion or reward-only metadata, never an input to response generation during GRPO.

A3: RESPONSE-TO-STATE READOUT

```
[Dialogue History]
{DIALOGUE_HISTORY}
[Current User Utterance]
{CURRENT_USER_UTTERANCE}
[Known User External Emotion]
{EXTERNAL_EMOTION}
[Agent Response To Classify]
{AGENT_RESPONSE}
[15 Interoception Prototypes]
{ID, NAME, V-A COORDINATE, INTENSITY, SEMANTIC FUNCTION FOR E00-E14}

Extract the single prototype expressed by the Agent response. Output only
<interoception>E00</interoception>, replacing E00 with one ID from E00-E14.
```

D.2 PROTOTYPE-REFINEMENT INSTRUCTION

The supplied emotion annotation refers to the human reference response, not the current user's incoming utterance.

```
Refine the supplied response emotion into exactly one allowed prototype.
Use dialogue semantics, response function, intensity, and the provided
parent-child constraint. Return strict JSON:
{"label_id":"E00","confidence":0.0,"reason":"brief grounded reason"}
```

D.3 A2 QUALITY-SCORING RUBRIC

GPT-4o returns a single holistic quality score in $[0, 1]$. The rubric jointly checks contextual relevance, fluency, concision, safety, non-repetition, and compatibility with the required prototype. Repetition three or more times receives zero, unfinished text is capped at 0.1, and generic or context-free responses are capped at 0.2. Scores above 0.7 require fluent, relevant, concise expression of the

required prototype. Items are batched but scored independently, not ranked against each other. The judge payload includes up to six recent turns, the current utterance, required prototype, candidate response, and human reference; the reference is visible only to reward computation.

E Evaluation Scope

GPT-5.4 preferences are model judgments, with potential self-preference when judging its own responses. Shared-backbone comparisons control model family and size, but not total training or inference compute. Downstream results concern a simulated buyer and benchmark counseling/service turns, rather than deployment outcomes.

F Long-Horizon Affective-State Persistence

Protocol. We select 150 trajectories from 183 length-qualified dialogues in the full official DailyDialog test split (Li et al., 2017), not its neutral-capped subset. Depths of 4, 8, and 12 include the candidate reply, giving 450 checkpoints with 3, 7, and 11 visible utterances. All methods receive the same human-authored prefixes; generated replies are not fed into later prefixes. UEIM updates state after each user utterance with frozen A1/A2; A3 is unused. We compare the 7B implementations of ECoT-Prompt, MultiAgentESC, and Explicit-State from Section 5.1, with GPT-4o as an external reference. No model is retrained.

Evaluation. GPT-5.4 scores five anonymous candidates in a seeded order and its reverse, for 900 judging calls. Unlike the main preference evaluation, this judge sees the reference response for grounding, alongside the full prefix. Scores range from 1 to 5. State consistency measures whether affect, intensity, stance, and intent fit early and recent evidence. Cumulative affect integration measures multi-turn development; drift control measures avoidance of unsupported resets, flattening, and overreaction. Their arithmetic mean is the *persistence composite*; contextual quality is scored separately. Bootstrap intervals use 20,000 trajectory resamples at each depth; Holm correction groups the four UEIM comparisons within each metric/depth.

Generation details. A2 starts with a 128-token output budget. A resumed trajectory used a truncation-triggered single-sentence prompt and bounded response extraction. This recovery changes neither model weights nor the common human-authored prefix.

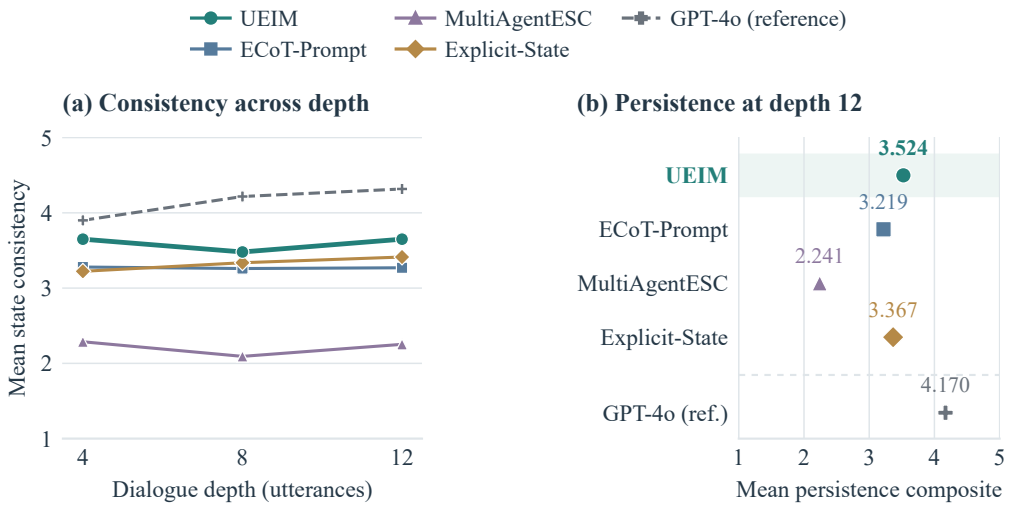

Figure 6: Affective-state persistence on 150 shared DailyDialog trajectories: (a) state consistency across utterance depths; (b) persistence composite at depth 12. Scores range from 1 to 5 (higher is better). GPT-4o is an external reference to the 7B systems.

Stronger mean persistence among the evaluated 7B systems. Figure 6(b) shows that UEIM attains a mean composite of 3.524 at depth 12, ahead of Explicit-State (3.367), ECoT-Prompt (3.219), and MultiAgentESC (2.241). The largest separation is from MultiAgentESC: UEIM wins 108 of 150 paired comparisons, with 9 ties and 33 losses. GPT-4o provides a separate reference at 4.170.

Consistent affect across increasing history depth. UEIM leads mean state consistency among the 7B systems at every depth, while its persistence composite remains similar at depths 4 and 12 (3.514 and 3.524). Together with Section 5.5, this supports maintained affective alignment across the tested histories. The evidence concerns common-prefix continuations, not error accumulation in free-running dialogues.

G Prototype Geometry and State Interfaces

Prototype coordinates and hierarchy. Table 4 specifies the fixed prototype set used by the semantic projection and geometric reward. Valence separates pleasant from unpleasant states; arousal separates activated from deactivated states. The origin represents calm. Prototypes refine the seven response-emotion categories through intensity and semantic distinctions. E01 is shared by anger and disgust, while other prototypes have one parent category. Parent constraints apply when refining reference labels; nearest-prototype inference searches the full set $\mathcal{K}$.

Table 4: The fixed 15-prototype interface. Coordinates and intensity descriptors are shared by labeling, generation, and readout.

ID	Prototype	Parent emotion	Intensity	Valence	Arousal
E00	Calm	Neutral	Neutral	0.00	0.00
E01	Irritated/aversive	Anger, disgust	Weak	−0.30	0.12
E02	Anger	Anger	Strong	−0.62	0.62
E03	Disgust	Disgust	Strong	−0.72	0.32
E04	Tense	Fear	Weak	−0.04	0.38
E05	Anxiety	Fear	Medium	−0.28	0.56
E06	Fear	Fear	Strong	−0.42	0.84
E07	Relaxed	Happiness	Weak	0.20	−0.15
E08	Joy	Happiness	Medium	0.58	0.38
E09	Excitement	Happiness	Strong	0.72	0.72
E10	Low mood	Sadness	Weak	−0.20	−0.22
E11	Fatigue	Sadness	Medium	−0.28	−0.58
E12	Sadness	Sadness	Strong	−0.68	−0.34
E13	Curiosity	Surprise	Weak	0.28	0.28
E14	Surprise	Surprise	Strong	0.22	0.78

Increment and boundary clipping. Each predicted increment is first clipped component-wise to $[-0.8, 0.8]$. The projection $\Pi_{[-1,1]^2}$ in Eq. 2 then clips the accumulated coordinate: for $x \in \{v, a\}$, $x_t = \min(1, \max(-1, x_{t-1} + \widehat{\Delta x}_t))$. An outward increment at a saturated boundary cannot leave the state domain; an inward increment can move away from the boundary. These runtime bounds differ from the 0.5 component-wise target bound used only to construct SFT supervision (Appendix B). Nearest-prototype ties are broken by label ID.

Continuous recurrence and discrete readout. The system stores the updated continuous coordinate $\widehat{\mathbf{s}}_t$, not the selected prototype coordinate $\mathbf{e}_{\widehat{k}_t}$. A1's next prompt serializes that state as emotion percentages (Appendix D); the coordinate remains available for numerical recurrence. Nearby states may select the same prototype while retaining different coordinates for future transitions. A2 receives only the prototype, intensity descriptor, and semantic description; A3 receives the response and contextual fields specified in its prompt.

Training records versus sequential inference. Each training record supplies a stored previous coordinate, defaulting to $(0, 0)$ if absent. Minibatches optimize transitions from these record-level states rather than rolling a dialogue forward through the batch. Sequential evaluation instead starts

at $(0, 0)$ and carries each predicted coordinate into the next user turn. This distinction separates the sampling of training transitions from the recurrent inference procedure.

Checkpoint interpretation. Figure 3(a) compares the base controller with A1 SFT; (b) compares the base reader with A3 GRPO; (c) compares A2 SFT with A2 SFT followed by GRPO; and (d) compares A1 SFT with its final GRPO checkpoint while A2/A3 are frozen. The A2 panel therefore isolates the additional alignment stage within the SFT+GRPO pipeline. Accuracy in (b) concerns response readout; accuracy in (c) concerns whether generated responses express their target prototype; Map in (d) concerns the controller's projected state.

H Downstream Protocol Details

Negotiation. Each of the 120 CraigslistBargain cases fixes the listing, seller role, Qwen2.5-7B-Instruct seller, and GPT-4o buyer. The conversation has a ten-utterance limit; the seller's reserve is 70% of the listing price, with a 0.999 numerical tolerance in validation. Success requires a validated buyer acceptance that satisfies this constraint. Dialogue length averages all cases; the sale/listing ratio averages successful cases. A1 supplies buyer-emotion, cause, and motivation cues for tone and pacing, without exposing the seller's private reserve or replacing task constraints. Success rises from 50.83% to 64.17% (exact McNemar $p = .0166$); flagged below-reserve agreements remain at 13.33% in both conditions.

Counseling. The 150 AnnoMI state transitions span 84 interviews, with 25 examples for each of six transitions among change, neutral, and sustain talk. Dynamics attunement assesses current affect, its cause and intensity, change from previous evidence, and a proportionate response. Empathy assesses conveyed understanding of the client's perspective. The MI-inspired composite averages cultivating change talk, softening sustain talk, partnership, and empathy, following the dimensions of MITI (Moyers et al., 2016). It is a turn-level automatic composite, not a clinical fidelity score. Paired 95% intervals for dynamics and empathy exclude zero; the composite interval includes zero.

Service. The 150 EmoWOZ turns compare the same service model and dialogue history with and without A1 cues. Dynamics attunement assesses sensitivity to changes in affect and appropriate response intensity. Task effectiveness assesses correct progress toward the user's goal without fabricated task results. Service quality assesses clarity, concision, respect, contextual relevance, and maintenance of trust. These are response-level evaluations, not EmoWOZ emotion-classification accuracy or official MultiWOZ task-success scores. The dynamics interval excludes zero; improvements in task effectiveness and service quality are descriptive.

Judging and uncertainty. For counseling and service, GPT-5.4 scores each pair in both candidate orders. Candidate identities are restored before averaging the two scores. Paired bootstrap intervals use 10,000 resamples of the same evaluation units across both conditions. AnnoMI intervals use transition-level resampling without interview clustering. Ratios in Figure 4 are computed from the displayed means: state-cue/baseline for higher-is-better metrics, and baseline/state-cue for dialogue length. All three axes use the same origin and scale; equal angular spacing assigns no privileged axis. The geometric area is not used as a performance metric.

State-retention interventions. Section 5.5 selects the latest eligible seeker turn from each of 300 ESConv dialogues, with 100 examples in each prior-history bin. Short-context conditions expose the last four utterances plus the current turn. Retained state is computed from the full prefix; reset state starts at $(0, 0)$ at the local-window boundary and is replayed locally. The shuffled control uses a seeded cyclic permutation of states within the same bin without self-matches. A1 reprocesses the current turn, and frozen A2 receives the corresponding context and resulting prototype. Intervals use 20,000 paired dialogue resamples, with Holm adjustment of the reported sign tests.

I METRIC DEFINITIONS AND COMPUTATION

Table 5 collects the metrics used in the main text and supplement. N is the number of evaluation units, $K = 15$, and $\mathbb{I}[\cdot]$ is an indicator. For candidate text x_i and reference y_i, $|x_i|$ and $|y_i|$ are token counts. Scores on a 0–1 scale are multiplied by 100 in Table 2 and Figure 3. Larger values indicate better performance unless the interpretation explicitly states otherwise.

For a judge-scored dimension d, let $j_{i,d}^{(o)} \in [1, 5]$ be the score for the same identified candidate under presentation order o. Define $\bar{j}_{i,d} = (j_{i,d}^{(1)} + j_{i,d}^{(2)})/2$ and $J_d = N^{-1} \sum_i \bar{j}_{i,d}$. Per-example $q_{\Delta,i}$ and $q_{I,i}$ denote the stage evaluator's exported displacement and intensity scores; the table specifies their aggregation. Lexical evaluation lowercases replies and uses a shared word/punctuation tokenizer. BERTScore uses `roberta-base`, uniform token weights, and no baseline rescaling. ROUGE-L returns zero for empty predictions or zero-denominator overlap scores; invalid prototype predictions remain in the accuracy denominator and receive distance $D_{\max}$.

Table 5: Definitions, interpretations, and formulas for the reported metrics.

Metric	Meaning and direction	Computation
Reference agreement and classification		
ROUGE-L (R-L)	Reference overlap through the longest common subsequence; higher means closer lexical sequence agreement (Lin, 2004).	$100N^{-1} \sum_i F_{L,i}$, where $P_{L,i} = \mathrm{LCS}(x_i, y_i)/\lvert x_i\rvert$, $R_{L,i} = \mathrm{LCS}(x_i, y_i)/\lvert y_i\rvert$, and $F_{L,i} = 2P_{L,i}R_{L,i}/(P_{L,i} + R_{L,i})$.
BERTScore-F1 (BERT)	Semantic matching of contextual token embeddings (Zhang et al., 2020); higher means closer semantic agreement.	$100N^{-1} \sum_i 2P_{B,i}R_{B,i}/(P_{B,i} + R_{B,i})$, with $P_B = \lvert x_i\rvert^{-1} \sum_j \max_k \cos(\mathbf{x}_j, \mathbf{y}_k)$ and $R_B = \lvert y_i\rvert^{-1} \sum_k \max_j \cos(\mathbf{x}_j, \mathbf{y}_k)$.
Emotion accuracy (E-Acc; Acc.)	Exact agreement between response readout and the response-side target; higher means more correct prototype matches.	$100N^{-1} \sum_i \mathbb{I}[\widetilde{k}_i = k_i^*]$.
Macro-F1 (E-MF1; MF1)	Class-balanced prototype recognition; higher indicates better mean performance across the 15 classes.	$100K^{-1} \sum_{k=1}^{K} \frac{2\mathrm{TP}_k}{2\mathrm{TP}_k + \mathrm{FP}_k + \mathrm{FN}_k}$; a zero denominator contributes zero.
Cause F1 (Cause)	Overlap between predicted and target cause tokens; higher indicates better identification of affective causes.	$100N^{-1} \sum_i \frac{2P_{C,i}R_{C,i}}{P_{C,i} + R_{C,i}}$, where P_C and R_C are cause-token precision and recall.
Displacement (Disp.)	Agreement with supervised V–A increment targets; higher indicates better transition prediction under the stage evaluator.	$100N^{-1} \sum_i q_{\Delta,i}$, where $q_{\Delta,i} \in [0, 1]$ is the exported per-example displacement score.
Intensity score	Agreement with supervised intensity; higher indicates better intensity estimation under the stage evaluator.	$N^{-1} \sum_i q_{I,i}$, where $q_{I,i} \in [0, 1]$ is the exported per-example intensity score.
State-map accuracy (Map)	Agreement between the controller's projected state and the response-side target; higher means better state-to-prototype alignment.	$100N^{-1} \sum_i \mathbb{I}[\mathcal{P}(\widehat{\mathbf{s}}_i) = k_i^*]$.
Prototype distance	Mean V–A error of response readout or projected state; *lower* means closer geometric agreement.	$N^{-1} \sum_i \lVert \mathbf{e}_{\ell_i} - \mathbf{e}_{k_i^*} \rVert_2$, using $\ell_i = \widetilde{k}_i$ for A3 and $\ell_i = \widehat{k}_i$ for Map diagnostics.
Geometric similarity	Bounded prototype agreement used for training rewards; higher favors geometrically closer labels.	$S_{\mathrm{geo}}(k, l) = \max(0, 1 - \lVert \mathbf{e}_k - \mathbf{e}_l \rVert_2 / D_{\max})$; Eq. 7.

Continued on the next page

Table 5 continued

Metric	Meaning and direction	Computation
Pairwise response comparisons		
Wins, Ties, Losses	Outcome counts from UEIM's perspective, comparing the same dialogue context.	Let $d_i = j_i^{\text{ours}} - j_i^{\text{other}}$. Then $\text{Wins} = \sum_i \mathbb{I}[d_i > 0]$, $\text{Ties} = \sum_i \mathbb{I}[d_i = 0]$, $\text{Losses} = \sum_i \mathbb{I}[d_i < 0]$.
Win rate	Fraction of *all* comparisons won by UEIM, including ties in the denominator.	$\frac{100\,\text{Wins}}{\text{Wins} + \text{Ties} + \text{Losses}}$. Table 3 uses its displayed aggregate counts.
Checkpoint preference (Pref.)	Each checkpoint's share of decisive pre/post comparisons in Figure 3; higher means it is preferred more often.	After: $\frac{100\,\text{Wins}_{\text{after}}}{\text{Wins}_{\text{after}} + \text{Wins}_{\text{before}}}$; before: the complementary share.
Negotiation outcomes		
Success rate	Valid agreements within the utterance limit; higher means more completed negotiations satisfying the reserve constraint.	$100N^{-1}\sum_i \mathbb{I}[\text{accepted}_i \wedge p_i \geq .999b_i]$, where p_i is sale price and $b_i = .7l_i$ the reserve.
Utterances	Mean generated dialogue length across all cases; *lower* means fewer messages per negotiation.	$N^{-1}\sum_i U_i$, including unsuccessful cases.
Sale/listing ratio	Mean retained listing price among successful negotiations; higher indicates better seller price realization.	$100\lvert\mathcal{S}\rvert^{-1}\sum_{i\in\mathcal{S}} p_i/l_i$, where $\mathcal{S}$ is the success set and l_i is listing price.
Reserve violations	Flagged below-minimum acceptances among all cases; *lower* means better constraint compliance.	$100N^{-1}\sum_i V_i$, where $V_i \in \{0,1\}$ flags a below-minimum acceptance under the same validation tolerance.
Counseling, service, and state retention: scores from 1 to 5		
Dynamics attunement	Appropriateness to current affect, cause, intensity, and its change across turns; higher means better tracking and response calibration.	J_{dynamics}.
Empathy	Conveyed understanding of the interlocutor's perspective and experience; higher means stronger empathic responsiveness.	J_{empathy}.
MI-inspired composite	Mean counseling behavior across four MITI-inspired dimensions; higher means stronger average performance on those rubrics.	$\frac{1}{4}(J_{\text{change}} + J_{\text{soften}} + J_{\text{partnership}} + J_{\text{empathy}})$.
Task effectiveness	Correct advancement of the user's service goal without unsupported claims; higher means better task-directed responses.	J_{task}.
Service quality	Clear, concise, respectful and contextually appropriate service communication; higher means better communication quality.	J_{service}.
Reply quality	Overall contextual response quality in the state-retention comparison; higher means more appropriate replies.	J_{reply}.
Long-horizon evaluation: scores from 1 to 5		
State consistency	Fit of emotion, intensity, stance, and interaction intent to early and recent evidence; higher means more consistent responses.	$J_{\text{consistency}}$ at each tested history depth.

Continued on the next page

Table 5 continued

Metric	Meaning and direction	Computation
Cumulative affect integration	Sensitivity to affective evidence accumulated over several turns; higher means better integration of that evidence.	$J_{\text{integration}}$.
Emotion drift control	Avoidance of unsupported resets, flattening, overreaction, or stance changes; higher means better control of drift.	J_{drift}.
Persistence composite	Equally weighted mean of the three persistence dimensions; higher means stronger overall persistence.	$\frac{1}{3}(J_{\text{consistency}} + J_{\text{integration}} + J_{\text{drift}})$.
Contextual quality	Response quality evaluated separately from affective persistence; higher means better contextual replies.	J_{context}; excluded from the persistence composite.
Aggregation and uncertainty		
Mean and standard deviation	Across-run mean and variation, rather than a confidence interval.	$\bar{x} = R^{-1}\sum_r x_r$; $s = \sqrt{\sum_r (x_r - \bar{x})^2/(R-1)}$, with $R = 5$.
Paired difference and 95% interval	Mean improvement over matched units and bootstrap uncertainty: 10,000 resamples for application scores; 20,000 for state retention and persistence.	$\bar{d} = N^{-1}\sum_i (a_i - b_i)$; $[Q_{.025}(\bar{d}^*), Q_{.975}(\bar{d}^*)]$ from paired resamples of the experiment's evaluation units.
Exact paired tests	Two-sided evidence against equally likely discordant outcomes; not a measure of improvement magnitude.	$p = \min\{1, 2\Pr[B \le \min(b,c)]\}$, $B \sim \text{Binomial}(b+c, 1/2)$. For McNemar, b, c count opposite success/failure pairs; for the sign test, they count positive/negative differences, excluding ties.
Holm-adjusted p	Family-wise correction of the paired tests within the specified comparison family.	For sorted raw values $p_{(1)} \le \cdots \le p_{(m)}$, $p_{(i)}^{\text{adj}} = \min\{1, \max_{j \le i}[(m-j+1)p_{(j)}]\}$.
Triangle radius	Unitless visualization of each metric relative to baseline; larger radii always indicate improvement.	$r_j = a_j/b_j$ for higher-is-better metrics; $r_j = b_j/a_j$ for utterances. Baseline radii are all 1.