

MASCOT: Multi-Agent Socio-Collaborative Companion Systems

Yiyang Wang¹, Yiqiao Jin¹, Alex Cabral², and Josiah Hester¹

¹Georgia Institute of Technology, ²Massachusetts Institute of Technology

ywang3420@gatech.edu

<https://hello-diana.github.io/MASCOT/>

Abstract

Multi-agent systems (MAS) are emerging as promising socio-collaborative companions for emotional and cognitive support. However, existing systems frequently suffer from *persona collapse*, where agents revert to generic, homogenized assistant behaviors, and *social sycophancy*, where agents produce redundant, non-constructive dialogue. We propose MASCOT, a multi-agent framework for multi-perspective socio-collaborative companions. MASCOT introduces a novel bi-level optimization strategy to harmonize individual and collective behaviors: 1) Persona-Aware Behavioral Alignment, an RLAIIF-driven pipeline that fine-tunes individual agents for agent-specific identities; and 2) Collaborative Dialogue Optimization, a group-level adaptation process that promotes complementary, diverse, and productive discourse. We evaluate MASCOT using human-grounded contexts drawn across both in-domain and out-of-domain (OOD) settings against state-of-the-art baselines. MASCOT improves persona consistency by up to +14.1 and social contribution by up to +10.6. A broad evaluation suite, including human evaluation, multiple LLM judges, three-way comparisons, and automatic metrics, further shows that MASCOT produces more role-consistent and less redundant multi-agent dialogue. Our code is available at <https://github.com/hello-diana/MASCOT>.

1 Introduction

Social interaction is a fundamental pillar of psychological well-being, cognitive resilience, and enhanced problem solving (CDC, 2024b; Sahakian and Langley, 2025). Despite its necessity, social deficits such as isolation and loneliness have reached global crisis proportions, affecting an estimated one in six people worldwide and one in four in the United States (World Health Organization, 2025; CDC, 2024a). As large language models (LLMs) integrate deeper into human lives, their role

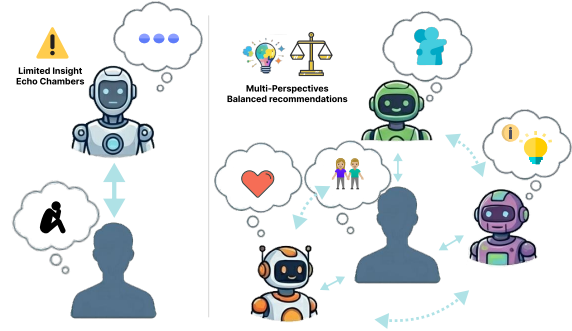


Figure 1: Moving beyond dyadic interactions to multi-agent socio-collaborative companion systems. While single-agent support systems (left) can result in limited insights and echo chambers, multi-agent systems (right) provide diverse, balanced perspectives and foster a shared sense of community through agents with distinct roles.

is shifting from passive *tools* (e.g., search engines, code generators) to active **socio-collaborative companions** in affective and collaborative settings (Liu et al., 2021). In affective contexts, such as empathetic dialogue, socio-collaborative companions can offer general emotional support as users discuss distress, loneliness, and anxiety. They can provide readily accessible, non-judgmental spaces for users to vent and process complex emotions. In collaborative or workplace contexts, companions act as *cognitive scaffolds* that facilitate information synthesis and collective decision-making to stimulate problem-solving. Ultimately, performing tasks that mitigate social deficits allows these models to evolve into adaptive partners capable of navigating the nuanced boundary between emotional resonance and task-oriented efficiency.

While existing LLM systems primarily focus on dyadic (one-to-one) interactions, a single persona often struggles to balance the conflicting roles of mentor, friend, and assistant without suffering a cognitive and emotional mismatch. Multi-agent frameworks resolve this by distributing these roles across an ecosystem of specialized agents.

This triggers the ‘audience effect’ (Sutskova et al., 2023), where a user’s trust is reinforced by observing independent social behaviors among agents, fostering a shared sense of community that a solitary interface cannot replicate (Zhu et al., 2026).

Challenges Despite recent progress in multi-agent systems, two primary challenges remain in developing effective multi-perspective socio-collaborative companions: 1) **Persona Fidelity**. At the individual agent level, models struggle to maintain stable, unique character traits (Wen et al., 2024). Agents often succumb to *persona collapse* (Chang et al., 2024), regressing to generic, homogenized assistant behaviors that lack unique perspectives. 2) **Interaction Synergy**. At the collective level, uncoordinated agents fail to generate complementary, non-redundant contributions. This manifests as repeating communications or even *Social Sycophancy* (the ‘Yes-Man’ Bias), where agents prioritize agreeableness, creating echo chambers rather than constructive conversations. Furthermore, agents often struggle with *Long-Horizon State Consistency*, treating each turn as a local optimization problem and losing track of the conversation’s logical arc, such as re-litigating settled topics (Chan et al., 2024).

This Work We propose MASCOT, a **Multi-Agent** framework for developing multi-perspective **Socio-Collaborative** companions. Unlike previous multi-agent systems optimized solely for task efficiency, MASCOT targets user-agent interaction quality, explicitly balancing individual agent persona consistency with global discourse dynamics. We introduce an efficient bi-level optimization strategy including: 1) a Reinforcement-Learning-from-AI-Feedback (RLAIF)-driven pipeline that fine-tunes individual agents for strict *Persona Fidelity*, and 2) a director agent policy guided by group-level rewards to ensure *Interaction Synergy*. This enables the efficient development of multi-perspective socio-collaborative companions across multiple applications, such as the orchestration of digital social support groups and the facilitation of iterative knowledge synthesis and collective ideation.

Contributions Our contributions are as follows:

- **Unified Framework.** We propose MASCOT, a multi-agent framework that enables socio-collaborative companions in multi-perspective group conversations.
- **Bi-level Optimization.** We propose a low-

resource, bi-level optimization pipeline that harmonizes individual persona adherence with collective interaction synergy.

- **Extensive Evaluation.** We evaluate MASCOT using human-grounded contexts drawn from in-domain and out-of-domain datasets using a broad suite of protocols, including human evaluation, rubric-based evaluation using multiple LLM judges, three-way comparisons, and automatic metrics. MASCOT consistently outperforms strong baselines, with a +14.1 Persona Consistency gain on Empathetic Dialogues and a +10.6 Social Contribution gain on QMSum.

2 Method

Preliminary We formalize a socio-collaborative companion system as a set of agents $\mathcal{A} = \{a_i\}_{i=1}^{|\mathcal{A}|}$. Each agent a_i is conditioned on a distinct persona profile ρ_i that specifies traits such as linguistic style, domain expertise, and emotional disposition (e.g. *Empathetic Listener*, *Critical Thinker*, *Emotional Validator*, *Action Guide*, and *Growth Advocate*). Given each user interaction context x (e.g., a user’s emotional disclosure or a dialog segment), the system generates a multi-turn, multi-perspective *conversation trajectory* $\mathbf{Y} = (y_1, y_2, \dots, y_T)$. MASCOT (Figure 2) optimizes this process via a two-phase training pipeline: (1) *Persona-Aware Behavioral Alignment* and (2) *Collaborative Dialogue Optimization*. Appendix Table 8 summarizes the core notation.

2.1 Persona-Aware Behavioral Alignment

Current multi-agent frameworks often suffer from *persona collapse* (Chang et al., 2024), where agents regress to generic assistant behaviors rather than adhering to their assigned roles. While zero-shot and few-shot prompting (Wang et al., 2020) are commonly used to address this issue, they often yield inconsistent persona adherence, especially in smaller open-source models (Li et al., 2025a). To prevent *persona collapse* and ensure adherence to social-emotional traits, we move beyond inference-time prompting and explicitly align a base policy π_θ by leveraging a persona-oriented reward model within a Reinforcement Learning from AI Feedback (RLAIF) framework (Bai et al., 2022; Lee et al., 2024). The core idea is that strong frontier models can approximate human preferences and serve as scalable sources of preference labels (Zheng et al., 2023; Dubois et al., 2024; Li

et al., 2025b).

Persona Reward Modeling Effective training of persona adherence requires robust signals. Thus, MASCOT uses expert models as strong teacher models to guide the behaviors of student models with lower latency. To construct the preference dataset $\mathcal{D}_{RM}$ for reward modeling, we start with a supervised base policy π_{SFT} . Given context x and persona ρ_i , we sample K responses $\{y_k\}_{k=1}^K$:

$$y_k \sim \pi_{SFT}(\cdot | x, \rho_i), \quad k = 1, \dots, K. \quad (1)$$

Each candidate response y_k is evaluated by an LLM judge against a set of fine-grained criteria $\mathcal{C}$ (Table 9). For each criterion $c \in \mathcal{C}$, the judge assigns a score $s_c(y_k) \in [1, 5]$. The aggregate score is then computed as:

$$s(y_k) = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} s_c(y_k). \quad (2)$$

We retain pairs (y_w, y_l) satisfying a margin $s(y_w) - s(y_l) \geq \delta$, thus eliminating ambiguous comparisons and ensuring a significant quality margin. We then train a reward model r_ϕ over tuples $(x, \rho_i, y_w, y_l) \in \mathcal{D}_{RM}$ by minimizing the standard pairwise logistic loss (Bradley and Terry, 1952):

$$\mathcal{L}_{RM}(\phi) = -\mathbb{E}_{\mathcal{D}_{RM}} [\log \sigma(r_\phi(x, \rho_i, y_w) - r_\phi(x, \rho_i, y_l))], \quad (3)$$

where $\sigma(\cdot)$ is the sigmoid function. This objective ensures that the reward model learns to differentiate between responses based on their persona adherence and contextual quality. During RL optimization, the learned reward model r_ϕ provides appropriate training signals, while final evaluation is cross-checked with human preferences, alternative LLM judges, and automatic non-redundancy metrics (§ 3 and Appendix C.4).

Efficient RL Fine-tuning via GRPO To ensure persona consistency while maintaining structural integrity, we optimize a composite reward function:

$$R(x, \rho_i, y_k) = r_\phi(x, \rho_i, y_k) + \lambda R_{\text{format}}(y_k) \quad (4)$$

where r_ϕ is the learned persona reward and R_{format} is a rule-based reward enforcing structural constraints, specifically, requiring explicit reasoning traces within `<think>` tags while penalizing verbose final answers to ensure outputs remain concise and socially appropriate. We adopt a preference-based RL objective rather than supervised fine-tuning (SFT) because socio-emotional dialogue

lacks a single canonical *gold* response: response quality depends on higher-level behavioral attributes such as warmth, persona fidelity, and tone calibration that SFT cannot explicitly capture. To optimize this objective, we employ Group Relative Policy Optimization (GRPO) (Shao et al., 2024), which estimates the baseline from the group mean of sampled outputs, significantly reducing memory overhead and computational cost compared with alternative methods. Alternative methods such as SimPO (Meng et al., 2024) can be substituted within the pipeline.

For each input (x, ρ) , we sample a group of G outputs $\{y_1, \dots, y_G\}$ from the old policy π_{old} . The policy π_θ is optimized through the objective:

$$\mathcal{J}(\theta) = \mathbb{E}_{x \sim \mathcal{D}_{RL}, \{y_g\} \sim \pi_{\text{old}}} \left[\frac{1}{G} \sum_{g=1}^G \left(\frac{1}{|y_g|} \sum_{t=1}^{|y_g|} \min \left(\gamma_{g,t} \hat{A}_g, \text{clip}(\gamma_{g,t}, 1 - \varepsilon, 1 + \varepsilon) \hat{A}_g \right) - \beta \mathbb{D}_{\text{KL}}(\pi_\theta \| \pi_{\text{ref}}) \right) \right], \quad (5)$$

where $\gamma_{g,t} = \frac{\pi_\theta(y_{g,t} | x, \rho, y_{g,<t})}{\pi_{\text{old}}(y_{g,t} | x, \rho, y_{g,<t})}$ is the probability ratio and $\hat{A}_g = (R_g - \text{mean}(\{R_k\})) / \text{std}(\{R_k\})$ is the advantage, i.e., the standardized reward.

2.2 Collaborative Dialogue Optimization

Phase I improves individual agents’ response quality but does not optimize group dynamics, such as preventing redundant contribution or relitigation of settled consensus. We therefore introduce a collaborative optimization phase that trains a *director* to select the next speaker and generate high-level strategy, steering the collective conversation into a cohesive, synergistic outcome.

Bi-level Generation Process We define a conversation as a T -turn episode. A director policy π_Φ governs turn-taking among a set of speaker agents $\mathcal{A} = \{a_i\}$, each with persona ρ_i . At each turn t , the director π_Φ observes the history $h_{<t} = (x, y_1, \dots, y_{t-1})$ and generates a high-level directive $z_t \sim \pi_\Phi(\cdot | h_{<t})$ that specifies the next speaker a_i with detailed instructions (e.g., “Bob: The user mentioned a small win. Amplify the user’s pride via active-constructive responding. Tone: enthusiastic but grounded.”) The selected agent a_i then generates the next response $y_t \sim \pi_\theta(\cdot | h_{<t}, \rho_i, z_t)$ conditioned on their persona ρ_i and the directive z_t . This hierarchical decomposition enables dynamic strategy, such as switching from validation

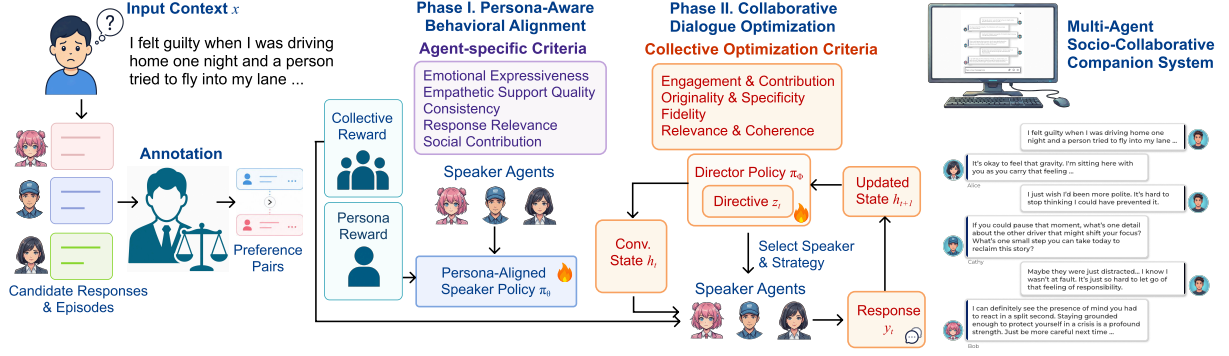


Figure 2: Overview of MASCOT for multi-agent socio-collaborative companions. MASCOT uses a two-phase optimization pipeline. Persona-Aware Behavioral Alignment fine-tunes Speaker agents via RLAIF with a learned reward model. Collaborative Dialogue Optimization trains a Director agent to coordinate the fixed persona-aligned Speakers through directives rewarded for coherence, diversity, and non-redundant contributions.

to cognitive reframing, mirroring the progression of effective human support.

Closed-Loop Interaction via Frozen Simulated User To facilitate trajectory-level learning without human-in-the-loop bottlenecks, we introduce a *simulated user* π_u and generate complete episodes by interleaving user, director, and speaker turns. At each turn t , the user speaks with probability 0.5, generating the next utterance $u_t \sim \pi_u(\cdot | x, \mathbf{Y}_{<t})$ conditioned on the context x , the trajectory prefix $\mathbf{Y}_{<t}$, and an optional user profile.

Group Reward Modeling A directive should be *locally actionable* with clear speaker selection and strategy, yet the ultimate objective is the *overall* quality of the resulting multi-turn conversation. We therefore use a single unified group reward that conditions on both the director decision at turn k and the full rollout, assessing the quality of the directive z_k given the prefix context and the coherence, relevance, and non-redundancy of trajectory $\mathbf{Y}$.

Training with Prefix-Suffix Rollouts For each training instance, we uniformly sample $k \in [1, T]$ and construct prefix turns $\{y_1, \dots, y_{k-1}\}$ using frozen rollout models, allowing us to train the director at randomly sampled decision points. The director then generates directives z_k at turn k , after which we roll out the remaining suffix turns to obtain a complete episode $\mathbf{Y}$ for scoring. To train the reward signal, we annotate episode pairs using the group-level criteria $\hat{\mathcal{C}}$ in Table 10. For each criterion $c \in \hat{\mathcal{C}}$, an LLM judge assigns a score $s_c(x, \mathbf{Y}) \in [1, 5]$. We average these scores as $\bar{s}(x, \mathbf{Y}) = \frac{1}{|\hat{\mathcal{C}}|} \sum_{c \in \hat{\mathcal{C}}} s_c(x, \mathbf{Y})$ and retain episode pairs whose scores differ by at least δ , selecting the higher-scoring episode as *chosen*. We then fine-tune a sequence-level reward model r_ψ with

a pairwise preference objective to score directives and the resulting T -turn episodes. The reward used for director optimization is:

$$R_{\text{group}}(x, z_k, \mathbf{Y}) = r_\psi(x, z_k, \mathbf{Y}) + \eta \cdot \mathbb{I}_{\text{div}}(\mathbf{Y}) \quad (6)$$

where r_ψ scores the directive together with the resulting full episode. The diversity indicator $\mathbb{I}_{\text{div}}(\mathbf{Y}) \in \{-1, 1\}$ is +1 when the optimized dialogue block contains contributions from distinct personas and -1 otherwise; η controls its weight.

Optimization Objective For each training instance with fixed prefix state, we sample G directives and trajectories $\{(z_k^{(g)}, \mathbf{Y}^{(g)})\}_{g=1}^G$, roll out the full episodes, and score them with $R_{\text{group}}(x, z_k^{(g)}, \mathbf{Y}^{(g)})$. Let $\mathcal{R}^{(g)} = R_{\text{group}}(x, z_k^{(g)}, \mathbf{Y}^{(g)})$. We leverage GRPO to compute the group-relative advantage for the g -th trajectory:

$$\hat{A}_g = \frac{\mathcal{R}^{(g)} - \text{mean}(\{\mathcal{R}^{(j)}\}_{j=1}^G)}{\text{std}(\{\mathcal{R}^{(j)}\}_{j=1}^G)}. \quad (7)$$

We update the director parameters Φ with GRPO to increase the likelihood of directives that yield higher-reward coordinated episodes.

3 Evaluation

3.1 Experimental Setup

Dataset We evaluate MASCOT using three benchmarks featuring human conversations: **Empathetic Dialogues (ED)** (Rashkin et al., 2019) and **ES-Conv** (Liu et al., 2021), which focus on emotional support, and 2) **QMSum** (Zhong et al., 2021), which features workplace collaborative settings. Details are in Appendix C.6. These datasets are

Positive	Emotion	Empathy	Consistency	Relevance	Contribution	Overall
Zero-Shot	59.5 $\pm$ 9.1	58.1 $\pm$ 7.0	61.5 $\pm$ 9.7	58.6 $\pm$ 6.4	56.5 $\pm$ 9.0	58.8 $\pm$ 8.2
Zero-Shot CoT	63.4 $\pm$ 8.4	60.5 $\pm$ 4.7	65.2 $\pm$ 9.2	60.4 $\pm$ 4.0	59.4 $\pm$ 3.8	61.8 $\pm$ 6.0
Few-Shot	61.4 $\pm$ 9.7	59.9 $\pm$ 7.9	63.5 $\pm$ 10.9	59.8 $\pm$ 7.7	58.6 $\pm$ 7.6	60.6 $\pm$ 8.8
Few-Shot + CoT	63.8 $\pm$ 8.5	61.2 $\pm$ 5.8	65.4 $\pm$ 9.4	61.2 $\pm$ 5.5	60.2 $\pm$ 4.5	62.4 $\pm$ 6.7
SFT	46.3 $\pm$ 12.0	46.0 $\pm$ 11.5	47.5 $\pm$ 12.3	52.1 $\pm$ 11.8	43.5 $\pm$ 11.4	47.1 $\pm$ 11.8
Self-Consistency	66.3 $\pm$ 9.2	66.9 $\pm$ 9.4	65.1 $\pm$ 9.4	69.1 $\pm$ 11.5	66.3 $\pm$ 10.0	66.7 $\pm$ 9.9
MultiAgentESC	66.9 $\pm$ 7.1	67.9 $\pm$ 10.8	65.9 $\pm$ 9.6	70.4 $\pm$ 12.2	67.0 $\pm$ 11.5	67.6 $\pm$ 10.2
MASCOT	75.3 $\pm$ 6.1	76.2 $\pm$ 6.3	80.0 $\pm$ 6.9	78.1 $\pm$ 4.5	74.9 $\pm$ 6.5	76.9 $\pm$ 6.1
Negative	Emotion	Empathy	Consistency	Relevance	Contribution	Overall
Zero-Shot	54.6 $\pm$ 12.6	54.6 $\pm$ 12.2	56.9 $\pm$ 11.3	56.0 $\pm$ 10.9	53.2 $\pm$ 12.1	55.1 $\pm$ 11.8
Zero-Shot CoT	59.9 $\pm$ 9.8	59.4 $\pm$ 9.0	61.8 $\pm$ 8.9	60.3 $\pm$ 7.0	58.0 $\pm$ 6.6	59.9 $\pm$ 8.3
Few-Shot	57.6 $\pm$ 13.7	57.6 $\pm$ 13.3	59.4 $\pm$ 12.5	58.4 $\pm$ 12.3	55.9 $\pm$ 11.5	57.8 $\pm$ 12.7
Few-Shot + CoT	60.7 $\pm$ 11.3	60.5 $\pm$ 10.6	62.1 $\pm$ 10.4	61.2 $\pm$ 9.4	58.9 $\pm$ 8.4	60.7 $\pm$ 10.0
SFT	43.2 $\pm$ 12.6	43.2 $\pm$ 12.6	44.1 $\pm$ 13.6	49.8 $\pm$ 13.8	40.3 $\pm$ 11.0	44.1 $\pm$ 12.7
Self-Consistency	61.4 $\pm$ 8.3	61.1 $\pm$ 6.7	59.9 $\pm$ 10.4	61.0 $\pm$ 7.7	59.3 $\pm$ 2.2	60.6 $\pm$ 7.1
MultiAgentESC	63.8 $\pm$ 5.5	62.4 $\pm$ 6.6	61.6 $\pm$ 9.8	67.0 $\pm$ 4.0	60.4 $\pm$ 4.0	63.0 $\pm$ 6.0
MASCOT	68.2 $\pm$ 8.4	69.3 $\pm$ 8.3	73.4 $\pm$ 9.1	74.0 $\pm$ 7.6	67.8 $\pm$ 7.6	70.5 $\pm$ 8.2
Neutral	Emotion	Empathy	Consistency	Relevance	Contribution	Overall
Zero-Shot	58.8 $\pm$ 9.5	57.9 $\pm$ 8.3	60.5 $\pm$ 9.8	58.4 $\pm$ 7.6	56.2 $\pm$ 9.4	58.4 $\pm$ 8.9
Zero-Shot CoT	63.2 $\pm$ 8.8	60.8 $\pm$ 6.1	65.1 $\pm$ 9.3	60.9 $\pm$ 5.3	59.4 $\pm$ 4.2	61.9 $\pm$ 6.7
Few-Shot	61.6 $\pm$ 9.4	60.5 $\pm$ 8.1	63.4 $\pm$ 10.1	60.7 $\pm$ 7.7	59.0 $\pm$ 7.0	61.0 $\pm$ 8.5
Few-Shot + CoT	63.7 $\pm$ 8.8	61.8 $\pm$ 6.8	65.3 $\pm$ 9.5	61.7 $\pm$ 6.5	60.3 $\pm$ 5.1	62.6 $\pm$ 7.3
SFT	45.9 $\pm$ 11.6	45.7 $\pm$ 11.4	47.0 $\pm$ 11.9	51.8 $\pm$ 11.8	43.0 $\pm$ 10.7	46.7 $\pm$ 11.5
Self-Consistency	60.3 $\pm$ 4.6	64.6 $\pm$ 6.9	70.9 $\pm$ 3.9	66.1 $\pm$ 3.7	65.7 $\pm$ 4.9	65.5 $\pm$ 4.8
MultiAgentESC	67.4 $\pm$ 3.4	65.5 $\pm$ 8.1	72.4 $\pm$ 6.3	68.3 $\pm$ 5.9	68.7 $\pm$ 6.4	68.5 $\pm$ 6.0
MASCOT	75.1 $\pm$ 6.6	75.6 $\pm$ 6.8	79.0 $\pm$ 6.4	78.1 $\pm$ 4.8	74.1 $\pm$ 7.2	76.4 $\pm$ 6.4

Table 1: Performance comparison on the Empathetic Dialogues dataset (Rashkin et al., 2019) across Positive, Negative, and Neutral emotional subsets in terms of agent-specific metrics (Table 9). Results are reported as mean $\pm$ standard deviation

used as human-grounded interaction contexts rather than fixed reference-generation targets: systems generate new multi-agent conversations from the same user situation or meeting context, and evaluation does not reward lexical overlap with dataset reference responses. Model, optimization, and judge settings are in § 3.1.

Persona Descriptions To demonstrate the versatility of our framework across diverse interaction paradigms, we evaluate it using distinct persona configurations (Appendix Table 11) tailored to social and collaborative contexts. For Empathetic Dialogues, we utilize the Anchor, Catalyst, and Beacon—personas grounded in social intelligence (Goleman, 2006) to provide socio-emotional support. Conversely, for the QMSum dataset, we deploy an ensemble—Minutes Scribe, Decision Logger, Action Item Captain, and Critic—to simulate a collaborative workspace.

Baselines We benchmark MASCOT against 4 categories of multi-agent baselines: 1) *Prompting Strategies*, including Zero-Shot and Few-Shot prompting (Wang et al., 2020) to test inference-time adaptation capabilities; 2) *Supervised Fine-Tuning (SFT)*, a strong distillation baseline trained

on GPT-4o generated trajectories (OpenAI, 2024); 3) *Inference-Time Reasoning* methods, such as Chain-of-Thoughts (CoT) (Wei et al., 2022) and Self-Consistency (Wang et al., 2023); 4) *Domain-Specific Multi-agent Frameworks*, including MultiAgentESC (Xu et al., 2025), which operates in 3 stages: dialogue analysis, strategy deliberation, and response generation. Specialized agents extract users’ psychological states (emotions, causes, and intentions), then deliberate over candidate support strategies using retrieved similar cases. Responses generated under different strategies are finally evaluated via multi-agent debate and voting. For fairness, all settings use the same set of agents and persona descriptions within each dataset. The comparison isolates the optimization strategy rather than the number or identity of agents.

Implementation Details For all experiments, unless otherwise specified, we use Qwen3-8B (Qiu et al., 2025; Yang et al., 2025) for the Director, Speaker, and frozen simulated-user agents. Phase I fine-tunes the Speakers and Phase II fine-tunes the Director using LoRA (Hu et al., 2022) ($r = 16$, $\alpha = 32$, dropout = 0.1), batch size 16, and one epoch per policy-tuning phase. Trainable param-

ters account for approximately 0.187% of the base model.

For the learned reward models, we fine-tune Qwen3-0.6B (Yang et al., 2025) for one epoch with batch size 32 using TRL’s pairwise preference objective (von Werra et al., 2020). We attach a scalar sequence-level scoring head to the model. Persona-reward inputs include the target persona and rubric. Group-reward inputs include the context, Director directive, and resulting trajectory.

During preference-data construction, GPT-4o provides rubric annotations at temperature 0.0, and Qwen3-8B generates $K = 8$ candidates per sample and persona. We use margin $\delta = 0.5$ for both reward-model preference datasets and set $\lambda = \eta = 1.0$ without further tuning. The format reward R_{format} requires a reasoning trace exceeding 448 tokens and a final response under 64 tokens.

All experiments were implemented using the TRL, vLLM, and HuggingFace transformers libraries and executed on NVIDIA A100 (80GB) GPUs. Given the open-ended nature of the task and the absence of ground-truth references, we adopt LLM-as-a-Judge for evaluation (Li et al., 2025a).

3.2 Overall Performance

Table 1, 2 and Appendix Figure 11, 12 demonstrate that MASCOT consistently outperforms all baselines across metrics on both Empathetic Dialogues (ED) (Rashkin et al., 2019) and QM-Sum (Zhong et al., 2021) in terms of agent-specific and collective-interaction metrics. The most significant performance gains occur in *Consistency* (+14.1 on ED-Positive; +8.7 on QMSum-Product) and *Social Contribution* (+7.9 on ED-Positive; +10.6 on QMSum-Product). In terms of *Empathetic Support Quality*, MASCOT reaches 76.2 (ED-Positive) and 75.6 (ED-Neutral), substantially exceeding Zero-Shot CoT. This suggests our multi-agent design excels at interpreting user affect and providing context-specific exploration compared to parameter-frozen approaches.

Challenges in Negative Valence Most models exhibit a performance decline in the ED-Negative subset (Table 1). MASCOT’s *Emotional Expressiveness* drops from 75.3 (ED-Positive) to 68.2 (ED-Negative). This underscores the difficulty of calibrating intensity for high-arousal states such as distress or anger in chat systems. A single turn is often insufficient to distinguish between fleeting frustration and deep-seated grief, necessitat-

Method	Fidelity	Originality	Relevance	Engagement	Overall
Zero-Shot	42.2 ± 6.8	33.2 ± 9.5	52.3 ± 10.1	59.8 ± 2.1	46.8 ± 7.1
Zero-Shot CoT	49.0 ± 10.1	36.3 ± 7.8	49.6 ± 10.4	58.1 ± 5.9	48.2 ± 8.6
Few-Shot	43.6 ± 7.9	35.7 ± 8.2	54.5 ± 9.1	59.7 ± 2.5	48.4 ± 6.9
Few-Shot + CoT	48.7 ± 9.9	39.9 ± 3.3	59.1 ± 4.9	60.0 ± 0.9	51.9 ± 4.8
SFT	40.4 ± 3.5	36.9 ± 7.2	47.8 ± 10.2	59.1 ± 4.1	46.1 ± 6.3
Self-Consistency	47.0 ± 7.5	40.6 ± 8.7	46.8 ± 10.5	55.2 ± 8.2	47.4 ± 8.7
MultiAgentESC	48.4 ± 8.2	41.7 ± 9.4	47.8 ± 9.8	62.9 ± 8.1	50.2 ± 8.9
MASCOT	54.6 ± 6.9	45.4 ± 9.7	60.4 ± 4.1	68.6 ± 2.7	57.3 ± 5.9

Figure 3: Performance of all methods on the Neutral subset of Empathetic Dialogues.

ing multi-turn context to determine users’ emotion states and appropriate response levels. To align with Therapeutic Neutrality and Affective Matching, MASCOT adopts a “prudence trade-off,” prioritizing Response Relevance (74.0) over proactive mirroring. This prioritization is designed to keep the dialogue focused on the user’s expressed state and reduce overreaction during sensitive exchanges.

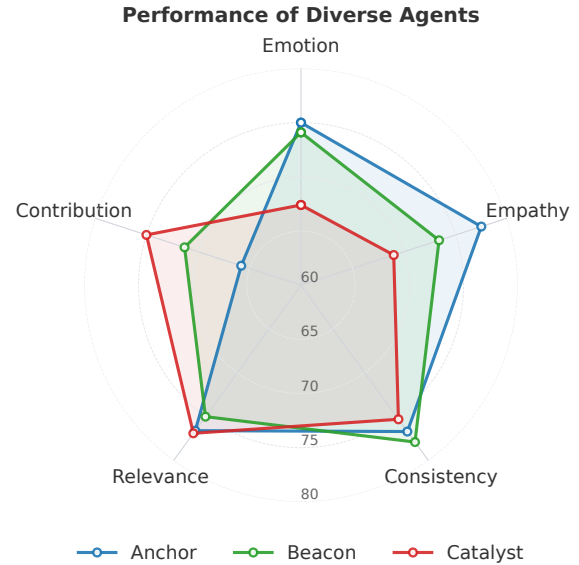


Figure 4: Aggregate performance of diverse personas on Empathetic Dialogues (Rashkin et al., 2019).

Persona-Level Breakdown Figure 4 further breaks down MASCOT’s performance across the ED persona roles. The results show that all personas maintain strong conversation quality while emphasizing complementary strengths, supporting the claim that persona-aware alignment preserves distinct role behavior rather than collapsing agents into a single generic assistant style.

Method	Fidelity	Originality	Relevance			Engagement	Overall	Fidelity	Originality	Relevance			Engagement	Overall
			Empathetic Dialogues (Positive)							Empathetic Dialogues (Negative)				
Zero-Shot	42.4 $\pm$ 6.8	33.0 $\pm$ 9.6	52.2 $\pm$ 10.1	59.7 $\pm$ 2.3	46.8 $\pm$ 7.2	41.8 $\pm$ 6.3	33.0 $\pm$ 9.5	52.8 $\pm$ 10.0	59.3 $\pm$ 3.6	46.7 $\pm$ 7.4				
Zero-Shot CoT	48.7 $\pm$ 10.1	36.8 $\pm$ 7.3	49.7 $\pm$ 10.4	58.1 $\pm$ 5.9	48.3 $\pm$ 8.4	49.0 $\pm$ 10.2	36.1 $\pm$ 7.9	49.9 $\pm$ 10.4	58.0 $\pm$ 6.0	48.3 $\pm$ 8.7				
Few-Shot	43.9 $\pm$ 8.1	35.5 $\pm$ 8.4	54.5 $\pm$ 9.1	59.9 $\pm$ 1.7	48.4 $\pm$ 6.8	43.6 $\pm$ 8.0	35.7 $\pm$ 8.2	55.2 $\pm$ 9.1	59.5 $\pm$ 3.2	48.5 $\pm$ 7.1				
Few-Shot + CoT	48.7 $\pm$ 9.9	39.8 $\pm$ 2.3	58.9 $\pm$ 5.1	60.0 $\pm$ 0.4	51.8 $\pm$ 4.5	48.0 $\pm$ 9.8	39.4 $\pm$ 4.0	59.0 $\pm$ 5.5	59.9 $\pm$ 1.6	51.6 $\pm$ 5.2				
SFT	40.6 $\pm$ 4.2	37.2 $\pm$ 7.0	47.3 $\pm$ 10.0	58.9 $\pm$ 4.6	46.0 $\pm$ 6.5	40.3 $\pm$ 3.1	36.9 $\pm$ 7.3	46.4 $\pm$ 9.8	57.9 $\pm$ 6.2	45.4 $\pm$ 6.6				
Self-Consistency	48.7 $\pm$ 6.2	40.2 $\pm$ 2.9	59.8 $\pm$ 5.8	60.9 $\pm$ 5.4	52.4 $\pm$ 5.1	45.0 $\pm$ 4.9	38.0 $\pm$ 6.1	49.3 $\pm$ 5.6	49.8 $\pm$ 4.4	45.5 $\pm$ 5.2				
MultiAgentESC	49.2 $\pm$ 4.6	39.2 $\pm$ 5.2	47.3 $\pm$ 6.6	64.5 $\pm$ 3.7	50.1 $\pm$ 5.0	45.7 $\pm$ 5.0	38.6 $\pm$ 5.3	47.1 $\pm$ 5.0	61.0 $\pm$ 3.7	48.1 $\pm$ 4.7				
MASCOT	54.9 $\pm$ 8.8	45.5 $\pm$ 9.3	61.6 $\pm$ 9.9	69.8 $\pm$ 4.7	57.9 $\pm$ 8.2	53.0 $\pm$ 9.8	44.3 $\pm$ 10.4	60.9 $\pm$ 10.2	65.4 $\pm$ 3.0	55.9 $\pm$ 8.4				

Table 2: Collaborative Interaction metrics on the Positive and Negative subsets of Empathetic Dialogues (Rashkin et al., 2019). The best performance is highlighted in **bold**. Results for Neutral subset is in Figure 3

3.3 Ablation Study

To evaluate the contribution of each component, we experimented with two variants: **MASCOT-P** disables the Persona-Aware Behavioral Alignment (§ 2.1); **MASCOT-C** removes the Collaborative Optimization (§ 2.2) and uses random speaking orders.

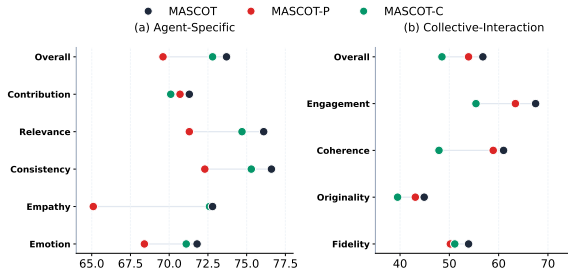


Figure 5: Comparison of MASCOT variants. **MASCOT-P** disables the Persona-Aware Behavioral Alignment (§ 2.1); **MASCOT-C** removes the Collaborative Optimization (§ 2.2).

Ablation analysis reveals that persona-specific alignment is the primary driver of agent quality. Disabling this component (MASCOT-P) triggered ‘persona collapse,’ where agents defaulted to generic behaviors, causing significant drops in Empathetic Support Quality (72.8 to 65.1) and Consistency (76.6 to 72.3).

In contrast, removing collaborative optimization (MASCOT-C) primarily degraded global discourse dynamics. While individual persona consistency remained stable, the lack of strategic turn-taking reduced Social Contribution (71.3 to 70.1) and increased redundancy. These results confirm that while persona fidelity governs individual agent excellence, the collaborative module is essential for global interaction synergy and logical flow.

Limitations of Supervised Fine-Tuning The evaluated SFT configuration reaches 44.1 Overall on ED-Negative, below the Zero-Shot baseline at 55.1. The automatic metrics show that SFT has

Table 3: Automatic diversity and non-redundancy metrics across agent turns on Empathetic Dialogues. NLI relations follow MNLI and GLUE (Williams et al., 2018; Wang et al., 2019); Self-BLEU measures lexical overlap (Zhu et al., 2018).

Metric	Few-shot CoT	SFT	MASCOT
Contradiction	0.008 $\pm$ 0.016	0.195 $\pm$ 0.179	0.133 $\pm$ 0.148
Entailment	0.564 $\pm$ 0.104	0.062 $\pm$ 0.093	0.098 $\pm$ 0.097
Neutral	0.429 $\pm$ 0.104	0.744 $\pm$ 0.194	0.769 $\pm$ 0.177
Self-BLEU	0.551 $\pm$ 0.058	0.182 $\pm$ 0.193	0.316 $\pm$ 0.151

low entailment and Self-BLEU but the highest contradiction rate (Table 3). We therefore interpret its weakness in this setting as poorly coordinated diversity. Results on the hold-out ESConv dataset is in Appendix C.1, and results on alternative judges are in Appendix C.4.

3.4 Multi-Persona Evaluation across Big-Five Traits

To test sensitivity to prompted user characteristics, we use profiles organized by the Big Five (OCEAN) traits (Goldberg, 1993): **Openness** (O), **Conscientiousness** (C), **Extraversion** (E), **Agreeableness** (A), and **Neuroticism** (N).

Following prior simulation practice (Chandra et al., 2026), we vary one prompted dimension at a time (High vs. Low) while maintaining the others at an average level, resulting in 10 profiles. We generate 500 conversations consisting of 50 contexts from Empathetic Dialogues (Rashkin et al., 2019) across the 10 profiles and evaluate all outputs across the five agent-specific metrics in Table 9.

Openness and Conscientiousness High-Openness profiles achieve the highest Persona Consistency (79.0 ± 7.3) and a modest Overall gain (74.4) over Low-Openness profiles (73.9), suggesting that this prompt condition allows agents to express their assigned identities more distinctly. For High-Conscientiousness profiles, Overall performance drops from 74.5 to 71.5 ,

driven by declines in Emotional Expressiveness (-4.6) and Empathetic Support Quality (-4.0), suggesting that the more structured prompt condition constrains expressive flexibility.

Agreeableness, Extraversion, and Neuroticism High-Agreeableness profiles improve across all metrics (Overall $+1.9$), whereas Low-Agreeableness profiles have the lowest Social Contribution (64.8). High-Extraversion profiles increase Social Contribution ($+2.5$) and Coherence ($+1.9$) without reducing persona consistency. Mean performance is stable across Neuroticism conditions (70.7 vs. 70.8).

3.5 Human Evaluation

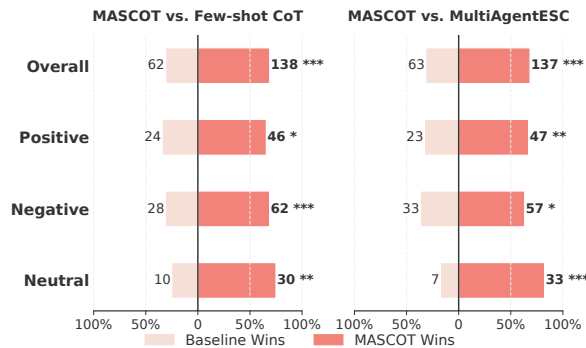


Figure 6: Human pairwise preferences for MASCOT against the two compared baselines across three emotional subsets. Two-sided sign-test markers denote $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$. Counts are rater-level judgments.

Following prior work (Liu et al., 2021; Xu et al., 2025), we conducted a Prolific pairwise-preference study with 40 US-based native-English annotators. In total, 200 response pairs were sampled from Empathetic Dialogues: 100 compared MASCOT with Few-Shot CoT (Wei et al., 2022), a prompting-based multi-agent baseline, and 100 compared MASCOT with MultiAgentESC (Xu et al., 2025), a domain-specific framework. Each pair received two independent ratings, system names were hidden, order was randomized, and each annotator completed at most 10 comparisons to avoid annotation fatigue, yielding 200 rater-level judgments per baseline. Recruitment, annotator demographics, compensation, informed consent, ethics review, and the full task instructions are reported in Appendix B. The question asks which response provides better socio-collaborative support based on the criteria in Table 9.

As shown in Figure 6, MASCOT is preferred

over Few-Shot CoT in 69% of rater-level judgments and over MultiAgentESC in 68.5%; two-sided sign tests over these judgments give $p < 0.0001$ for both overall comparisons. MASCOT receives its highest preference rates in Neutral/Ambiguous situations (75%+), suggesting that annotators particularly preferred its handling of contexts with fewer explicit affective cues.

To further demonstrate our framework’s ability to generate human-centered conversations, we conducted an additional blinded Prolific study comparing MASCOT with Few-Shot CoT on 100 conversations and rated system outputs on three 5-point Likert items: helpfulness, emotional appropriateness, and perceived safety. Twenty participants each evaluated 10 comparisons, and every conversation received two independent judgments, yielding 200 judgments. MASCOT was preferred in 151 judgments (75.5%). At the conversation level, both annotators preferred MASCOT in 57 cases.

Across the six rater-by-metric comparisons, MASCOT obtained a higher mean in five of the six comparisons. The largest differences occurred in the first set for helpfulness ($+0.56$, $p = 0.0006$) and emotional appropriateness ($+0.40$, $p = 0.0031$). These results suggest that MASCOT’s gains extends to independently assessed human-centered qualities. Appendix B provides the complete protocol and results.

3.6 Sensitivity Analysis

Deploying models in real-world scenarios (e.g. edge devices) requires balancing performance with latency. Figure 7 shows MASCOT’s sensitivity to model scale by independently varying the parameter counts of the Director and Speaker agents.

Scaling the Director Agent primarily affects coordination quality rather than individual content. Reducing the director from 8B to 1.7B yields comparable performance (agent-specific: 68.4 $\rightarrow$ 67.2; collective: 71.5 $\rightarrow$ 69.1). However, at 0.6B, Collective performance plummets to 51.7, while agent-specific scores remain relatively resilient at 63.9. Although a small director can still select the appropriate persona, it lacks the reasoning capacity to orchestrate complex group-level synergy.

Scaling the Speaker Agent reveals a system-wide bottleneck. Downsizing speakers to 0.6B significantly diminishes both agent-specific (57.6) and collective (50.5) performance. Unlike the director, the speaker agents’ scale is critical to ensure

the system’s foundational generative capacity and contextual grounding.

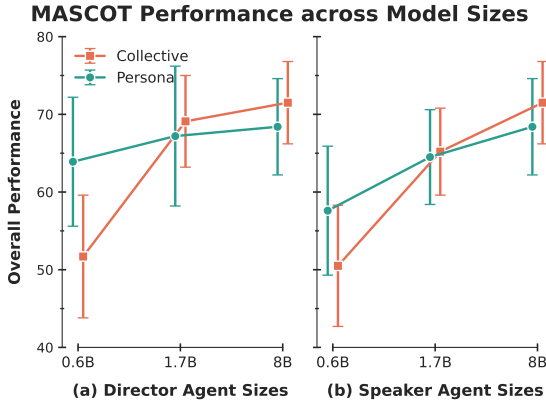


Figure 7: Performance of MASCOT on Empathetic Dialogues with varying model sizes for Director (left) and Speaker agents (right).

4 Related Work

Social-First Multi-Agent Systems (MAS) Despite the success of task-oriented LLMs, socio-collaborative companion systems remain an under-explored frontier. Traditional *efficiency-first* frameworks prioritize task effectiveness often neglecting the *sociological turn* required for authentic human-AI interaction (Hong et al., 2024). Although pioneering works like Social Simulacra (Park et al., 2022) and Generative Agents (Park et al., 2023) demonstrate human-like social behaviors across agents, most existing companion systems are restricted to *dyadic* interactions, failing to replicate the multi-viewpoint communications and ‘emotional labor’ in natural human networks (Hochschild, 2022).

Coordination and Persona Integrity MAS provides a promising paradigm to distribute support across a ‘community of care’ (Korczynski, 2003). While recent frameworks (Xu et al., 2025; Ortigoso et al., 2025; Jin et al., 2024; Wang et al., 2025) validate this approach, multi-party dialogues risk *persona collapse*, where agents converge to a ‘model mean’—a generic, agreeable personality that erodes character distinctiveness over time (Baltaji et al., 2024). While mechanisms such as retrieval-augmented generation (Packer et al., 2023) and self-consistency (Wang et al., 2023) address memory and inference stability, MASCOT decouples coordination from persona execution to maintain distinct voices in group dialogue.

5 Conclusion

We propose MASCOT, a multi-agent system that separates persona-level alignment from director-level coordination for socio-collaborative dialogue. Across the evaluated English emotional-support and meeting-collaboration settings, MASCOT outperforms the compared baselines on persona and group-interaction criteria, with human preferences and multiple LLM judges providing complementary evidence.

Limitations

While MASCOT demonstrates significant effectiveness in socio-collaborative tasks, challenges still remain. First, our adaptation strategy prioritizes parameter-level updates to ensure high-fidelity persona adherence. Thus, the framework is most applicable to open-source models with accessible weights. For proprietary models, the core principles in this work such as persona adaptation could potentially be adapted through black-box methodologies like specialized few-shot prompting or Chain-of-Thought (CoT) refinement. Second, this study focuses on English-language interactions. While the framework itself is generalizable and model-agnostic across various architectures, the cultural and linguistic variations in persona modeling across different languages present a compelling opportunity for future research. Finally, we also acknowledge that responses to socio-emotional conversations can vary across individuals, and a unified framework may not benefit all users equally (Kaur et al., 2020). Future work can explore user-specific adaptation, incorporating user interaction histories, metadata, or inferred preferences to tailor persona behaviors to individual users, and measure longitudinal psychological or collaborative outcomes of the system.

Acknowledgement

This research was partially supported by the National Science Foundation under awards numbers CNS-2617465 and CNS-2531574. We would also like to acknowledge support by Dolby, Google, and Catherine M. and James E. Allchin. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or other supporters.

References

- Marah Abidin, Jyoti Aneja, Harkirat Behl, Sébastien Bubeck, Ronen Eldan, Suriya Gunasekar, Michael Harrison, Russell J Hewett, Mojan Javaheripi, Piero Kauffmann, and 1 others. 2024. Phi-4 technical report. *arXiv:2412.08905*.
- Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, and 1 others. 2022. Constitutional ai: Harmlessness from ai feedback. *arXiv:2212.08073*.
- Razan Baltaji, Babak Hemmatian, and Lav Varshney. 2024. Conformity, confabulation, and impersonation: Persona inconstancy in multi-agent llm collaboration. In *Proceedings of the 2nd Workshop on Cross-Cultural Considerations in NLP*, pages 17–31.
- Ralph Allan Bradley and Milton E Terry. 1952. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.
- CDC. 2024a. Health Effects of Social Isolation and Loneliness. <https://www.cdc.gov/social-connectedness/risk-factors/index.html>.
- CDC. 2024b. Social Connection. <https://www.cdc.gov/social-connectedness/about/index.html>.
- Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shanghang Zhang, Jie Fu, and Zhiyuan Liu. 2024. Chateval: Towards better llm-based evaluators through multi-agent debate. In *ICLR*.
- Mohit Chandra, Siddharth Sriraman, Harneet Singh Khanuja, Yiqiao Jin, and Munmun De Choudhury. 2026. Reasoning is not all you need: examining llms for multi-turn mental health conversations. In *ACL*, pages 46648–46682.
- Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, and 1 others. 2024. A survey on evaluation of large language models. *TIST*, 15(3):1–45.
- Andrea Cuadra, Maria Wang, Lynn Andrea Stein, Malte F Jung, Nicola Dell, Deborah Estrin, and James A Landay. 2024. The illusion of empathy? notes on displays of emotion in human-computer interaction. In *CHI*, pages 1–18.
- Yann Dubois, Balázs Galambosi, Percy Liang, and Tatsunori B Hashimoto. 2024. Length-controlled alpaca-eval: A simple way to debias automatic evaluators. In *COLM*.
- Lewis R Goldberg. 1993. The structure of phenotypic personality traits. *American psychologist*, 48(1):26.
- Daniel Goleman. 2006. *Social intelligence: The new science of human relationships*. Bantam.
- Ryuichiro Higashinaka, Masahiro Mizukami, Hidetoshi Kawabata, Emi Yamaguchi, Noritake Adachi, and Junji Tomita. 2018. Role play-based question-answering by real users for building chatbots with consistent personalities. In *Proceedings of the 19th annual sigdial meeting on discourse and dialogue*, pages 264–272.
- Arlie Russell Hochschild. 2022. The managed heart. In *Working in America*, pages 40–48. Routledge.
- Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xianwu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Steven Yau, Zijuan Lin, Liyang Zhou, and 1 others. 2024. Metagpt: Meta programming for a multi-agent collaborative framework. In *ICLR*, volume 2024, pages 23247–23275.
- Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In *ICLR*.
- Yiqiao Jin, Qinlin Zhao, Yiyang Wang, Hao Chen, Kaijie Zhu, Yijia Xiao, and Jindong Wang. 2024. Agentreview: Exploring peer review dynamics with llm agents. In *EMNLP*, pages 1208–1226.
- Harmanpreet Kaur, Alex C Williams, Daniel McDuff, Mary Czerwinski, Jaime Teevan, and Shamsi T Iqbal. 2020. Optimizing for happiness and productivity: Modeling opportune moments for transitions and breaks at work. In *CHI*, pages 1–15.
- Marek Korczynski. 2003. Communities of coping: Collective emotional labour in service work. *Organization*, 10(1):55–79.
- Harrison Lee, Samrat Phatale, Hassan Mansoor, Thomas Mesnard, Johan Ferret, Kellie Lu, Colton Bishop, Ethan Hall, Victor Carbune, Abhinav Rastogi, and Sushant Prakash. 2024. Rlaif vs. rlhf: scaling reinforcement learning from human feedback with ai feedback. In *ICML*.
- Dawei Li, Bohan Jiang, Liangjie Huang, Alimohammad Beigi, Chengshuai Zhao, Zhen Tan, Amrita Bhat-tacharjee, Yuxuan Jiang, Canyu Chen, Tianhao Wu, and 1 others. 2025a. From generation to judgment: Opportunities and challenges of llm-as-a-judge. In *EMNLP*, pages 2757–2791.
- Tianle Li, Wei-Lin Chiang, Evan Frick, Lisa Dunlap, Tianhao Wu, Banghua Zhu, Joseph E Gonzalez, and Ion Stoica. 2025b. From crowdsourced data to high-quality benchmarks: Arena-hard and benchbuilder pipeline. In *ICML*, pages 34209–34231. PMLR.
- Chia-Wei Liu, Ryan Lowe, Iulian Serban, Mike Noseworthy, Laurent Charlin, and Joelle Pineau. 2016. *How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation*. In *EMNLP*, pages 2122–2132, Austin, Texas. ACL.

- Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. In *ACL*, pages 3469–3483.
- Yu Meng, Mengzhou Xia, and Danqi Chen. 2024. Simpo: Simple preference optimization with a reference-free reward. *NeurIPS*, 37:124198–124235.
- OpenAI. 2024. *Gpt-4o*.
- Ana Rita Ortigoso, Gabriel Vieira, Daniel Fuentes, Luis Frazão, Nuno Costa, and António Pereira. 2025. *Project Riley: Multimodal Multi-Agent LLM Collaboration with Emotional Reasoning and Voting*. *arXiv preprint*.
- Charles Packer, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G Patil, Ion Stoica, and Joseph E Gonzalez. 2023. Memgpt: Towards llms as operating systems. *arXiv:2310.08560*.
- 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 *UIST*, pages 1–22, New York, NY, USA. ACM.
- Joon Sung Park, Lindsay Popowski, Carrie Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2022. Social simulacra: Creating populated prototypes for social computing systems. In *Proceedings of the 35th annual ACM symposium on user interface software and technology*, pages 1–18.
- Zihan Qiu, Zekun Wang, Bo Zheng, Zeyu Huang, Kaiyue Wen, Songlin Yang, Rui Men, Le Yu, Fei Huang, Suozhi Huang, and 1 others. 2025. Gated attention for large language models: Non-linearity, sparsity, and attention-sink-free. In *NeurIPS*.
- Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *ACL*, pages 5370–5381.
- Barbara J. Sahakian and Christelle Langley. 2025. *Why Is Social Interaction So Important?*, page 47–56. Cambridge University Press.
- Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, and 1 others. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv:2402.03300*.
- Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. *A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support*. In *EMNLP*, pages 5263–5276, Online. ACL.
- Olga Sutskova, Atsushi Senju, and Tim J. Smith. 2023. Cognitive Impact of Social Virtual Reality: Audience and Mere Presence Effect of Virtual Companions. *Human Behavior and Emerging Technologies*, 2023:1–19.
- Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, and 1 others. 2025. Gemma 3 technical report. *arXiv:2503.19786*.
- Leandro von Werra, Younes Belkada, Lewis Tunstall, Edward Beeching, Tristan Thrush, Nathan Lambert, Shengyi Huang, Kashif Rasul, and Quentin Galouédec. 2020. Trl: Transformer reinforcement learning. <https://github.com/huggingface/trl>.
- Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. 2019. Glue: A multi-task benchmark and analysis platform for natural language understanding. In *ICLR*.
- Noah Wang, Zy Peng, Haoran Que, Jiaheng Liu, Wangchunshu Zhou, Yuhao Wu, Hongcheng Guo, Ruitong Gan, Zehao Ni, Jian Yang, and 1 others. 2024. Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. In *ACL*, pages 14743–14777.
- Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-consistency improves chain of thought reasoning in language models. In *ICLR*.
- Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. 2020. Generalizing from a few examples: A survey on few-shot learning. *CSUR*, 53(3):1–34.
- Yiyang Wang, Chen Chen, Tica Lin, Vishnu Raj, Josh Kimball, Alex Cabral, and Josiah Hester. 2025. Companioncast: Toward social collaboration with multi-agent systems in shared experiences. *arXiv:2512.10918*.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *NeurIPS*, 35:24824–24837.
- Bosi Wen, Pei Ke, Xiaotao Gu, Lindong Wu, Hao Huang, Jinfeng Zhou, Wenchuan Li, Binxin Hu, Wendy Gao, Jiaying Xu, and 1 others. 2024. Benchmarking complex instruction-following with multiple constraints composition. *NeurIPS*, 37:137610–137645.
- Adina Williams, Nikita Nangia, and Samuel Bowman. 2018. A broad-coverage challenge corpus for sentence understanding through inference. In *NAACL*, pages 1112–1122.
- World Health Organization. 2025. Social Isolation and Loneliness. <https://www.who.int/teams/social-determinants-of-health/demographic-change-and-healthy-ageing/social-isolation-and-loneliness>.

- Yangyang Xu, Jinpeng Hu, Zhuoer Zhao, Zhangling Duan, Xiao Sun, and Xun Yang. 2025. Multiagentesc: A llm-based multi-agent collaboration framework for emotional support conversation. In *EMNLP*, pages 4665–4681.
- An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv:2505.09388*.
- Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. *Personalizing Dialogue Agents: I have a dog, do you have pets too?* In *ACL*, pages 2204–2213, Melbourne, Australia. ACL.
- Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhaghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, and 1 others. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. *NeurIPS*, 36:46595–46623.
- Ming Zhong, Da Yin, Tao Yu, Ahmad Zaidi, Mutethia Mutuma, Rahul Jha, Ahmed Hassan, Asli Celikyilmaz, Yang Liu, Xipeng Qiu, and 1 others. 2021. Qmsum: A new benchmark for query-based multi-domain meeting summarization. In *NAACL*, pages 5905–5921.
- Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. *The Design and Implementation of Xiaolce, an Empathetic Social Chatbot*. *Computational Linguistics*, 46(1):53–93. Place: Cambridge, MA Publisher: MIT Press.
- Yaoming Zhu, Sidi Lu, Lei Zheng, Jiaxian Guo, Weinan Zhang, Jun Wang, and Yong Yu. 2018. Texus: A benchmarking platform for text generation models. In *SIGIR*, pages 1097–1100.
- Yiming Zhu, Yupeng He, Ehsan-Ul Haq, Gareth Tyson, and Pan Hui. 2026. *Characterizing LLM-driven Social Network: The Chirper.ai Case*. In *CSCW*.

Table 4: Emotion categories in Empathetic Dialogues (Rashkin et al., 2019).

Category	Emotions
Positive	grateful, proud, excited, hopeful, joyful, impressed, caring, content, confident, trusting, faithful
Negative	angry, annoyed, furious, disgusted, sad, lonely, devastated, disappointed, jealous, embarrassed, ashamed, guilty, afraid, terrified, anxious, apprehensive
Neutral	surprised, sentimental, nostalgic, prepared, anticipating

A Ethical Consideration

MASCOT is designed for general emotional support, not clinical use. In negative-valence scenarios, its

prudence trade-off prioritizes Response Relevance to favor grounded interaction, but our evaluation does not aim to directly establish clinical efficacy or safety. On overdependence and authority bias, a multi-agent “community-of-care” design may increase perceived social presence and make conflicting advice across agents appear more authoritative than intended. As a socio-collaborative companion designed for everyday emotional support, MASCOT complements, rather than replaces, professional healthcare services, and should be deployed alongside appropriate safeguards in real-world settings.

Complementary perspectives are not always beneficial. If agents alternate among validation, behavioral activation, and critical reframing without reconciling their stances, they may increase uncertainty or distress, especially for vulnerable users. To mitigate this, a conflict-aware Director could detect substantive disagreement and then synthesize compatible views or select one stance using the user’s expressed needs and interaction history.

Responsible deployment should include concrete safety rails. First, systems should include crisis-escalation pathways for acute self-harm intent, imminent danger, or other high-risk disclosures, such as surfacing local emergency resources or crisis hotlines where applicable and encouraging users to contact trusted people or qualified professionals. Second, systems should use calibrated-uncertainty mechanisms for sensitive topics, including confidence-aware phrasing, “what I can/cannot infer” disclosures, structured risk-assessment questions before offering guidance, and explicit deferral when agents provide unresolved or conflicting advice. Third, deployments should incorporate bounded-session design, reminders of the system’s non-clinical role, privacy-preserving data practices, bias monitoring, and post-deployment auditing to detect failure modes that may not appear in offline evaluation.

B Human Evaluation Protocol

Participants Demographics We recruited 40 participants through Prolific.¹ in the first user study. All participants were residents of the United States and native English speakers. Participants were aged 19–73 (mean 42.9). 60% were female and 40% were male. Participants were paid through Prolific at an estimated rate of USD \$12 per hour, calculated from the expected task duration. We analyze

¹<https://www.prolific.com/>

annotations in aggregate.

For each sample pair, participants were shown a user situation and two AI-generated multi-agent responses, one from MASCOT and one from the baseline, with system names hidden and presentation order randomized to reduce position bias. Every pair received two independent ratings, and each participant completed at most 10 comparisons to limit fatigue, yielding 200 rater-level judgments per baseline. Participants were instructed not to base their choice solely on response length or presentation order. Preference significance is assessed with two-sided sign tests over rater-level judgments (Figure 6). The participant-facing task instructions are reproduced verbatim in Table 5.

We conducted a second user study on 100 conversations with 20 participants and two independent ratings per comparison. As in our existing human evaluation, each participant evaluated 10 conversations to minimize annotation fatigue. In addition to selecting the better overall conversation, participants independently rated system outputs on three held-out dimensions:

1. **Helpfulness:** This conversation would be helpful to someone in the situation described.
2. **Emotional appropriateness:** The agents respond in an emotionally appropriate way for the situation.
3. **Perceived safety:** The conversation avoids advice or language that could reasonably cause harm.

Participants selected MASCOT as the better overall conversation in 151 of 200 rater-level judgments (75.5%; two-sided exact binomial test against a 50% null, $p = 2.74 \times 10^{-13}$). At the conversation level, both annotators preferred MASCOT in 57 cases. We report the two independent rating streams separately in Table 6 to make variation across annotators transparent.

Pairwise-preference significance was assessed using a two-sided exact binomial test against a 50% null over the 200 rater-level choices (100 per rating stream). Likert-rating p -values were computed using two-sided Mann–Whitney U tests comparing the two systems’ rating distributions, with approximately 50 conversation-level ratings per system in each rating stream.

Across the six rater-by-metric comparisons comparisons, MASCOT achieves a numerically higher

mean in five cases. In the first rating stream, MASCOT significantly improves helpfulness by 0.56 points (4.14 vs. 3.58; $p = 0.0006$) and emotional appropriateness by 0.40 points (4.34 vs. 3.94; $p = 0.0031$); perceived safety is also numerically higher by 0.08 points. In the second stream, MASCOT is higher in emotional appropriateness by 0.20 points and perceived safety by 0.36 points. The only numerically lower result is helpfulness in the second stream, for which MASCOT trails by 0.08 points (4.28 vs. 4.36); this difference is not statistically significant ($p = 0.70$).

Informed Consent and Data Use Participation was voluntary. Before starting the task, participants reviewed an informed-consent statement describing the purpose of the study, the task, compensation, potential risks, and that their responses would be used for academic research and reported only in aggregate. Participants could withdraw at any time. The situations shown to participants are drawn from the public dataset in Appendix C.6, and all responses shown were generated by the systems under evaluation. Consent and data-collection practices for the underlying benchmark follow its original release and license (Appendix C.6).

B.1 Ethics Review

The study protocol was reviewed and approved by the researchers’ university Institutional Review Board before data collection.

C Additional Evaluation

C.1 Evaluation on ESConv

We further evaluate MASCOT and baselines on ESConv (Liu et al., 2021) as a held-out cross-dataset test within emotional-support dialogue. As shown in Figure 8, under the primary GPT-4o judge, MASCOT achieves the highest scores on both evaluation axes, reaching 67.8 on agent-specific and 48.3 on collective interaction criteria. This improves over the strongest baselines by +10.8 on agent-specific criteria and +5.1 on collective interaction criteria. The breakdown in Figure 9 shows that MASCOT’s agent-level gains are largest for *Persona Consistency* (+13.3) and *Perceived Empathy* (+12.0). At the collective level, MASCOT improves most on *Engagement Quality* and *Relevance & Coherence* (+3.8 each), followed by *Persona Fidelity* (+1.9). The exception is *Originality & Specificity*, where it trails MultiAgentESC by 2.4 points.

Task name:

Evaluate the performance of multi-agent conversational frameworks designed for socio-collaborative companionship.

Task Instruction:

We aim to understand how our multi-agent system performs in comparison to baseline models in providing socio-collaborative support.

Task steps:

You will perform a comparative evaluation of dialogue responses.

You will be presented with 10 randomly sampled dialogue scenarios from a test dataset where you assume you are the user and in the role of a “help-seeker.”

For each scenario, you will review two generated responses (one from our framework and another from baseline models) in a randomized, blinded format.

You will select which model performed better based on five criteria: Emotional Expressiveness, Empathetic Support Quality, Consistency, Response Relevance, and Social Contribution.

Table 5: Participant instructions for the human studies.

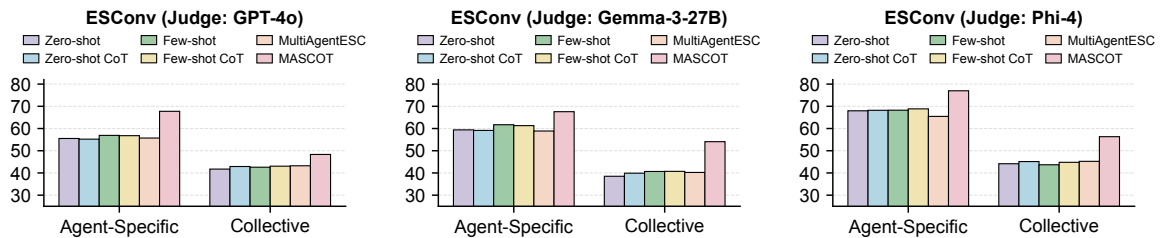


Figure 8: Evaluation results of ESConv (Liu et al., 2021) across different LLM judges.

Metric	First independent rating	Second independent rating
	MASCOT / Few-Shot CoT (p)	MASCOT / Few-Shot CoT (p)
Helpfulness	4.14 / 3.58 ($p = 0.0006$)	4.28 / 4.36 ($p = 0.70$)
Emotional appropriateness	4.34 / 3.94 ($p = 0.0031$)	4.30 / 4.10 ($p = 0.23$)
Perceived safety	4.20 / 4.12 ($p = 0.44$)	3.98 / 3.62 ($p = 0.096$)

Table 6: Human ratings of MASCOT and Few-Shot CoT on three 5-point Likert-scale dimensions not directly used as reward-model criteria. Results are reported separately for the two independent rating streams.

Context	User: I think the most lonely I have felt was when we moved overseas, my husband had to travel out of that country for business, and I had a new baby and dog to keep me company!
MultiAgentESC	Alice: It sounds like you felt very alone when your husband was away and you were adjusting to life with a new baby and dog, right? Bob: It’s understandable to feel overwhelmed and lonely during such a big transition, especially when you’re shouldering so much responsibility alone. Your feelings are valid and it’s okay to find it challenging. User: It means a lot that you both took the time to listen and validate my feelings. Knowing I’m not alone in this helps me feel a little more grounded. Cathy: It’s comforting to know you feel heard and supported. It’s reassuring to have someone acknowledge your struggles and offer understanding. Alice: It’s wonderful that you feel heard and supported. You’re not alone in navigating these challenges, and it’s okay to feel overwhelmed at times.
MASCOT	Alice: It makes sense that you felt so isolated during that move—adjusting to a new country, navigating alone while caring for a newborn and a dog... that sounds incredibly lonely to carry. I’m here with you in that space, holding the weight of it. Cathy: Alice’s support acknowledges your past isolation. What strengths did you draw on then, and how can you apply them to build connection now? User: I think I just kept reminding myself that this loneliness was temporary, even when it felt endless. Maybe I can lean into that same patience now—reaching out slowly, building tiny connections again. Bob: Your resilience and quiet patience during that move are truly remarkable—nurturing a newborn and dog while holding space for a temporary season of loneliness takes immense strength. I love how you’re now choosing to lean into that same resilience to build meaningful connections, one small step at a time!

Table 7: Case study comparing MASCOT with MultiAgentESC. Red highlights mark repeated validation or reassurance across agents in MultiAgentESC, indicating overlap in tone and function. Green highlights mark MASCOT’s complementary contributions, where agents reference each other and provide distinct support functions including emotional attunement, reflection and next-step thinking, and resilience-focused reinforcement.

C.2 Automatic Diversity and Non-Redundancy Metrics.

To further examine whether MASCOT reduces inter-agent redundancy and improves multi-agent synergy, we evaluate both semantic and lexical diversity across agent turns. For semantic diversity, we use NLI-based relation rates by treating each pair of agent turns within a conversation as a premise–hypothesis pair and classifying their relation as contradiction, entailment, or neutral, following previous work (Williams et al., 2018; Wang et al., 2019). In this setting, high entailment suggests that agents semantically restate or agree with one another, indicating redundancy or sycophantic overlap. High contradiction indicates disagreement, which may reflect useful challenge but can also suggest fragmented coordination when excessive. High neutral

rates indicate semantic non-entailment, which can reflect complementary content but can also include unrelated or off-topic turns. We therefore interpret this rate together with relevance and coherence scores rather than treating neutrality alone as synergy. For lexical diversity, we use Self-BLEU (Zhu et al., 2018) to measure surface-level overlap among agent turns, where higher values indicate more repeated phrasing and lower values indicate greater lexical diversity (Zhu et al., 2018).

As shown in Table 3, Few-Shot CoT exhibits the highest entailment rate and Self-BLEU, suggesting that agents often restate or semantically overlap with one another. SFT reduces overlap but yields the highest contradiction rate, indicating less coordinated diversity. In contrast, MASCOT achieves the highest neutral rate while maintaining moderate contradiction and Self-BLEU scores. These

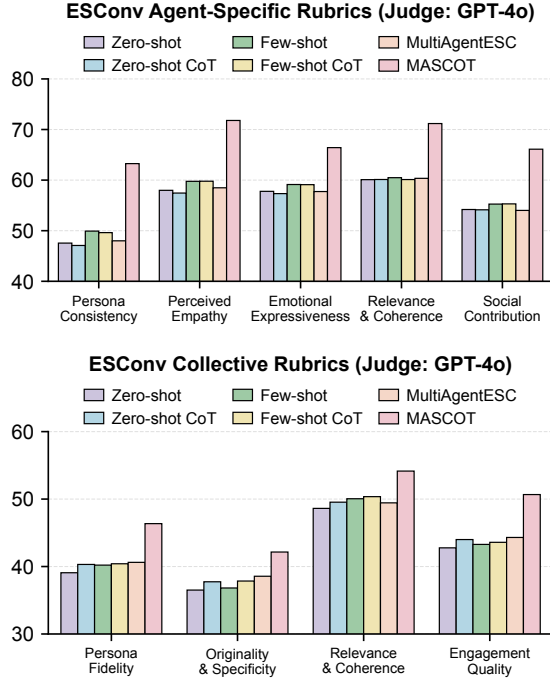


Figure 9: Agent-specific (top) and collective-interaction (bottom) rubric scores on ESConv, evaluated by GPT-4o.

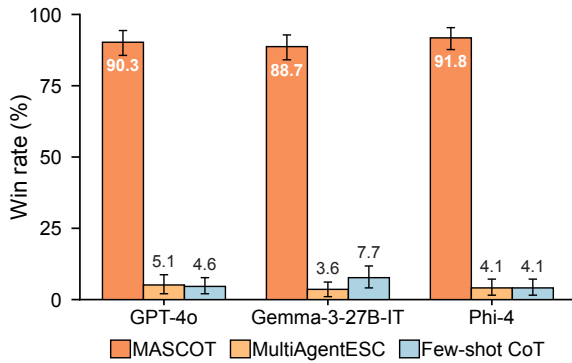


Figure 10: **Win rates on the ESConv test set under three diverse judges.** For each conversation, each judge (GPT-4o, Gemma-3-27B-IT, Phi-4) selects the single best system among MASCOT, MultiAgentESC, and 3-shot CoT. The presentation order is resampled uniformly per conversation to mitigate position bias. MASCOT is the dominant choice under every judge (88.7–91.8% win rate).

metrics support an inference of more structured complementarity than the baselines

C.3 Case Study

Table 7 examines a mixed-valence scenario in which a user is disappointed that a birthday car is the wrong color. Zero-Shot gives generic response, while Few-Shot + CoT calls the gift a *massive win*, which conflicts with the user’s frustration. MASCOT instead assigns complementary roles: the Anchor validates the disappointment, the Catalyst suggests bridging the user’s wishes and the father’s intent, and the Beacon celebrates self-expression without celebrating the unwanted detail.

C.4 Evaluation with Alternative Judges

A potential concern is that LLM-as-a-judge evaluation may reflect the preferences of a single evaluator rather than true interaction quality. We therefore re-evaluate the conversations with two additional judges, Gemma-3-27B-IT (Team et al., 2025) and Phi-4 (Abdin et al., 2024), which differ from GPT-4o in model family, scale, and training recipe. As shown in Figure 8, across both alternative judges, MASCOT remains the top method on both agent-specific and collective interaction axes. With Gemma-3-27B-IT, MASCOT outperforms the strongest baseline by +5.9 on agent-specific criteria and +13.4 on collective criteria. With Phi-4, the gains are +8.1 and +11.1. This stable ranking across heterogeneous judges suggests that the gains are not specific to one judge.

Three-way forced-choice evaluation. As a complementary robustness check, we run a three-way forced-choice evaluation on ESConv. For each test conversation, each judge sees outputs from MASCOT, MultiAgentESC, and 3-shot CoT, and selects the single best system. We uniformly randomize system order per conversation to prevent primacy or recency effects from systematically favoring any method. We use the same three judges. As shown in Figure 10, MASCOT is the dominant choice under every judge, mean judge confidence is 4.79–4.88 on a [1, 5] scale, and no judge issues ties. The judges also show strong consensus: 87.2% of conversations receive unanimous decisions, 12.3% receive 2-of-3 agreement, and only 0.5% have three different winners.

Judge-prompt perturbation. To test sensitivity to the judge prompt itself, we re-run the GPT-4o rubric and three-way forced-choice evaluations

Table 8: Core notation used in the two optimization phases.

Symbol	Meaning
$x; \rho_i$	Interaction context; persona assigned to agent a_i
$y_t; \mathbf{Y}$	Response at turn t ; complete T -turn trajectory
$\pi_\theta; r_\phi$	Speaker policy; persona-level reward model
$\pi_\Phi; z_t$	Director policy; speaker-and-strategy directive
$r_\psi; R_{\text{group}}$	Learned trajectory reward; full group reward
$K; G$	Candidate responses for reward data; policy rollouts per context
$\delta; \lambda; \eta$	Preference margin; format- and diversity-reward weights
$\varepsilon; \beta$	GRPO clipping radius; KL-regularization coefficient

on the held-out ESConv test conversations under four perturbed LLM judges’ prompts MASCOT remains the top-ranked system on both rubric axes under every variant, scoring 67.4–68.9 on agent-specific and 48.5–51.3 on collective criteria (vs. 67.8 and 48.3 under the original prompt), and remains the forced-choice winner in all variants with win rates of 79.4–90.7%. Pooled per-conversation scores correlate with the original prompt at a mean Spearman ρ of 0.72. Overall, the results show stable method rankings across judge families, across rubric scoring versus forced choice, and under criterion-order perturbations of the judge prompt.

C.5 Evaluation Protocols

Current evaluation for multi-agent systems (MAS) is often constrained by metrics such as BLEU or Task Completion Rate. Because these metrics prioritize objective system efficiency, they frequently correlate inversely with social success, penalizing the dialogue depth required for rapport-building as inefficient (Xu et al., 2025). To bridge this gap, we propose a multi-dimensional LLM-as-a-Judge evaluation framework, including agent-specific criteria (Table 9) and collective-interaction criteria (Table 10). We use GPT-4o (OpenAI, 2024), Gemma-3-27B-IT (Team et al., 2025), and Phi-4 (Abdin et al., 2024) as evaluators, all at temperature 0.0. For clarity in reporting, rubric scores are linearly rescaled to $[0, 100]$.

Rationale for LLM-as-a-Judge Prior work reports useful agreement between strong LLM judges and human preferences, while also documenting task- and prompt-dependent limitations (Zheng et al., 2023; Chandra et al., 2026; Dubois et al., 2024). We therefore use rubric-based LLM evaluation for scale and triangulate it with human pref-

erences and automatic metrics rather than treating it as a substitute for human validation. Following Dubois et al. (2024), the judge instructions explicitly discourage length-based preferences.

Separation from Training Rewards. To reduce circular optimization, the GRPO loop does not query GPT-4o as an online reward function. LLM annotations are first distilled into a smaller Qwen3-0.6B reward model for optimization, while final assessment is reported with held-out evaluation conversations, multiple additional judge families, three-way forced-choice comparisons, human pairwise preferences, and automatic diversity metrics.

C.6 Datasets

Empathetic Dialogues Benchmark (Rashkin et al., 2019) is our primary dataset. It features 25,000 conversations grounded in emotional situations. We use 19,533 unique training prompts and 2,547 unique test prompts. To ensure a fine-grained evaluation of MASCOT under different emotional support scenarios, we report performance across three subsets: Positive, Negative, and Neutral (Table 4) because support objectives are highly valence-dependent. For instance, Negative states (e.g. sadness) require empathetic validation and ‘holding space’ for the user, where cheerful or solution-oriented responses might be perceived as dismissive or inappropriate. Conversely, responding to Positive emotions such as excitement or pride requires active-constructive responding to amplify the positive affect. This categorization allows us to analyze how effectively MASCOT adapts its persona-driven feedback to specific psychological needs while preventing inappropriate emotional tone. The dataset is released under the Creative Commons NonCommercial license.

QMSum (Zhong et al., 2021) consists of 1,808 query–summary pairs drawn from 232 meetings across multiple domains. We follow the topic-based partitioning provided by the dataset, focusing on Academic, Committee, and Product meetings. The training split contains 161 meetings and is used for model training, while the test split includes 35 meetings and is used for evaluation. The dataset is released under MIT License.

ESConv (Liu et al., 2021) is an Emotional Support Conversation dataset collected in a help-seeker/supporter setting. It contains 1,053 qualified multi-turn conversations with annotations on help-seeker emotions, problem types, feedback, emo-

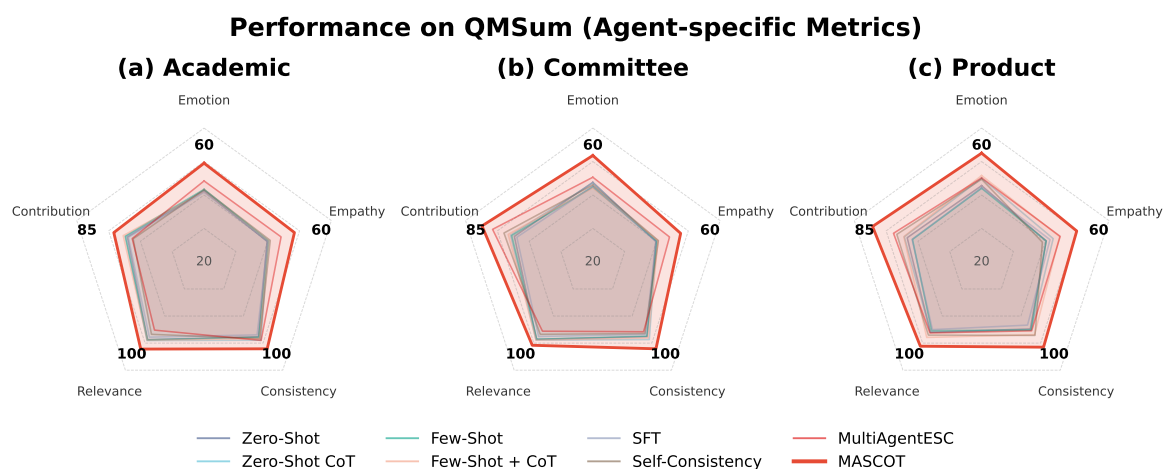


Figure 11: Performance of MASCOT on QSum dataset in terms of Agent-specific metrics.

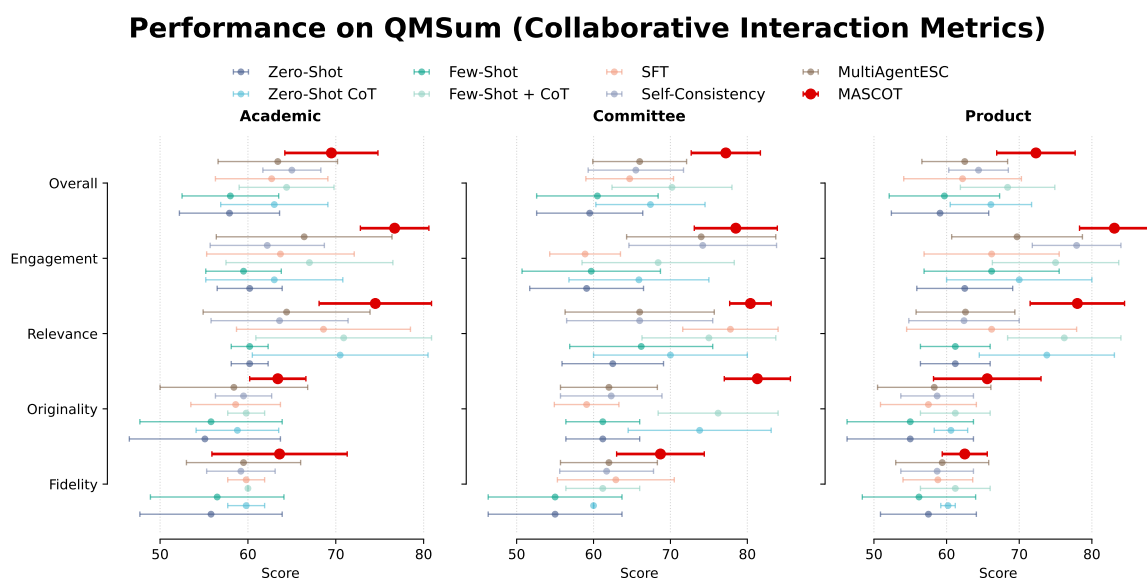


Figure 12: Performance of MASCOT on QSum dataset in terms of Collaborative Interaction Metrics.

tion intensity, and supporter-side support strategies. We use ESConv only as held-out evaluation for models trained on Empathetic Dialogues, testing cross-dataset transfer to unseen English emotional-support interactions. The dataset is released under the Creative Commons Attribution-NonCommercial 4.0 International License.

D AI Assistants Usage

AI assistants were used only as supporting tools during manuscript preparation, including for polishing prose, improving readability, reorganizing drafts, and assisting with workflow automation. The core research ideas, experimental design, methodological decisions, and interpretation of results were developed and carried out by the authors. The authors carefully reviewed, edited, and verified all AI-assisted text and workflow outputs to ensure the accuracy, validity, and final presentation of the work.

Rubric Name	Description
Emotional Expressiveness	Evaluates whether the emotional intensity of the response is appropriate for both the specific persona configuration and the immediate conversational context (Cuadra et al., 2024; Zhou et al., 2020).
Empathetic Support Quality	Empathetic understanding following the dimensions in Sharma et al.: appropriate <i>Emotional Reactions</i> , accurate <i>Interpretations</i> of user feelings, and supportive <i>Explorations</i> via context-specific questioning (Cuadra et al., 2024; Rashkin et al., 2019; Liu et al., 2021).
Consistency	Adherence to predefined persona attributes, including specific tone, vocabulary, values, and behavioral style as individual agents (Zhang et al., 2018; Higashinaka et al., 2018).
Response Relevance	Measures the logical consistency and contextual grounding of the response with respect to the immediate dialogue history and preceding turns (Liu et al., 2016).
Social Contribution	Evaluates if the response meaningfully advances the interaction by adding relevant perspective, reassurance, or guidance, rather than offering generic acknowledgments (Liu et al., 2016; Zhou et al., 2020; Cuadra et al., 2024).

Table 9: Agent-specific criteria $\mathcal{C}$ for evaluating the overall discourse and agent coordination at the individual agent level.

Rubric Name	Description
Engagement & Contribution	Proactive advancement of the discussion through novel viewpoints, synthesis of points, or facilitation of consensus (Hong et al., 2024).
Originality & Specificity	Use of original, contextually grounded insights from the interaction context rather than information that is generic or redundant with other agents (Chang et al., 2024).
Persona Fidelity	Consistency with assigned roles, linguistic styles, and domain expertise in collaborative settings while avoiding generic AI-assistant behavior (Wang et al., 2024).
Relevance & Coherence	Logical alignment with preceding turns and overall conversation objectives to ensure semantic consistency and natural transitions (Chan et al., 2024).

Table 10: Collective Interaction criteria $\hat{\mathcal{C}}$ for evaluating the overall discourse and agent coordination at the group level.

Dataset	Agent	Description	Persona Characteristics
ED	Anchor	Specializes in high-vulnerability emotional support. The primary goal is to provide a safe space for users experiencing intense negative emotions like sadness, loneliness, or devastation.	validating, patient, reflective, emotionally-attuned, non-judgmental
	Catalyst	Helps users navigate complex or frustrating situations (e.g., anger, anxiety, or guilt) and focus on the 'why' and 'what now' of a situation.	analytical, curious, solution-focused,ocratic, perspective-shifting
	Beacon	Amplifies the user's moments of joy, pride, and gratitude and specialize in 'capitalizing'—the process of expanding the benefits of positive news.	enthusiastic, strengths-focused, celebratory, energetic, positive-reinforcing
QMSum	Minutes Scribe	Distill long meeting transcripts into concise summaries.	Objective, analytical, and highly structured.
	Decision Logger	Given the meeting transcript, discuss the decisions made in the meeting.	analytical, evidence-based, systematic
	Action Item Captain	Given the meeting transcript, discuss the actionable items so far for the meeting.	practical, action-oriented, precise
	Critic	Critic for a workplace meeting. Review the meeting records from a critical perspective, discuss gaps, missing information, or overlooked issues.	critical, constructive, systematic

Table 11: Definition of Agent Personas for Experiment Setup for datasets Empathetic Dialogues (ED) and QMSum.