SOCIALAGENT: FROM COLLABORATIVE REASONING TO SELF-IMPROVING SOCIAL INTELLIGENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Artificial Social Intelligence (ASI) refers to the ability of machines to understand and respond to human social behavior, constituting an important dimension of general artificial intelligence. Such ability depends on understanding how different participants perceive the same situation and drawing on experience from previous interactions. However, current multimodal large language models (MLLMs) largely reason about each interaction in isolation, without explicitly reconciling participants' perspectives or accumulating reusable social knowledge. We introduce SocialAgent, a framework that develops social reasoning through collaborative experience acquisition and its subsequent internalization into a single model. SocialAgent-Scaffold combines Role-based Collaboration (RoC) with an Evolving Experience Pool (EEP) without parameter updates. RoC reasons from individual participants' perspectives and reconciles their complementary evidence, producing explicit demonstrations of perspective taking. The EEP abstracts successful reasoning processes into reusable role and reasoning templates, turning individual solutions into strategies that can guide new interactions. Although this collaboration provides structured reasoning demonstrations, repeatedly simulating multiple roles incurs substantial inference cost. SocialAgent-Evolve therefore internalizes this experience into a single-agent MLLM through Strategy-Mediated Self-Refinement (SMSR). Its Strategy-guided Learning stage trains the student on scaffold-generated trajectories paired with their guiding EEP strategies. The trajectories demonstrate how to solve concrete cases, while the strategies make generalizable reasoning principles explicit. This pairing enables the student to learn perspective taking without replaying multi-role collaboration at inference time. Since imitation alone leaves gaps beyond the demonstrated solutions, Multi-strategy Policy Optimization then targets the student's remaining failures, using alternative EEP strategies to guide exploration of diverse reasoning paths. Together, these stages connect learning from collaborative demonstrations with failure-driven self-refinement, without relying on a stronger external teacher. Extensive experiments on social reasoning benchmarks demonstrate that SocialAgent achieves state-of-the-art performance on both open-ended generation and multiple-choice question answering.

1 INTRODUCTION

Artificial Social Intelligence (ASI) (Bainbridge et al., 1994; Dautenhahn, 2007) refers to the ability of machines to understand, interpret, and respond to human social behavior, and constitutes an important dimension of general artificial intelligence. Beyond recognizing observable actions, ASI requires reasoning about the latent intentions, beliefs, and emotions that shape human interactions. Here, social interactions refer to context-dependent exchanges of verbal or nonverbal behavior among people. In unconstrained videos, these mental states must be inferred from nuanced multimodal cues, including facial expressions, gestures, dialogue, and contextual relationships. Such cues are often ambiguous, as the same smile may convey affection, embarrassment, or sarcasm depending on the participants and the surrounding situation. Benchmarks such as Social-IQ (Zadeh et al., 2019) capture this challenge through video question answering tasks that require interpreting social dynamics beyond surface-level perception.

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.

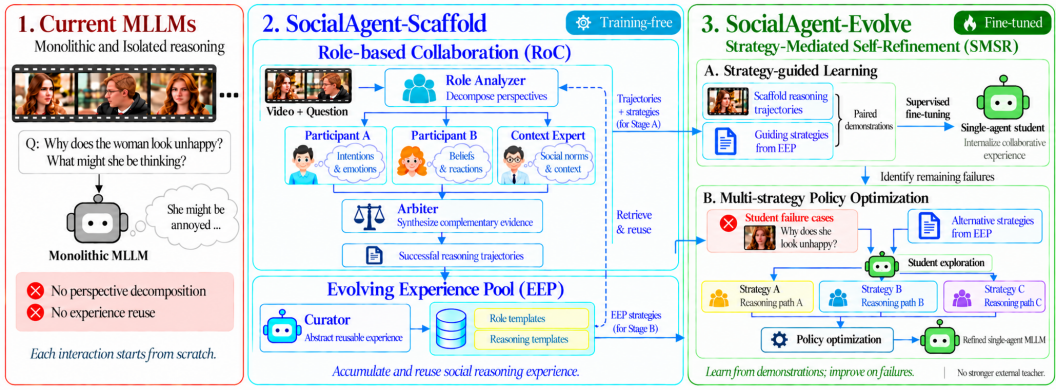

Figure 1: Overview of SocialAgent. Left: standard MLLM inference offers limited explicit support for perspective decomposition and experience reuse. Middle: SocialAgent-Scaffold acquires social reasoning experience without parameter updates. Role-based Collaboration (RoC) integrates complementary perspectives, while an Evolving Experience Pool (EEP) retains reusable role and reasoning templates. Together, they provide concrete reasoning demonstrations and generalizable strategies. Right: SocialAgent-Evolve uses this experience to develop a single-agent MLLM through Strategy-Mediated Self-Refinement (SMSR). Strategy-guided Learning internalizes demonstrations paired with their guiding strategies, and Multi-strategy Policy Optimization uses alternative EEP strategies to guide exploration on the model's remaining failures.

Human social reasoning draws on two closely related capacities, perspective taking and experience accumulation (Decety & Jackson, 2006; Bandura & Walters, 1977). Perspective taking allows people to consider what different participants know, believe, and intend, while accumulated experience provides reusable ways to interpret unfamiliar interactions. These capacities motivate artificial social reasoning systems that explicitly consider distinct viewpoints and reuse insights across situations. However, in standard inference settings, multimodal large language models (MLLMs) (Yang et al., 2025; Achiam et al., 2023; Liu et al., 2024; Lin et al., 2024; Maaz et al., 2024) typically process each interaction as an isolated, monolithic reasoning problem. Multiple perspectives must be reconciled within a single response, and useful reasoning patterns are not explicitly retained for subsequent examples. As illustrated in the left panel of Figure 1, this setup offers limited explicit support for systematic perspective decomposition and cumulative experience reuse, hindering robust social inference across diverse situations.

We introduce SocialAgent, a unified framework for acquiring social reasoning experience through collaboration and internalizing that experience into a single model. Explicit multi-role reasoning provides a way to examine and reconcile different perspectives, but repeating this process for every new case incurs substantial inference cost. SocialAgent therefore connects two complementary stages. SocialAgent-Scaffold makes perspective taking explicit and accumulates reusable experience without parameter updates. SocialAgent-Evolve uses the resulting demonstrations and strategies to train a single-agent MLLM, reducing the need to replay multi-role collaboration at inference time while enabling further improvement on cases the model still finds difficult.

SocialAgent-Scaffold combines Role-based Collaboration (RoC) and an Evolving Experience Pool (EEP) to support perspective taking and experience accumulation, respectively. In RoC, a role analyzer identifies relevant perspectives, role players examine the scene from these viewpoints, and an arbiter reconciles their evidence into a coherent answer. This process makes explicit how different participants can interpret the same behavior differently, providing demonstrations of perspective taking grounded in the observed scene. Complementing this reasoning process, the EEP retains reusable experience. A curator abstracts successful reasoning processes into role templates that specify useful perspectives and reasoning templates that describe how to organize evidence. These templates guide subsequent collaboration and provide strategies that can also be used in single-agent learning. This process thus produces two complementary forms of experience, concrete trajectories showing how to solve individual cases and reusable strategies explaining how to approach similar problems.

SocialAgent-Evolve turns this experience into single-agent capability through Strategy-Mediated Self-Refinement (SMSR). Strategy-guided Learning uses supervised fine-tuning on scaffold-generated trajectories paired with the EEP strategies that guided them. For example, a strategy may recommend comparing a speaker's apparent intention with a listener's reaction, while its paired trajectory demonstrates how dialogue and facial expressions support that comparison in a particular video. This pairing exposes the model to both a worked solution and the reasoning principle behind it, supporting perspective taking without explicitly simulating every role at test time. However, learning from these demonstrations does not directly address the trained model's remaining weaknesses. Multi-strategy Policy Optimization therefore uses alternative EEP strategies to guide exploration on cases the model still fails to solve, and uses the resulting reasoning paths for policy improvement. The order follows this learning dependency, with demonstrations establishing an initial reasoning policy and subsequent exploration targeting its observed failures. As a result, the EEP serves as reusable guidance for learning and exploration, allowing the model to build on collaborative experience without requiring a stronger external teacher.

Our contributions are threefold:

- We introduce SocialAgent, a unified framework that connects collaborative experience acquisition with single-agent learning for social reasoning. Its training-free stage, SocialAgent-Scaffold, combines RoC and EEP to integrate complementary perspectives and turn successful reasoning processes into reusable experience.
- We propose SocialAgent-Evolve to internalize and refine this experience through SMSR. It learns from reasoning demonstrations paired with reusable strategies and then uses alternative EEP strategies to improve its policy on remaining failures, without relying on a stronger external teacher.
- Extensive experiments on the challenging Social-IQ 2.0 (Zadeh et al., 2019) and MuMA-ToM (Shi et al., 2025) benchmarks demonstrate that SocialAgent achieves state-of-the-art performance, supporting the effectiveness of collaborative scaffolding and subsequent strategy-mediated single-agent refinement.

2 Related Work

2.1 Social Reasoning Benchmarks

Social-IQ (Zadeh et al., 2019) evaluates artificial social intelligence through question answering about human interactions in videos, while Social-IQ 2.0 provides updated question–answer annotations. MuMA-ToM (Shi et al., 2025) offers a complementary evaluation of multimodal Theory-of-Mind reasoning through videos and textual descriptions of household interactions. Its questions cover beliefs, social goals, and beliefs about others' goals. Existing approaches for social reasoning incorporate external knowledge (Natu et al., 2023; Park et al., 2020), emotional cues for contrastive learning (Xie & Park, 2023), or attention-based question–answer modeling (Pirhadi et al., 2023). F2F-CL (Wilf et al., 2023) decomposes interactions by speaking turns and applies graph contrastive learning. Looped Video Debating (LVD) (Mori et al., 2025) iteratively requests additional visual evidence from a VQA model to refine an LLM's answer. Social-IQ also supports evaluation of broader video understanding models (Wu et al., 2024b; Li et al., 2024). Our work focuses on organizing participant-specific perspectives and accumulating reusable reasoning experience across scenarios. SocialAgent-Scaffold abstracts collaborative reasoning into role and reasoning templates, providing procedural guidance on which perspectives to consider and how to interpret the available evidence, with evaluation on both Social-IQ 2.0 and MuMA-ToM.

2.2 Multimodal Large Language Models

Recent advances in multimodal large language models (MLLMs) have improved video question answering (VideoQA) (Chen & Dolan, 2011; Yu et al., 2019; Fei et al., 2024) by connecting multimodal perception with language-based reasoning. These models integrate multimodal information through feature alignment and cross-modal fusion (Yu et al., 2024; Wu et al., 2024a; Zhu et al., 2023). Representative approaches include Video-LLaMA (Zhang et al., 2023), which connects visual and audio encoders with a language model through Q-Former-based modules; Video-

ChatGPT (Maaz et al., 2024), which projects spatial and temporal video features into the language model; and Video-LLaVA (Lin et al., 2024), which aligns image and video representations with language. PG-Video-LLaVA (Munasinghe et al., 2023) further extends video understanding with pixel-level grounding. Recent studies also investigate social reasoning in multimodal settings. Li et al. (Li et al., 2025b) study belief-related attention representations on GridToM and apply targeted attention-head interventions during inference. VisionToM (Liu et al., 2026) introduces visual and task-oriented representation interventions for video-only mental-state reasoning. GazeCoT (Ye et al., 2026) incorporates estimated gaze information into visual and textual prompts for social question answering. VEGAS (Li et al., 2025a) introduces Generalist Instruction Fine-Tuning (GIFT) on crafted datasets to improve social video understanding. SocialAgent focuses on connecting collaborative reasoning experience with subsequent model improvement. SocialAgent-Scaffold organizes role-based collaboration and accumulates reusable reasoning strategies without parameter updates. SocialAgent-Evolve then internalizes this experience through strategy-conditioned learning and performs strategy-mediated refinement on the student's own failure cases.

3 METHOD

3.1 OVERVIEW OF SOCIALAGENT

SocialAgent connects collaborative experience acquisition with single-agent learning through two complementary stages (Figure 2). SocialAgent-Scaffold makes perspective taking explicit through multi-role collaboration and accumulates reusable experience without parameter updates. SocialAgent-Evolve internalizes and refines this experience, reducing the cost of repeated multi-role inference.

SocialAgent-Scaffold combines Role-based Collaboration (RoC) for perspective taking with an Evolving Experience Pool (EEP, $\mathcal{E}$) for experience accumulation. Given a video $\mathcal{V}$, subtitles $\mathcal{S}$, and a question $\mathcal{Q}$, a Role-Analyze Agent (A_{RA}) identifies perspectives, Role-Play Agents ($A_{RP}^{(k)}$) reason in parallel, and an Arbiter Agent (A_{AB}) integrates their evidence into an answer. A Knowledge Curator Agent (A_{KC}) abstracts successful reasoning into role and reasoning templates in $\mathcal{E}$, guiding subsequent collaboration and single-agent learning.

SocialAgent-Evolve internalizes this experience through Strategy-Mediated Self-Refinement (SMSR). Strategy-guided Learning first fine-tunes a single-agent student on scaffold-generated trajectories paired with their guiding EEP strategies, combining concrete demonstrations with explicit reasoning principles. Multi-strategy Policy Optimization then targets the student's remaining failures, extending GRPO with alternative EEP strategies to explore diverse reasoning paths and update the same model.

3.2 SOCIALAGENT-SCAFFOLD

As illustrated in the left of Figure 2, SocialAgent-Scaffold combines Role-based Collaboration (RoC) with an Evolving Experience Pool (EEP) to support social reasoning without parameter updates. RoC decomposes social interactions into role-specific perspectives and synthesizes their complementary evidence, while the EEP accumulates generalizable role and reasoning templates abstracted from successful reasoning processes.

3.2.1 ROLE-BASED COLLABORATION

Participants may interpret the same interaction differently because of their distinct beliefs, emotions, and goals. RoC makes these perspectives explicit through a three-tier pipeline comprising role analysis, parallel role play, and arbitration.

Role analysis. The Role-Analyze Agent (A_{RA}) identifies perspectives relevant to answering the question, including participant-specific viewpoints and contextual perspectives such as social norms. Given the question $\mathcal{Q}$ and retrieved experience templates $\mathcal{T}_{\text{ret}}$, it produces a structured set of role–perspective pairs:

$$\{(R_k, P_k)\}_{k=1}^{K} = A_{RA}(\mathbf{P}_{RA}(\mathcal{Q}, \mathcal{T}_{\text{ret}})), \tag{1}$$

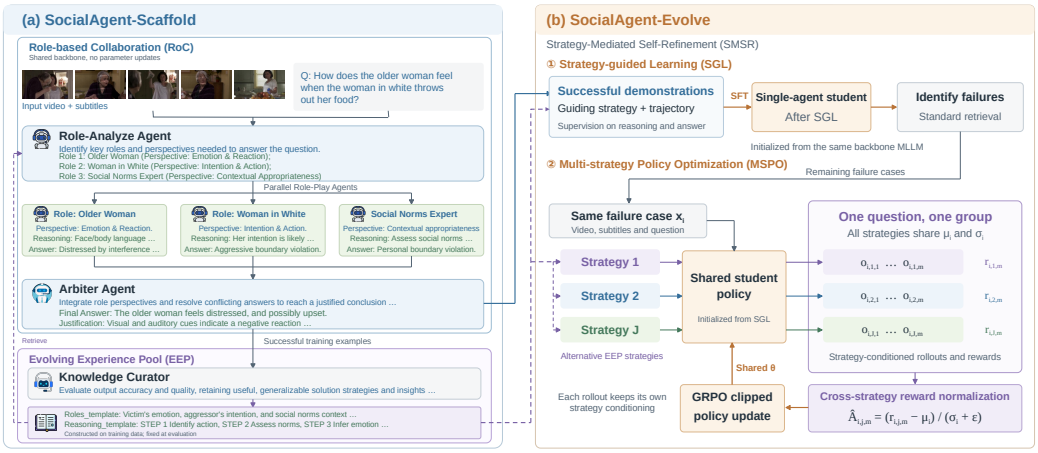

Figure 2: **Overview of SocialAgent.** (a) SocialAgent-Scaffold uses Role-based Collaboration (RoC) to analyze relevant perspectives, reason through parallel role-play agents, and synthesize an answer through arbitration. A knowledge curator abstracts successful training examples into role and reasoning templates in the Evolving Experience Pool (EEP), which guides subsequent reasoning. (b) SocialAgent-Evolve performs Strategy-Mediated Self-Refinement (SMSR) in two steps. Strategy-guided Learning trains a single-agent student on successful scaffold demonstrations paired with their guiding strategies. Multi-strategy Policy Optimization then uses alternative EEP strategies to generate rollouts for the student's remaining failure cases. Rollouts for the same question form one group for cross-strategy reward normalization and a shared-parameter GRPO policy update.

where R_k denotes a role, P_k specifies its reasoning focus, and $\mathbf{P}_{RA}$ is the role-analysis prompt. Retrieved role templates guide the selection of useful perspectives while allowing the decomposition to adapt to the current question.

Role play. Each Role-Play Agent ($A_{RP}^{(k)}$) examines the video $\mathcal{V}$ and subtitles $\mathcal{S}$ from its assigned perspective. This separation allows the agents to consider different interpretations of the same interaction before synthesis. Conditioned on its role assignment and retrieved reasoning templates, each agent produces an intermediate conclusion C_k and supporting rationale L_k:

$$(C_k, L_k) = A_{RP}^{(k)}\big(\mathbf{P}_{RP}^{(k)}(R_k, P_k, \mathcal{V}, \mathcal{S}, \mathcal{T}_{\text{ret}})\big). \tag{2}$$

The resulting findings, $\mathcal{F} = \{(C_k, L_k)\}_{k=1}^{K}$, retain perspective-specific interpretations and their supporting evidence, including inferences about participants' possible beliefs, emotions, and intentions.

Arbitration. Individual perspectives may be incomplete or mutually inconsistent. The Arbiter Agent (A_{AB}) therefore compares the conclusions and supporting rationales in $\mathcal{F}$, weighs conflicting interpretations, and synthesizes a final answer $\hat{\mathcal{A}}$ with justification $\mathcal{J}$:

$$(\hat{\mathcal{A}}, \mathcal{J}) = A_{AB}(\mathbf{P}_{AB}(\mathcal{F}))\,. \tag{3}$$

3.2.2 Evolving Experience Pool

RoC organizes reasoning within an interaction, while the EEP retains reusable experience across interactions. Specifically, the EEP stores abstract procedural templates that describe which perspectives to consider and how to organize social evidence.

Experience curation. The Knowledge Curator Agent (A_{KC}) examines the role assignments, role-specific findings, and final justification from a successful collaborative reasoning process. It abstracts this process into an experience entry:

$$\mathcal{T}_i = (\mathcal{K}_i, \mathcal{T}_{\mathcal{R},i}, \mathcal{T}_{\text{Steps},i})\,, \tag{4}$$

where $\mathcal{K}_i$ contains strategy keywords, $\mathcal{T}_{\mathcal{R},i}$ specifies useful roles and their reasoning contributions, and $\mathcal{T}_{\text{Steps},i}$ describes an abstract reasoning procedure. The curator replaces scene-specific details with placeholders such as `[Observed Action]` and `[Emotion Inferred]`, preserving generalizable reasoning structure rather than individual answers.

The EEP is constructed from training examples whose predicted answers agree with the ground truth. For each successful example, the curator adds the resulting experience entry to $\mathcal{E}$, progressively accumulating reusable role and reasoning templates.

Experience retrieval and reuse. Given a question $\mathcal{Q}$, we extract informative lexical tokens $\mathcal{K}_{\mathcal{Q}}$ and score each stored entry by keyword overlap:

$$s_i = |\mathcal{K}_{\mathcal{Q}} \cap \mathcal{K}_i| \,. \tag{5}$$

We retrieve a relevant experience template $\mathcal{T}_{\text{ret}}$ based on these scores. Its role and reasoning components are incorporated into the role-analysis and role-play prompts, respectively, enabling accumulated experience to guide the current collaborative reasoning process.

3.3 SocialAgent-Evolve

Although multi-role deliberation yields strong reasoning, explicitly simulating multiple characters incurs substantial inference cost. SocialAgent-Evolve turns the experience acquired by SocialAgent-Scaffold into single-agent capability through Strategy-Mediated Self-Refinement (SMSR). Strategy-guided Learning first trains the student on scaffold-generated trajectories paired with their guiding EEP strategies, allowing it to learn both concrete solutions and reusable reasoning principles. Building on this initial policy, Multi-strategy Policy Optimization uses alternative EEP strategies to guide exploration on the student's remaining failures, enabling further improvement beyond imitation of collaborative demonstrations.

3.3.1 Strategy-guided Learning

We initialize a single-agent student M_θ from the same backbone MLLM used by the scaffold agents, with initial parameters θ_0. We train it on scaffold-generated reasoning trajectories paired with their guiding strategies. Given the training set $\mathcal{D}_{\text{train}} = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i = (\mathcal{V}_i, \mathcal{S}_i, \mathcal{Q}_i)$ and y_i is the reference answer, SocialAgent-Scaffold produces a collaborative reasoning trajectory z_i^{sc} and an answer $\hat{y}_i^{\text{sc}}$ under a guiding strategy s_i^{sc}. The strategy consists of the role and reasoning templates used to guide the collaborative process. We retain successful demonstrations to construct

$$\begin{aligned} \mathcal{D}_{\text{SGL}} = \big\{ (x_i, s_i^{\text{sc}}, z_i^{\text{sc}}, \hat{y}_i^{\text{sc}}) \;\big|\; (x_i, y_i) \in \mathcal{D}_{\text{train}}, \\ \text{Correct}(\hat{y}_i^{\text{sc}}, y_i) = 1 \big\}, \end{aligned} \tag{6}$$

where Correct denotes the task-specific correctness criterion.

Starting from θ_0, we optimize the supervised objective

$$\mathcal{L}_{\text{SGL}}(\theta) = -\mathbb{E}_{(x,s,z,\hat{y})\sim\mathcal{D}_{\text{SGL}}} \left[\log p_\theta(z, \hat{y} \mid x, s) \right]. \tag{7}$$

The guiding strategy remains part of the input, and the loss is applied to the reasoning and answer tokens. Pairing each trajectory with its strategy exposes the student to both the reasoning process and the procedural guidance that shaped it. The resulting student, denoted $M_{\theta_{\text{SGL}}}$, learns to carry out reasoning informed by collaborative experience within a single-agent model.

3.3.2 Multi-strategy Policy Optimization

To address the student's remaining failures, we propose Multi-strategy Policy Optimization (MSPO), an extension of Group Relative Policy Optimization (GRPO) (Shao et al., 2024). In our setting, vanilla GRPO samples responses under a fixed retrieved strategy, which may restrict exploration to similar unsuccessful reasoning paths and make successful trajectories difficult to discover. When all responses receive identical rewards, their relative advantages vanish. MSPO broadens training-time exploration through alternative strategies from the EEP and uses shared-parameter learning to transfer the resulting experience across strategies.

Starting from $M_{\theta_{\mathrm{SGL}}}$, we collect training examples answered incorrectly under the standard retrieval procedure into $\mathcal{D}_{\mathrm{fail}}$. For each $(x_i, y_i) \in \mathcal{D}_{\mathrm{fail}}$, we retrieve alternative strategies $\mathcal{G}_i = \{s_{i,j}\}_{j=1}^{J_i}$ and sample student responses:

$$o_{i,j,m} = (z_{i,j,m}, \hat{y}_{i,j,m}) \sim p_{\theta_{\mathrm{old}}}(\cdot \mid x_i, s_{i,j}), \tag{8}$$

where m indexes rollouts under each strategy and the policy is initialized from θ_{SGL}. Different strategies guide the student toward alternative perspectives and evidence relations, expanding exploration beyond repeated sampling under a single strategy.

Each response receives a task-specific reward $r_{i,j,m} = r(\hat{y}_{i,j,m}, y_i)$. We pool all responses for the same question across strategies into one group and compute

$$\hat{A}_{i,j,m} = \frac{r_{i,j,m} - \mu_i}{\sigma_i + \epsilon}, \qquad \mu_i = \mathrm{mean}_{j,m}\, r_{i,j,m}, \quad \sigma_i = \mathrm{std}_{j,m}\, r_{i,j,m}, \tag{9}$$

where $\epsilon > 0$ ensures numerical stability. This shared baseline compares reasoning outcomes across strategies. We then optimize the GRPO clipped objective using these advantages, with each response's token-level probability ratio conditioned on its own guiding strategy:

$$\rho_{i,j,m,t}(\theta) = \frac{p_\theta(o_{i,j,m,t} \mid x_i, s_{i,j}, o_{i,j,m,<t})}{p_{\theta_{\mathrm{old}}}(o_{i,j,m,t} \mid x_i, s_{i,j}, o_{i,j,m,<t})}. \tag{10}$$

All strategy-conditioned responses update the same student parameters, enabling reasoning experience acquired under different strategies to contribute to transferable social reasoning capabilities. At test time, we retain the standard retrieval procedure and hold the retrieved strategy fixed during answer generation. MSPO thus aims to improve single-agent inference under this fixed-strategy setting through broader training-time exploration.

4 Experiment

4.1 Experimental Settings

Benchmarks. We evaluate SocialAgent on Social-IQ 2.0 (Zadeh et al., 2019) for open-ended and multiple-choice social video question answering, and MuMA-ToM (Shi et al., 2025) for multimodal Theory-of-Mind reasoning. MuMA-ToM covers belief inference, social goal inference, and inference of beliefs about others' goals in household interactions.

Metrics. For open-ended Social-IQ 2.0 evaluation, following Video-ChatGPT (Maaz et al., 2024) and VEGAS (Li et al., 2025a), GPT-3.5-turbo receives all four candidate options and evaluates answer correctness (Accuracy, %) and semantic alignment with the reference (Score, 1–5). For multiple-choice evaluation, we report index-matching accuracy on Social-IQ 2.0 and overall and per-category accuracy on MuMA-ToM.

Implementation details. All scaffold agents use Qwen3-VL-8B-Instruct (Bai et al., 2025) with role-specific prompts, eight sampled video frames, subtitles, and the question. The EEP is built only from training data and fixed during evaluation. Strategy-guided Learning uses rank-8 LoRA for five epochs with a learning rate of 10^{-4}. MSPO samples 4 responses per retrieved strategy for each question, jointly normalizes their rewards, and trains for 300 steps with a learning rate of 10^{-6}. Full details are provided in the supplementary material.

4.2 Main Results

Table 1 reports results on Social-IQ 2.0 and MuMA-ToM. Training-free methods are evaluated on both benchmarks, while supervised comparisons use the available Social-IQ 2.0 training split for adaptation.

Among training-free methods, SocialAgent-Scaffold achieves the best performance on every reported metric. It improves Social-IQ 2.0 open-ended and multiple-choice accuracy over the base MLLM by 7.08 and 7.55 percentage points, respectively, and outperforms Multi-agent Debate on all three metrics. These results support the benefits of structured collaboration and experience reuse.

Table 1: Performance on Social-IQ 2.0 and MuMA-ToM. Methods are grouped into Supervised Method and Training-free Method according to whether model parameters are updated.

Method	Social-IQ 2.0			MuMA-ToM			
	Open-ended		Multiple-choice	Belief	Social Goal	Belief of Goal	Overall
	Accuracy	Score	Accuracy				
Training-free Method							
Base MLLM (Bai et al., 2025)	39.04	3.15	63.49	72.28	50.00	40.52	49.78
Multi-step CoT (Wei et al., 2022)	40.82	3.24	65.52	74.10	51.62	41.88	51.20
Self-consistency (Wang et al., 2022)	42.15	3.31	66.65	76.35	52.94	43.10	53.02
Multi-agent Debate (Liang et al., 2024)	43.67	3.38	68.77	78.92	53.80	44.21	54.37
SocialAgent-Scaffold (Ours)	**46.12**	**3.52**	**71.04**	**81.68**	**54.95**	**46.57**	**56.33**
Supervised Method							
DeSIQ (Guo et al., 2023)	–	–	68.93				
Video-LLaVA (Lin et al., 2024)	49.50	3.40	79.17	–	–	–	–
VEGAS (Li et al., 2025a)	54.90	3.90	80.90	–	–	–	–
Qwen3-VL-8B + vanilla SFT	50.80	3.56	78.81	–	–	–	–
Qwen3-VL-8B + vanilla SFT + GRPO	52.70	3.72	80.36	–	–	–	–
SocialAgent-Evolve (Ours)	**56.10**	**4.01**	**83.49**	–	–	–	–

Table 2: Component ablation of SocialAgent-Scaffold.

Model Variant	Acc. (%)
Base MLLM	63.49
+ RA + RP	67.31
+ Arbiter	69.15
+ EEP (Full)	**71.04**

Table 3: Training-stage ablation of SocialAgent-Evolve.

Model Variant	Acc. (%)
SGL only	80.12
SGL + vanilla GRPO	81.47
SGL + MSPO (Full)	**83.49**

Table 4: MSPO sampling allocation. J: strategies; M: rollouts per strategy.

J	M	Acc. (%)
1	8	81.47
2	4	**83.49**
4	2	83.08

On MuMA-ToM, Scaffold raises overall accuracy from 49.78% to 56.33%, with gains across all question categories.

Among supervised methods, SocialAgent-Evolve achieves the highest reported Social-IQ 2.0 results, with 56.10% open-ended accuracy, a 4.01 score, and 83.49% multiple-choice accuracy. It exceeds vanilla SFT followed by GRPO on the same backbone by 3.40 percentage points, 0.29, and 3.13 percentage points, respectively, supporting strategy-mediated learning and refinement. It also surpasses VEGAS by 1.20 percentage points, 0.11, and 2.59 percentage points.

SocialAgent-Evolve also reduces inference time from SocialAgent-Scaffold's 72.29 s to 15.35 s, a $4.71\times$ speedup, while improving open-ended and multiple-choice accuracy from 46.12% and 71.04% to 56.10% and 83.49%. These results show that internalizing collaborative experience improves both accuracy and inference efficiency. We omit VEGAS latency because its implementation is unavailable for direct measurement. Both SocialAgent-Evolve and VEGAS use single-model generation without test-time parameter updates, suggesting inference costs of a similar order for comparably sized backbones under matched inference settings.

4.3 Ablation Study

Ablation of SocialAgent-Scaffold. Table 2 examines the contributions of the scaffold components on Social-IQ 2.0 multiple-choice QA. Introducing role analysis and parallel role play improves accuracy from 63.49% to 67.31%, suggesting that explicitly considering complementary perspectives benefits social reasoning. Adding arbitration further increases accuracy to 69.15%, supporting the value of synthesizing role-specific findings. Incorporating the EEP further brings the full scaffold to 71.04%. These cumulative improvements suggest that perspective decomposition, evidence synthesis, and experience reuse provide complementary benefits without parameter updates.

Ablation of SocialAgent-Evolve. Table 3 compares the learning stages of SocialAgent-Evolve. Strategy-Guided Learning alone achieves 80.12% accuracy, while subsequent vanilla GRPO improves it to 81.47%. Replacing vanilla GRPO with MSPO further raises accuracy to 83.49%, a gain of 2.02 percentage points. This comparison supports the use of alternative strategies and cross-strategy reward normalization for policy refinement. Since test-time inference retains the standard single-strategy retrieval procedure, the improvement is consistent with training-time exploration contributing to capabilities retained in the shared student parameters.

Sampling allocation in MSPO. Table 4 compares sampling allocations under a fixed budget of eight rollouts per question. Using two strategies with four rollouts each achieves the highest accuracy of 83.49%, exceeding eight rollouts under a single strategy by 2.02 percentage points. Increasing the number of strategies to four while reducing rollouts per strategy to two yields 83.08%. This pattern suggests a trade-off between exploring different reasoning strategies and sampling alternative responses within each strategy. Among the tested configurations, two strategies with four rollouts each provide the most effective allocation.

4.4 Qualitative Results

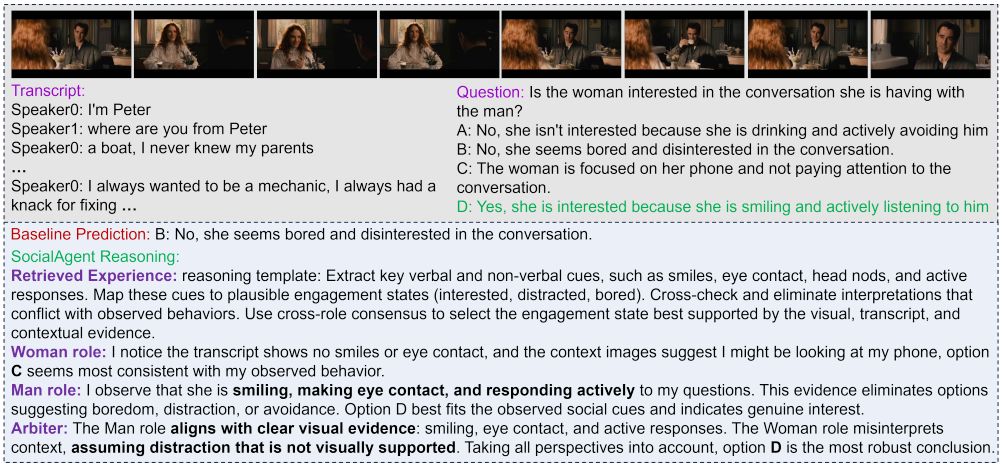

Figure 3: Visualization of baseline prediction and multi-role reasoning steps of our SocialAgent.

Figure 3 shows SocialAgent's reasoning for the question: "Is the woman interested in the conversation with the man?" The baseline model incorrectly predicts that she appears bored, failing to capture her engagement. SocialAgent retrieves prior experience emphasizing verbal and non-verbal cues such as smiles, eye contact, and active responses. The Woman role initially interprets her behavior as distraction due to phone usage, suggesting disinterest. In contrast, the Man role observes her smiling, maintaining eye contact, and responding actively, indicating engagement. The Arbiter integrates these perspectives and prioritizes evidence-supported cues, concluding that the woman is indeed interested in the conversation. This example highlights how SocialAgent combines cross-role reasoning with retrieved experience to interpret social cues more accurately while providing an interpretable decision process.

5 Conclusion

We present SocialAgent, a framework that acquires social reasoning experience through collaboration and internalizes it into a single-agent MLLM. SocialAgent-Scaffold combines role-based collaboration with an evolving experience pool to produce complementary reasoning demonstrations and reusable strategies without parameter updates, while SocialAgent-Evolve learns from this experience and targets remaining failures through strategy-guided policy optimization. Our work highlights the value of reusable reasoning strategies in connecting collaborative inference with model learning, offering a path toward more adaptable artificial social intelligence.

AI USE STATEMENT

Generative AI tools were used as general-purpose assistance in this work. Specifically, they were used to assist with debugging and writing small auxiliary bash scripts, to create and refine schematic figures, and to polish the manuscript for clarity and readability. All AI-assisted code and debugging suggestions were manually reviewed and tested by the authors. AI-assisted figures were manually inspected to ensure that they accurately represent the proposed method, and all AI-assisted text was reviewed to ensure that the scientific content and claims reflect the authors' intended meaning. We take full responsibility for the final content of this work, including all text, code, figures, and claims produced with the aid of generative AI.

ETHICS STATEMENT

Our research adheres to rigorous ethical guidelines. We verified the licenses of all software and datasets we used in this study and ensured full compliance with their terms. Furthermore, we have thoroughly assessed the project and do not anticipate any additional potential risks.

REPRODUCIBILITY STATEMENT

Section 3 provides a detailed description of SocialAgent, including Role-based Collaboration, the Evolving Experience Pool, and Strategy-Mediated Self-Refinement. The experimental settings and evaluation protocols are described in Section 4.1. Further implementation details are provided in the supplementary material: Section C presents the training details and hyperparameters, and Section D provides the complete prompt library used in our framework. We plan to release the code and associated scripts upon acceptance to facilitate reproducibility.

REFERENCES

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, et al. Qwen3-vl technical report. *arXiv preprint arXiv:2511.21631*, 2025.

William Sims Bainbridge, Edward E Brent, Kathleen M Carley, David R Heise, Michael W Macy, Barry Markovsky, and John Skvoretz. Artificial social intelligence. *Annual review of sociology*, 20(1):407–436, 1994.

Albert Bandura and Richard H Walters. *Social learning theory*, volume 1. Prentice-hall Englewood Cliffs, NJ, 1977.

David Chen and William B Dolan. Collecting highly parallel data for paraphrase evaluation. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pp. 190–200, 2011.

Kerstin Dautenhahn. A paradigm shift in artificial intelligence: why social intelligence matters in the design and development of robots with human-like intelligence. In *50 Years of Artificial Intelligence: Essays Dedicated to the 50th Anniversary of Artificial Intelligence*, pp. 288–302. Springer, 2007.

Jean Decety and Philip L Jackson. A social-neuroscience perspective on empathy. *Current directions in psychological science*, 15(2):54–58, 2006.

Hao Fei, Shengqiong Wu, Meishan Zhang, Min Zhang, Tat-Seng Chua, and Shuicheng Yan. Enhancing video-language representations with structural spatio-temporal alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):7701–7719, 2024.

Xiao-Yu Guo, Yuan-Fang Li, and Reza Haf. Desiq: Towards an unbiased, challenging benchmark for social intelligence understanding. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 3169–3180, 2023.

Hao Li, Hao Fei, Zechao Hu, Zhengwei Yang, and Zheng Wang. Vegas: Towards visually explainable and grounded artificial social intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 4707–4715, 2025a.

Xinyang Li, Siqi Liu, Bochao Zou, Jiansheng Chen, and Huimin Ma. From black boxes to transparent minds: Evaluating and enhancing the theory of mind in multimodal large language models. *arXiv preprint arXiv:2506.14224*, 2025b.

Yunxin Li, Xinyu Chen, Baotain Hu, and Min Zhang. Llms meet long video: Advancing long video comprehension with an interactive visual adapter in llms. *arXiv preprint arXiv:2402.13546*, 3(7), 2024.

Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. Encouraging divergent thinking in large language models through multi-agent debate. In *Proceedings of the 2024 conference on empirical methods in natural language processing*, pp. 17889–17904, 2024.

Bin Lin, Yang Ye, Bin Zhu, Jiaxi Cui, Munan Ning, Peng Jin, and Li Yuan. Video-llava: Learning united visual representation by alignment before projection. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 5971–5984, 2024.

Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.

Siqi Liu, Xinyang Li, Bochao Zou, Junbao Zhuo, Huimin Ma, and Jiansheng Chen. Video-only tom: Enhancing theory of mind in multimodal large language models. *arXiv preprint arXiv:2603.24484*, 2026.

Muhammad Maaz, Hanoona Rasheed, Salman Khan, and Fahad Khan. Video-chatgpt: Towards detailed video understanding via large vision and language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12585–12602, 2024.

Erika Mori, Yue Qiu, Hirokatsu Kataoka, and Yoshimitsu Aoki. Leveraging llms with iterative loop structure for enhanced social intelligence in video question answering. *arXiv preprint arXiv:2503.21190*, 2025.

Shehan Munasinghe, Rusiru Thushara, Muhammad Maaz, Hanoona Abdul Rasheed, Salman Khan, Mubarak Shah, and Fahad Khan. Pg-video-llava: Pixel grounding large video-language models. *arXiv preprint arXiv:2311.13435*, 2023.

Sanika Natu, Shounak Sural, and Sulagna Sarkar. External commonsense knowledge as a modality for social intelligence question-answering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3044–3050, 2023.

Jae Sung Park, Chandra Bhagavatula, Roozbeh Mottaghi, Ali Farhadi, and Yejin Choi. Visualcomet: Reasoning about the dynamic context of a still image. In *European Conference on Computer Vision*, pp. 508–524. Springer, 2020.

Mohammad Javad Pirhadi, Motahhare Mirzaei, and Sauleh Eetemadi. Just ask plus: Using transcripts for videoqa. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3082–3085, 2023.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Haojun Shi, Suyu Ye, Xinyu Fang, Chuanyang Jin, Leyla Isik, Yen-Ling Kuo, and Tianmin Shu. Muma-tom: Multi-modal multi-agent theory of mind. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 1510–1519, 2025.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*, 2022.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Alex Wilf, Martin Q Ma, Paul Pu Liang, Amir Zadeh, and Louis-Philippe Morency. Face-to-face contrastive learning for social intelligence question-answering. In *2023 IEEE 17th international conference on automatic face and gesture recognition (FG)*, pp. 1–7. IEEE, 2023.

Shengqiong Wu, Hao Fei, Xiangtai Li, Jiayi Ji, Hanwang Zhang, Tat-Seng Chua, and Shuicheng Yan. Towards semantic equivalence of tokenization in multimodal llm. *arXiv preprint arXiv:2406.05127*, 2024a.

Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. Next-gpt: Any-to-any multimodal llm. In *Forty-first International Conference on Machine Learning*, 2024b.

Baijun Xie and Chung Hyuk Park. Multi-modal correlated network with emotional reasoning knowledge for social intelligence question-answering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3075–3081, 2023.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Zhoutong Ye, Xutong Wang, Chengwen Zhang, Ruiwen Zhang, Mingze Sun, Qinwei Li, Chun Yu, and Yuanchun Shi. Gazecot: Unleashing social intelligence in multimodal llms with gaze-informed chain-of-thought reasoning. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, pp. 1–24, 2026.

Shoubin Yu, Jaehong Yoon, and Mohit Bansal. Crema: Multimodal compositional video reasoning via efficient modular adaptation and fusion. *arXiv preprint arXiv:2402.05889*, 1, 2024.

Zhou Yu, Dejing Xu, Jun Yu, Ting Yu, Zhou Zhao, Yueting Zhuang, and Dacheng Tao. Activitynet-qa: A dataset for understanding complex web videos via question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 9127–9134, 2019.

Amir Zadeh, Michael Chan, Paul Pu Liang, Edmund Tong, and Louis-Philippe Morency. Social-iq: A question answering benchmark for artificial social intelligence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8807–8817, 2019.

Hang Zhang, Xin Li, and Lidong Bing. Video-llama: An instruction-tuned audio-visual language model for video understanding. *arXiv preprint arXiv:2306.02858*, 2023.

Bin Zhu, Bin Lin, Munan Ning, Yang Yan, Jiaxi Cui, HongFa Wang, Yatian Pang, Wenhao Jiang, Junwu Zhang, Zongwei Li, et al. Languagebind: Extending video-language pretraining to n-modality by language-based semantic alignment. *arXiv preprint arXiv:2310.01852*, 2023.

A Limitations and Future Work

SocialAgent involves a multi-stage pipeline with multi-agent collaboration, which introduces additional training complexity and computational overhead. Future work will explore simplifying this pipeline while preserving its benefits. We have also not systematically investigated how different reward designs influence social reasoning in MLLMs, leaving opportunities to improve both reasoning quality and generalization.

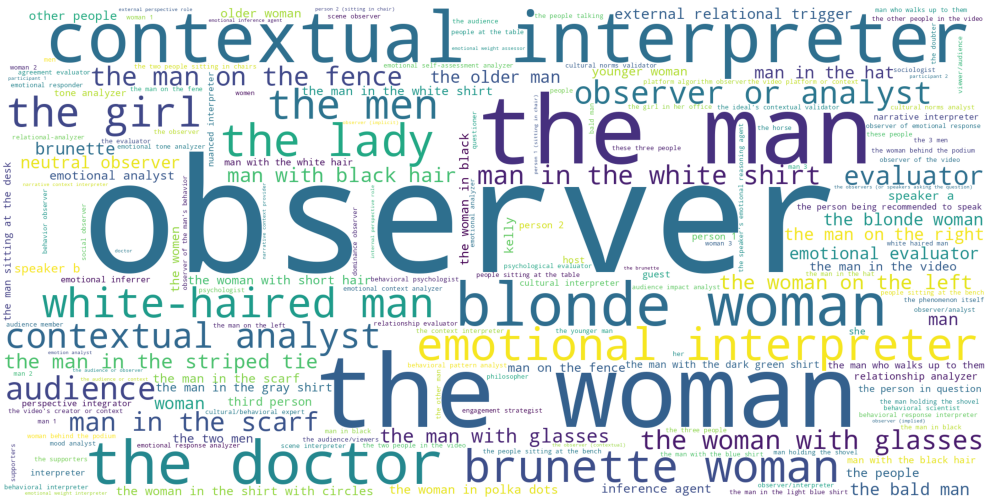

Figure 4: **Role diversity generated by the Role-Analyze Agent.** The distribution shows a dominant "observer" role alongside a long-tail of fine-grained contextual roles, reflecting the heterogeneous perspectives driving multi-agent reasoning.

B Further Analysis of SocialAgent

B.1 Role Diversity Analysis

To better understand how heterogeneous perspectives emerge within the multi-agent reasoning process, we analyze the roles generated by the Role-Analyze Agent. Figure 4 presents a word-cloud visualization of the roles generated across queries. While the generic persona "observer" appears most frequently, the Role-Analyze Agent consistently produces a long-tail of fine-grained and context-sensitive roles. These include semantic perspectives such as contextual interpreter or contextual analyst, emotion-oriented roles such as emotional interpreter, as well as highly specific situational identities (e.g., the woman, the man, the doctor, the woman with glasses, the man on the fence, etc.).

This rich role spectrum demonstrates that the system does more than assign superficial personas—it dynamically grounds roles in the visual and narrative context of each query. Such heterogeneity ensures that parallel role-play agents operate under diverse interpretative frames, leading to complementary evidence extraction and reducing the risk of narrow or biased reasoning. Ultimately, this diversity contributes to more robust and comprehensive multi-agent decision-making.

B.2 Experience Pool Analysis

To better understand the internal structure of the evolving experience pool (EEP), we analyze the final pool constructed from Social-IQ 2.0. The EEP functions as a repository of reusable multi-agent reasoning templates accumulated over time. Our analysis focuses on the curated reasoning primitives stored in the pool. Figure 5 visualizes the distribution of curated keywords using a word cloud. The EEP exhibits a clear concentration around meta-cognitive reasoning motifs such as consensus validation, role-framed interpretation, and evidence triangulation. These primitives correspond to reusable reasoning structures learned during training set exploration.

For example, one frequently observed template involves expert consensus. This pattern typically emerges in questions that require subjective interpretation of social signals, such as inferring a character's emotion, intention, or interpersonal relationship (e.g., "Why does the character feel this way?"). In these cases, the solver adopts role-framed reasoning where multiple expert perspectives analyze the dialogue or behavior cues, and the final decision is guided by agreement among these interpretations.

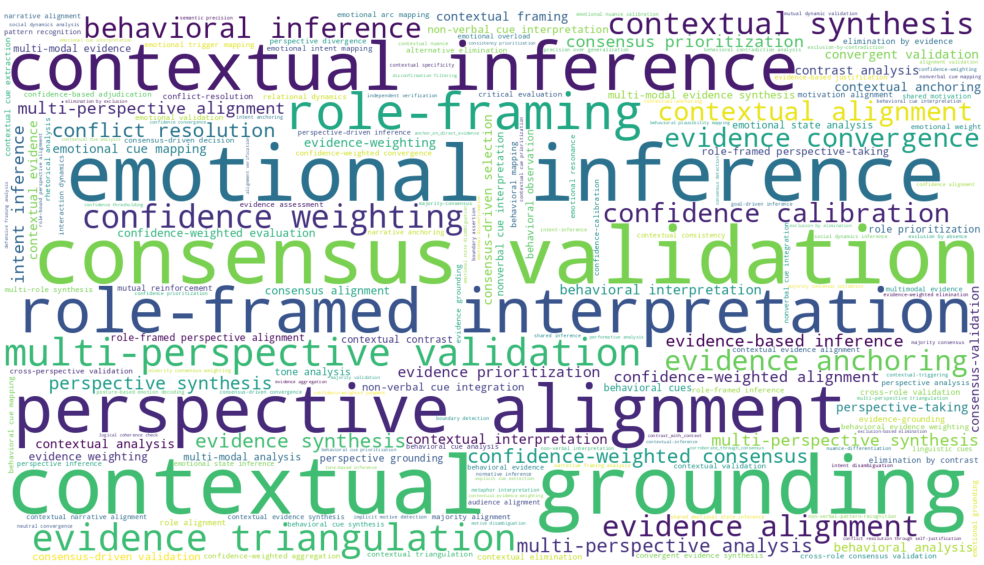

Figure 5: **Curated keyword distribution in the experience pool.** High-frequency reasoning primitives such as consensus validation, role-framed interpretation, and evidence triangulation indicate that the pool encodes structured, reusable meta-cognitive templates.

Another common template reflects evidence triangulation. This strategy is often triggered by questions that require reconstructing events or causal relations (e.g., "What happened next?" or "Why did the character act this way?"). The reasoning process integrates multiple sources of information—such as dialogue content, observable actions, and contextual cues—to validate a hypothesis through cross-checking evidence from different perspectives.

These high-frequency primitives indicate that the EEP does not merely store task-specific solutions, but rather accumulates structured, multi-perspective reasoning templates. Such templates can be retrieved and reused across tasks, enabling the system to progressively refine its collaborative reasoning strategies.

C Implementation Details

SocialAgent-Scaffold. All scaffold agents use Qwen3-VL-8B-Instruct with shared model parameters and role-specific prompts. Each video is represented by eight frames sampled at a fixed stride from its extracted frame sequence, with a maximum pixel budget of 131,072 per frame, together with subtitles and the question. The EEP is constructed exclusively from successful collaborative reasoning on the training split, with correctness determined by exact option matching. During scaffolding, we retrieve up to two experience entries based on lexical keyword overlap and incorporate their role and reasoning templates into the corresponding prompts. The resulting training-derived experience entries are consolidated into a frozen pool for student learning. No model parameters are updated during scaffolding, and the EEP remains fixed during evaluation. Since MuMA-ToM does not provide training question–answer pairs for experience construction, we evaluate SocialAgent-Scaffold on this benchmark without EEP.

SocialAgent-Evolve. The student is initialized from Qwen3-VL-8B-Instruct. Strategy-guided Learning pairs successful scaffold outputs with strategies retrieved from the finalized training pool. Targets contain `<reasoning>` and `<answer>` fields, with answer-only supervision when retained reasoning is unavailable. We apply LoRA to the language model's attention and MLP projections with rank 8, scaling factor 8, and dropout 0.05, while freezing the base model parameters, including the visual encoder. Supervised training uses a learning rate of 10^{-4}, an effective batch size of 16, and a loss computed only over assistant target tokens. The student proceeds to policy optimization after five SFT epochs.

(a) Role Analysis Agent

Analyze the following Video QA task and identify 2-3 specific expert roles needed to solve it accurately.

Question: <QUESTION>
Context Summary: <CONTEXT>

Respond ONLY in JSON:

```
{
  "roles": [
    {"role": "Expert Name",
     "description": "What this expert should focus on"}
  ]
}
```

Figure 6: Prompt for the Role-Analyze Agent.

For Multi-strategy Policy Optimization (MSPO), we identify training examples answered incorrectly by the student under single-strategy retrieval. For each question, we retrieve two strategies using weighted lexical overlap, excluding entries originating from the same video or question and filtering near-duplicates with a token-set Jaccard threshold of 0.85. We sample four responses per strategy with temperature 0.8, top-p of 0.95, and a maximum of 512 new tokens. A response receives a reward of 1 if it is valid, non-truncated, and selects the correct option, and 0 otherwise. The eight responses for each question form a single group for reward normalization across strategies. We optimize the shared student with the GRPO clipped objective, conditioning each response on its generating strategy. The learning rate is 10^{-6}, the clipping threshold is 0.2, and the KL coefficient is 0.01, with the initial SFT policy serving as the frozen reference. MSPO runs for 300 steps on four NVIDIA GeForce RTX 3090 GPUs using BF16 and 16 completions per update. At test time, the student retrieves one strategy from the frozen pool and generates a single response using greedy decoding.

D Prompt Library

We release the MLLM prompts used in our framework to support research transparency and reproducibility. As shown in Figure 6, the Role-Analyze Agent determines the role structure of the question. Each role is instantiated with its own reasoning prompt (Figure 7), and their outputs are integrated by the Arbiter Agent (Figure 8). Finally, the Knowledge Curator (Figure 9) extracts reusable templates for future tasks.

(b) Role Play Agent

You are acting as the: <ROLE_NAME>
Your focus: <ROLE_DESCRIPTION>

Relevant Strategies from Experience Pool:
- Strategy: <STRATEGY>
- Pattern: <REASONING_PATTERN>

Context: <CONTEXT>
Question: <QUESTION>
Choices: <CHOICES>

Instructions:
1. Provide your specialized analysis based on your role.
2. Consider the provided strategies if relevant.
3. Be concise but insightful.

Respond ONLY in JSON:

```
{
  "role": "<ROLE_NAME>",
  "specialized_analysis": "...",
  "tentative_answer": "your best guess"
}
```

Figure 7: Prompt for the Role-Play Agent.

(c) Arbiter Agent

You are the Arbiter. Your task is to synthesize the opinions of multiple experts to provide the final answer.

Question: <QUESTION>
Choices: <CHOICES>

Expert Opinions:
Expert [<ROLE_NAME>]: <SPECIALIZED_ANALYSIS> (Answer: <TENTATIVE_ANSWER>)

Instructions:
1. Evaluate the reasoning of each expert.
2. Resolve any contradictions.
3. Provide the final conclusion.

Respond ONLY in JSON:

```
{
  "reasoning": "Final synthesis of all expert views...",
  "answer": <index_int>
}
```

Figure 8: Prompt for the Arbiter Agent.

(d) Curator Agent

Analyze the successful QA process below and distill it into a reusable 'Experience Pool' entry.
Focus on HOW to solve similar problems, not the specific answer.

Task: <QUESTION>
Reasoning Used: <FINAL_REASONING>

Respond ONLY in JSON:

```
{
  "keywords": ["word1", "word2"],
  "strategy": "One sentence describing the high-level approach",
  "reasoning_pattern": "A template for logical steps to follow"
}
```

Figure 9: Prompt for the Knowledge Curator Agent.