COUPLING MACRO DYNAMICS AND MICRO STATES FOR LONG-HORIZON SOCIAL SIMULATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Social network simulation aims to model collective opinion dynamics in large populations, but existing LLM-based simulators primarily focus on aggregate dynamics and largely ignore individual internal states. This limits their ability to capture opinion reversals driven by gradual individual shifts and makes them unreliable over long-horizon simulations. Unlike existing methods that collapse dynamics into macro-only updates, we propose a social simulation framework, **M^{2}Sim**, which tightly couples macro-level collective dynamics with micro-level individual states. We explicitly model micro-level opinion states with a state transition mechanism, coupling **M**acro-level collective dynamics with **M**icro-level states. This allows the opinion-state distribution to evolve gradually, rather than triggering instant reactions, enabling the simulator to distinguish states close to or far from switching, capture opinion reversals, and maintain accuracy over long-horizon simulations. Across real-world events, our M^{2}Sim supports stable simulation of long-horizon social events with up to **40,000** interactions (compared to $\sim$300 in baseline MF-LLM), while reducing long-horizon KL divergence by **75.3%** (1.2490 $\rightarrow$ 0.3089) and reversal KL by **66.9%** (1.6425 $\rightarrow$ 0.5434), significantly mitigating the drift observed in MF-LLM.

1 INTRODUCTION

Simulating how collective behaviors and social structures emerge is central to understanding social diffusion, mobilization, and opinion formation (Squazzoni et al., 2014; Freeman, 2004). In **social network simulation (SNS)** (Gao et al., 2024), macroscopic public outcomes arise from the aggregation of microscopic actions (Li et al., 2017). This tight micro-macro coupling induces strong nonlinear feedback, such that small initiating groups, delayed evidence, or localized interventions can trigger tipping points and opinion reversals (Lu et al., 2018b; Nie et al., 2018).

Traditional paradigms, including *mechanistic models* (Meadows et al., 1974; Bouanan et al., 2019), *empirical and statistical models* (Kong et al., 2018; Lu & Nie, 2019), and *agent-based models* (Schelling, 1971; Grossi et al., 2005), capture certain macro-micro regularities but rely on static parameters or handcrafted behavioral rules. As a result, they struggle to represent evolving beliefs, delayed commitment, and feedback-driven opinion transitions at scale, motivating the need for scalable, *state-aware* frameworks (Schelling, 2006).

Recent studies show that large language models (LLMs) (Guo et al., 2025; Wei et al., 2025) can endow social agents with reasoning, perception, and interaction capabilities, enabling simulations from small-scale communities (Park et al., 2023; 2024) to multi-scene social systems (Wang et al., 2025; Zhang et al., 2025b). Frameworks such as *GA-S^3* (Zhang et al., 2025c), *Oasis* (Yang et al., 2025b), and *AgentSociety* (Piao et al., 2025) rely on fully LLM-driven agents to generate rich behaviors. To improve scalability and macro-level coherence, *MF-LLM* (Mi et al., 2025) introduces mean-field modeling into LLM-based simulation, coupling individual actions with aggregated macro signals through iterative feedback. While *MF-LLM* yields coherent short-term trajectories, it over-compresses individual dynamics into macro-level signals without per-agent latent state modeling. Consequently, actions are treated as instantaneous reactions rather than state-changing events, which hampers capturing the turning-point timing of opinion reversals and realistic long-horizon collective dynamics.

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.

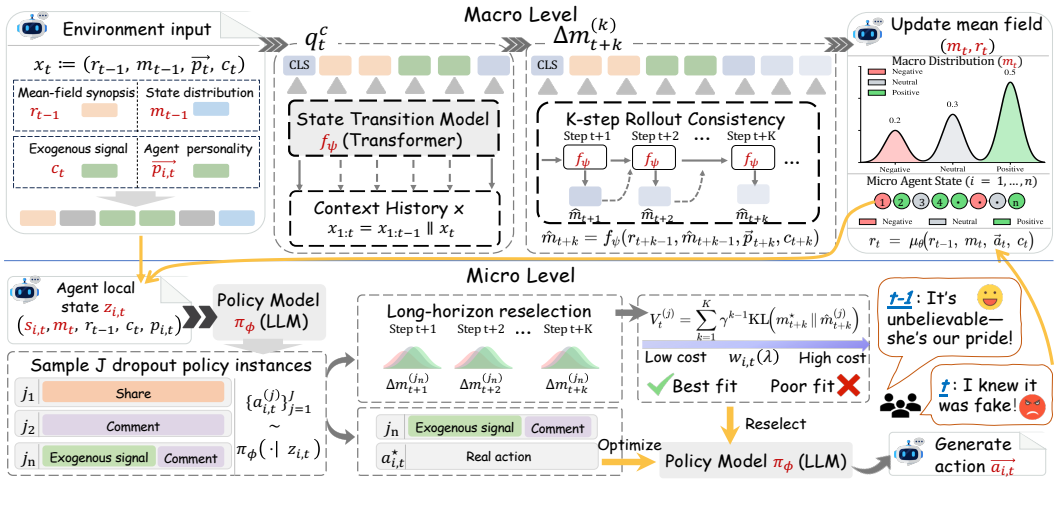

Figure 1: **M^{2}Sim overview. (i) Macro level**, an event-level state transition model (Event Transformer) updates the distributional mean field and is trained with a K-step rollout consistency objective to improve long-horizon trajectory fidelity. **(ii) Micro level**, a LoRA-tuned policy LLM generates actions conditioned on agent states and the mean-field signals; we sample J dropout-induced policy instances, obtain candidate actions, and score them by a K-step mean-field surrogate (distributional KL to future trajectories), yielding soft weights for action reselection and optimizing the policy via a combined auxiliary text loss and discounted long-horizon prediction loss.

These gaps highlight three core challenges in social simulation. **(C1) Micro-Macro Decoupling**: existing approaches focus on either micro-level individual behaviors or macro-level collective dynamics in isolation, failing to capture the tightly coupled co-evolution of both aspects in real social systems. **(C2) Long-Horizon Dynamics Degradation**: achieving *long-horizon* simulation is challenging because single-scale rollouts accumulate errors over time, leading to variance damping, majority lock-in, and unrealistic trajectories under delayed evidence accumulation and gradual opinion transitions. **(C3) Unresolved Opinion Reversals**: reversal events impose a stricter requirement on long-horizon simulations. The model must not only sustain long rollouts but also capture timely and significant opinion reversals when influenced by exogenous signals. Without tightly coupled macro-state signals and micro-level opinion states of current participants, simulations exhibit inertia that misses turning points or drift that overshoots, hindering realistic opinion transitions.

Contributions. To address challenges (C1-C3), we make the following contributions:

1. M^{2}Sim: A new social dynamic simulation framework that couples macro-level collective dynamics and micro-level agent states. We introduce **M^{2}Sim**, which provides macro-level collective signals through a mean field and generates individual actions conditioned on micro-level opinion states, coupling their evolution through an explicit state-transition mechanism to address the micro-macro decoupling challenge (**C1**).

2. Coupled state-transition and rollout modeling for long-horizon dynamics and opinion reversals. To address **C2** and **C3**, M^{2}Sim couples a *macro-level* state transition model with *micro-level* multi-step rollout-based action reselection. At the macro level, the transition model learns the long-term evolution of the state distribution, reducing drift, variance damping, and majority lock-in. At the micro level, agents condition on this evolving distribution and use multi-step rollouts to evaluate candidate actions by their future impacts. This enables agents to adjust their actions in response to exogenous signals, overcoming inertia and capturing the timing and magnitude of opinion reversals.

3. Empirical Validation. Experiments on diverse real-world social events demonstrate that **M^{2}Sim** consistently outperforms existing methods: it significantly reduces long-horizon KL divergence by **75.3%** and reversal KL by **66.9%**, while mitigating the drift commonly observed in *MF-LLM*. Additionally, **M^{2}Sim** maintains the scalability advantages in long-horizon scenes, enabling stable simulations even with up to **40,000** interactions (vs. $\sim$300).

2 BACKGROUND AND MOTIVATION

We formalize collective dynamics as a state-aware decision process (§2.1) and show that omitting micro-level states induces self-reinforcing rollouts that miss opinion reversals (Prop. 2.1).

2.1 PROBLEM AND NOTATION DEFINITION

We consider a social group of N agents and simulate its collective decision-making over $[0, T]$. All agents share a discrete action space $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$ (e.g., comment, share) and state space $\mathcal{S} = \{s_1, s_2, \ldots, s_m\}$. The action and state of agent i at time t are $a_{i,t}$ and $s_{i,t}$, with aggregated vectors $\vec{a}_t = [a_{1,t}, \ldots, a_{N,t}]$ and $\vec{s}_t = [s_{1,t}, \ldots, s_{N,t}]$. Each agent has a personality descriptor $p_{i,t}$, collected as $\vec{p}_t = [p_{1,t}, \ldots, p_{N,t}]$, which modulates its decision behavior. Although $\mathcal{S}$ and $\mathcal{A}$ are written as finite sets for convenience, they correspond in practice to latent or semantic categories derived from open-ended interactions. To characterize macroscopic dynamics, we introduce a mean-field representation $m_t \in \Delta^{m-1}$, defined as the macro state distribution over $\mathcal{S}$ at time t, where

$$\Delta^{m-1} = \left\{ x \in \mathbb{R}^m_{\geq 0} : \sum_{s \in \mathcal{S}} x(s) = 1 \right\}$$

denotes the probability simplex over $\mathcal{S}$. The evolution of m_t captures how decentralized agent decisions give rise to collective behavioral trends, which we also refer to as the *public mood* in the context of social group simulation. The system is driven by exogenous sianals c_t, which are generated endogenously within the simulator by designated information-source agents (e.g., official accounts or media reporters) at step t and then broadcast to the public. These externally grounded messages influence agent states and actions and thereby shape the joint evolution of individual actions and the mean-field trajectory $\{m_t\}_{t=0}^T$, capturing both short-term responses and long-term collective dynamics in social propagation. In addition, we maintain a textual mean-field synopsis $r_t \in \mathbb{T}$, which provides a natural-language summary of recent collective trajectories and exogenous signals; together, (r_t, m_t) constitute the macros.

2.2 DYNAMICS OF PREVIOUS METHOD

The existing simulation dynamics evolve by two parameterized models. First, the action model $\pi_\phi(a_{i,t}|r_t, p_{i,t}, c_t)$, it samples the actions of i-th agent based on the current public mood (mean-field) r_t, the personality of the agent $p_{i,t}$ at time t. Second a mean-field summarizer $\mu_\theta(r_{t+1}|r_t, \mathbf{a}_t)$. It summarize the next public mood (r_{t+1}) based on the current actions $\mathbf{a}_t$ and the current public mood (r_t). The model iteratively updates $\mathbf{a}_t$ and r_t, which stops until convergence or the reaching the terminal time T. It is obvious that this dynamic is state-ignored.

2.3 LIMITATIONS OF STATE-IGNORED DYNAMICS.

A key parameter that tracks evolution in $\{m_t\}_{t=0}^T$ is $f(m) := \mathbb{E}[m_{t+1}|m_t]$, which indicates the expected m_{t+1} given the previous m_t. The $\delta m(s) := \mathbb{E}[m_{t+1}(s)|m_t] - m(s)$ indicates the expected changes of the mean field over a specific state s. We find the following proposition w.r.t. the majority state $s^* = \arg\max_{s \in \mathcal{S}} m(s)$ under very mild assumptions: (1). The mean-field summarizer μ_θ faithfully summarize the m_t with a bounded error ϵ for some very small $\epsilon \geq 0$, i.e.,

$$\mathbb{E}[m_{t+1}(s)|m_t, \mathbf{a}_t] \geq \hat{q}(\mathbf{a}_t)(s) - \epsilon, \tag{1}$$

where the $\hat{q}(\mathbf{a}_t)$ is the empirical distribution of $\mathcal{S}$ yet projected from actions $\mathcal{A}$; (2). The agents, "on average", align slightly toward the current majority of the public mood, i.e.,

$$\mathbb{E}[\hat{q}(\mathbf{a}_t)(s^*)|m_t = m] \geq m(s^*) + \eta, \tag{2}$$

for some $\eta > 0$.

Proposition 2.1 (Self-strengthening of state-ignore dynamics)**.** *After the last exogenous signals c_t, the $\delta m(s^*) \geq \eta - \epsilon \geq 0$, for any $\eta \geq \epsilon$.*

Remark 2.2. In practice, the aggregation error ϵ is empirically small because μ_θ is prompted and trained to produce a faithful summary (sometimes even deterministic). Consequently, the one-step expected majority drift $\delta m(s^*) \geq 0$ in the scenario that a clear majority of public mood exists. This implies a systematic tendency to reinforce the current majority after the last exogenous signal. Though, this does not rule out reversals of personal or public mood, yet reversals become increasingly rare and typically require atypical realizations of the action-sampling noise.

3 METHODOLOGY

We now introduce the **M^{2}Sim** framework (§3.1) to couple macro-level states with micro-level agent states (Fig. 1). Then, we present **LCT** (§3.2), a trajectory-aware algorithm for realistic reversal and long-horizon simulation. The overall procedure is summarized in Algorithm 1.

3.1 M^{2}SIM FRAMEWORK

To model public mood dynamics at scale, we introduce **M^{2}Sim**, where the mean field captures the macro-level state distribution and a state-transition mechanism models micro-level agent dynamics.

MACRO-LEVEL MEAN FIELD

Recall from §2.1 that the macroscopic mean field at time t is given by the pair

$$(r_t, m_t), \quad r_t \in \mathbb{T},\ m_t \in \Delta^{m-1},$$

where $\mathbb{T}$ denotes the space of textual summaries and Δ^{m-1} is the probability simplex over the state space $\mathcal{S}$. We view these as two synchronized channels:

- **Textual synopsis r_t:** a free-form natural-language summary of recent trajectories, in the same spirit as the memory string in *MF-LLM*. It aggregates salient information about past micro-level interactions and exogenous signals into a compact macro-level description.
- **Distributional synopsis m_t:** a macro-level distribution over $\mathcal{S}$ that retains accumulated opinion-state information from past participation and incorporates updates from current participants. It explicitly tracks the relative proportions of latent opinion states. When $|\mathcal{S}|$ is large, semantically related states may be merged into coarse buckets to keep m_t low-dimensional.

Recursive update. At each time step $t \geq 1$, the macroscopic signal (r_{t-1}, m_{t-1}) and exogenous sianal c_t drive the evolution of internal states and the public mood.

First, f_ψ generates current participants' states and updates the distributional mean field:

$$\big(\vec{s}_t,\ m_t\big) \ = \ f_\psi\big(m_{t-1},\ r_{t-1},\ \vec{p}_t,\ c_t\big), m_t \in \Delta^{m-1}, \tag{3}$$

where $n_t \in \{0, \ldots, N\}$ denotes the number of active participants at time t, and $\vec{s}_t = [s_{1,t}, \ldots, s_{n_t,t}]$ and $\vec{p}_t = [p_{1,t}, \ldots, p_{n_t,t}]$ are their opinion states and personality descriptors, respectively.

After actions $\vec{a}_t = [a_{1,t}, \ldots, a_{n_t,t}]$ have been realised at time t, the textual synopsis is updated by the *MF-LLM* summariser μ_θ:

$$r_t \ = \ \mu_\theta\big(r_{t-1},\ m_t,\ \vec{a}_t,\ c_t\big). \tag{4}$$

In the degenerate case where internal states and the distributional mean field are ignored, the effective state collapses to the textual synopsis r_t. Concretely, dropping $\vec{s}_t$ and m_t, we update r_t only from $(r_{t-1}, \vec{a}_t, c_t)$ and condition each participant's action on $(r_{t-1}, p_{i,t}, c_t)$. This recursion coincides with macro-only two-LLM mean-field simulators, where a single textual mean field both conditions agent behaviour and summarizes collective trajectories.

Warm-up phase. Following *MF-LLM*, we employ a short warm-up phase of length t_{warm} in which the textual synopsis r_t is updated using ground-truth trajectories $\mathbf{a}_t^*$ and corresponding states, providing a stable initial macro description. Unlike the baseline, we *also* initialize the distributional channel from empirical data: m_0 is set to the historical state frequency over $\mathcal{S}$ rather than a uniform prior. This ensures that the mean field starts from a realistic macro configuration before simulated dynamics take over.

MICRO-LEVEL AGENTS

Given the macroscopic mean field (r_{t-1}, m_t), each agent acts based on its opinion state, personality, and macro-level information. For agent i, we denote this combined decision context by $z_{i,t} = (s_{i,t}, r_{t-1}, m_t, c_t, p_{i,t}) \in \mathcal{Z}$.

All agents share the same action space $\mathcal{A}$ and follow a stochastic policy

$$\pi_\phi(a_{i,t} \mid z_{i,t}) \;=\; \pi_\phi\big(a_{i,t} \mid s_{i,t},\, r_{t-1},\, m_t,\, c_t,\, p_{i,t}\big) \tag{5}$$

which maps the current decision context to a distribution over actions, capturing that decisions are shaped jointly by micro-level opinion and personality $(s_{i,t}, p_{i,t})$ and macro-level information (r_{t-1}, m_t, c_t).

Temporal evolution is not modelled via identity-specific kernels, but through the macro-level transition encoded in the mean field. At each time step, the state model f_ψ generates current participants' states and updates m_t from the previous macroscopic mean field, participant personalities, and the exogenous signal, retaining historical opinion-state information. Current participants act according to π_ϕ in equation 5. This recursive coupling between (r_{t-1}, m_t, c_t) and agent policies links micro-level actions with macro-level dynamics; in the degenerate case where internal states and the distributional mean field are ignored, actions depend only on $(r_{t-1}, p_{i,t}, c_t)$ and textual updates on $(r_{t-1}, \vec{a}_t, c_t)$, recovering prior macro-only two-LLM simulators as a special case.

Environment transition dynamics. We summarize the updates within step t and the preparation for step $t+1$ through an environment transition. In M^{2}Sim, this transition jointly describes updates to the distributional channel, the textual channel, the exogenous signal, and the active agent pool:

$$\big(m_t,\; r_t,\; \vec{p}_{t+1},\; c_{t+1}\big) \;=\; \mathcal{T}\big(\vec{p}_t,\; m_{t-1},\; r_{t-1},\; \vec{a}_t,\; c_t\big). \tag{6}$$

3.2 Long-Horizon Consistency Training

While *MF-LLM* (the textual synopsis r_t) provides a scalable macroscopic summary, its fidelity can be improved from both macro and micro perspectives. We propose **LCT** to optimize (1) *macro-level:* the state transition model, to extract predictive macroscopic signals; and (2) *micro-level:* the policy model, to generate behaviorally realistic actions conditioned on these signals.

State Transition Model

We model macro dynamics as learning a conditional law over *mean-field sequences*. Let $m_t \in \Delta^{m-1}$ denote the distributional mean field at step t, and define the event-level context

$$x_t := (r_{t-1},\, m_{t-1},\, \vec{p}_t,\, c_t).$$

We instantiate the mean-field state transition model f_ψ as a temporal sequence Transformer that maps the context history to a prediction of the current mean field:

$$\hat{m}_t \;=\; f_\psi\big(x_{1:t}\big) \in \Delta^{m-1},$$

where self-attention provides a flexible mechanism to capture delayed evidence accumulation by selectively attending to relevant past contexts.

Given empirical trajectories $\{m_t^\star\}_{t=1}^T$, we train f_ψ by matching the *entire* predicted sequence to the observed sequence using a KL-based sequence loss:

$$\mathcal{L}_{\text{seq}}(\psi) \;=\; \mathbb{E}\left[\sum_{t=1}^{T} \mathrm{KL}(m_t^\star \,\|\, \hat{m}_t)\right], \hat{m}_t = f_\psi(x_{1:t}). \tag{7}$$

Since simulation quality depends on multi-step evolution, we further impose a *rollout consistency* objective over a short horizon K. Starting from the empirical state $m_t^\star$, we generate a K-step rollout by recursively feeding the model with its own predictions:

$$\hat{m}_{t+k} = f_\psi\big(r_{t+k-1},\, \hat{m}_{t+k-1},\, \vec{p}_{t+k},\, c_{t+k}\big), \tag{8}$$

where the rollout is initialized at $\hat{m}_t = m_t^\star$ and recursively applied for $k = 1, \ldots, K$.

We then penalize the discrepancy between the rolled-out predictions and the observed future distributions:

$$\mathcal{L}_{\text{roll}}(\psi) \;=\; \mathbb{E}\left[\sum_{t=1}^{T-K}\sum_{k=1}^{K} \mathrm{KL}\Big(m_{t+k}^\star \,\|\, \hat{m}_{t+k}^{(k)}\Big)\right]. \tag{9}$$

The final training objective combines sequence fitting and rollout consistency,

$$\mathcal{L}_{\text{trans}}(\psi) \;=\; \mathcal{L}_{\text{seq}}(\psi) \;+\; \alpha\, \mathcal{L}_{\text{roll}}(\psi), \tag{10}$$

encouraging f_ψ to both match observed trajectories and remain stable under short self-conditioned rollouts, which is crucial for reproducing realistic delayed-commitment and opinion reversal.

POLICY MODEL OPTIMIZATION

The policy model π_ϕ governs microscopic decision-making conditioned on private states and the macroscopic mean-field signal. We adopt a factorized structure over the active agent set at step t,

$$\pi_\phi(\vec{a}_t \mid \vec{z}_t) \;=\; \prod_{i=1}^{n_t} \pi_\phi(a_{i,t} \mid z_{i,t}), \tag{11}$$

which is scalable while allowing heterogeneous behaviors through agent-specific inputs.

Long-horizon action reselection with dropout policy sampling. Direct RL for LLM policies is often bottlenecked by action sampling, since exploring diverse behaviors requires many autoregressive rollouts. To enable efficient long-horizon optimization, we shift exploration from token trajectories to *latent policy instances* by sampling dropout subnetworks. Concretely, we sample a dropout variable λ (implemented by in-place dropout in the forward pass), which induces a conditional policy $\pi_\phi(\cdot \mid \lambda)$. We then score each λ using a long-horizon mean-field surrogate computed via the state transition model, avoiding explicit text rollouts. The full derivation is provided in Appendix B.2.

Training objective. We define a long-horizon cost for a candidate policy instance as

$$V_\phi(\vec{a}, \lambda) = \sum_{k=1}^{K} \gamma^{k-1} \mathrm{KL}\big(m^\star_{t+k} \,\|\, \hat{m}_{t+k}(\pi_\phi(\vec{a}), \lambda)\big)\,, \tag{12}$$

where K is the rollout horizon and $\gamma \in (0, 1]$ is the discount factor. We optimize a weighted prediction loss $\mathcal{L}_{\mathrm{pred}}$ using soft weights over dropout samples:

$$\mathcal{L}_{\mathrm{pred}} = \sum_{t=1}^{T} \sum_{i=1}^{n_t} \mathbb{E}_{\lambda \sim p_{\mathrm{trivial}}(\lambda)} \Big[w_{i,t}(\lambda)\, V_\phi(\vec{a}_{i,t}, \lambda) \Big], \tag{13}$$

where $\vec{a}_{i,t} \sim \pi_\phi(\cdot \mid \lambda)$ denotes the candidate action of agent i at step t under dropout sample λ. Here $p_{\mathrm{trivial}}(\lambda)$ is a tractable reference distribution for dropout sampling, and $w_{i,t}(\lambda)$ is the soft weight over sampled λ, computed as a softmax on the corresponding long-horizon cost.

To stabilize training, we add an auxiliary text-supervision loss

$$\mathcal{L}_{\mathrm{text}}(\phi) = -\mathbb{E}_{\lambda \sim p_{\mathrm{trivial}}(\lambda)} \left[\sum_{t=1}^{T} \sum_{i=1}^{n_t} \log \pi_\phi\big(a^\star_{i,t} \mid \lambda\big) \right]. \tag{14}$$

The final objective is

$$\mathcal{L}_{\mathrm{total}}(\phi) = \mathcal{L}_{\mathrm{pred}}(\phi) + \alpha\, \mathcal{L}_{\mathrm{text}}(\phi). \tag{15}$$

4 EXPERIMENT

4.1 SETTINGS

Details are in the Appendix C; here we list what is used.

Model. Our framework contains two trainable components. **(1) State transition model.** We use an event-level causal Transformer to predict the macro state distribution m_t from historical mean-field signals. **(2) Policy model.** Following *MF-LLM*, we adopt **Qwen2-1.5B** as the backbone and fine-tune it with LoRA, attaching a lightweight K-step predictor to support long-horizon action selection. **Training.** We train for 1 epoch with learning rate 1×10^{-5} (other detailed settings in Appendix C.1).

Dataset. We follow *MF-LLM* (Mi et al., 2025) and use the WEIBO corpus (Ma et al., 2016). To stress-test reversal and delayed-commitment dynamics, we additionally crawl reversal events from Weibo and Douyin and convert them into the same *MF-LLM* event-centric format. For WEIBO, we select test events with trajectory length > 1000 and evaluate both (i) the default rollout horizon and (ii) a long-horizon rollout over the full trajectory. For the reversal set we simulate full trajectories.

Table 1: **Comparison across simulation settings.** We report distributional and classification metrics under SHORT-HORIZON, LONG-HORIZON, and REVERSAL simulations. Lower is better (↓) for all metrics except Micro/Macro-F1 (↑). Values denote mean with improvement rate (%) relative to **Direct LLM**. **Bold** indicates the best and underline indicates the second best.

Method	KL Div.↓	Wass. Dist.↓	DTW↓	Macro F1↑	Micro F1↑	NLL Loss↓
SHORT-HORIZON (DEFAULT STEPS)						
Direct LLM	0.1101	0.1661	0.1578	0.5805	0.6953	4.0674
Social Retrieval	0.1051 (4.54%)	0.1578 (5.00%)	0.1500 (4.94%)	0.5829 (0.41%)	0.6905 (-0.69%)	3.9532 (2.81%)
MF-LLM	0.0492 (55.31%)	0.1062 (36.06%)	**0.0944 (40.18%)**	0.5861 (0.96%)	0.6975 (0.32%)	3.9336 (3.29%)
M^{2}Sim (Ours)	**0.0453 (58.86%)**	**0.1006 (39.43%)**	0.0995 (36.95%)	**0.5897 (1.58%)**	**0.7082 (1.86%)**	**3.9156 (3.73%)**
LONG-HORIZON (FULL STEPS)						
Direct LLM	1.8300	0.3746	0.4003	0.3922	0.5621	4.6141
Social Retrieval	1.6554 (9.54%)	0.3515 (6.17%)	0.3816 (4.67%)	0.4014 (2.35%)	0.5962 (6.07%)	4.5294 (1.84%)
MF-LLM	1.2490 (31.75%)	0.3251 (13.21%)	0.2886 (27.91%)	0.4643 (18.38%)	0.6153 (9.46%)	3.9198 (15.05%)
M^{2}Sim (Ours)	**0.3089 (83.12%)**	**0.1773 (52.67%)**	**0.1666 (58.38%)**	**0.4805 (22.51%)**	**0.6393 (13.73%)**	**3.7655 (18.39%)**
REVERSAL (FULL STEPS)						
Direct LLM	5.4883	0.4535	0.4377	0.3531	0.4837	4.6469
Social Retrieval	3.5539 (35.26%)	0.4432 (2.27%)	0.3978 (9.12%)	0.3690 (4.50%)	0.5017 (3.72%)	4.2002 (9.61%)
MF-LLM	1.6425 (70.08%)	0.2763 (39.06%)	0.2425 (44.59%)	0.4158 (17.76%)	0.5721 (18.27%)	3.8841 (16.41%)
M^{2}Sim (Ours)	**0.5434 (90.10%)**	**0.2127 (53.10%)**	**0.1986 (54.63%)**	**0.4533 (28.38%)**	**0.6065 (25.39%)**	**3.8384 (17.40%)**

Evaluation Metrics. We adopt a micro-to-macro evaluation protocol. (i) Micro level, we annotate each individual action into one of 8 evaluation dimensions. (ii) Macro level, we compute the action distribution over n_t actions at each timestep t and report (1) **KL Divergence**, (2) **Wasserstein Distance**, (3) **DTW**, (4) **NLL**, (5) **Macro-F1**, and (6) **Micro-F1**. Details are provided in Appendix C.2.

Baselines. We compare M^{2}Sim with representative LLM-based baselines: (1) *Direct LLM (Gao et al., 2023; Park et al., 2023)*, which conditions a vanilla LLM only on the individual profile and event topic/context; (2) *Social Retrieval (Piao et al., 2025; Zhang et al., 2025b)*, which augments Direct LLM with retrieved peer responses (the k most recent and k most popular comments); (3) *MF-LLM (Mi et al., 2025)*, the two-LLM mean-field simulator with a textual synopsis for summarization.

4.2 COMPARISON WITH BASELINES

The main results are summarized in Table 1 (full in Appendix D.1); we make three observations.

(1) Strong short-horizon performance. Under SHORT-HORIZON (DEFAULT STEPS) simulation, *MF-LLM* already achieves strong distributional and classification accuracy, indicating that a textual mean field is sufficient for capturing near-term macro trends. M^{2}Sim yields only a modest but consistent improvement on top of MF-LLM, suggesting that explicit state distributions and stateful dynamics mainly contribute beyond the short-horizon regime.

(2) Long-horizon robustness. Under LONG-HORIZON (FULL STEPS) simulation, all baselines, including *Direct LLM*, *Social Retrieval*, and *MF-LLM*, degrade, reflecting compounding errors when the rollout extends beyond the default window. While *MF-LLM* improves over *Direct LLM* on distributional metrics (e.g., KL $1.8300 \rightarrow 1.2490$, **31.7%** reduction; DTW $0.4003 \rightarrow 0.2886$, **27.9%** reduction), its long-rollout drift remains pronounced. In contrast, M^{2}Sim remains stable and improves over *MF-LLM* (KL $1.2490 \rightarrow 0.3089$, **75.3%** reduction; DTW $0.2886 \rightarrow 0.1666$, **42.3%** reduction), demonstrating stronger long-horizon consistency in tracking collective trajectories.

(3) Reversal opinion dynamics. The reversal setting is the most challenging because it requires models to capture non-monotonic trend changes and delayed commitment. *MF-LLM* improves over SOCIAL RETRIEVAL on trajectory metrics (e.g., KL $3.5539 \rightarrow 1.6425$, **53.8%** reduction), yet its reversal tracking remains noticeably misaligned. M^{2}Sim further strengthens reversal fidelity (KL $1.6425 \rightarrow 0.5434$, **66.9%** reduction; DTW $0.2425 \rightarrow 0.1986$, **18.1%** reduction) and improves classification accuracy (e.g., Micro-F1 $0.5721 \rightarrow 0.6065$, **6.0%** increase), indicating that explicit distributional signals and long-horizon consistency training better support reversal dynamics over extended timelines.

Table 2: **Ablation study on M^{2}Sim.** We evaluate the contribution of LCT-State, LCT-Policy and Sampling under SHORT-HORIZON, LONG-HORIZON, and REVERSAL simulations. Green = improvement; Red = degradation; **Bold** = best.

Method	KL Div.↓	Wass. Dist.↓	DTW↓	Macro F1↑	Micro F1↑	NLL Loss↓
SHORT-HORIZON (DEFAULT STEPS)						
M^{2}Sim (Ours)	0.0453	**0.1006**	0.0995	0.5897	**0.7082**	**3.9156**
w/o LCT-State	0.0449 (+0.88%)	0.1120 (-11.33%)	**0.0802** (+19.40%)	0.5825 (-1.22%)	0.6828 (-3.59%)	4.0064 (-2.32%)
w/o LCT-Policy	0.0491 (-8.39%)	0.1051 (-4.47%)	0.0856 (+13.97%)	0.5809 (-1.49%)	0.6962 (-1.69%)	3.9403 (-0.63%)
w/o Sampling	0.0649 (-43.27%)	0.1115 (-10.83%)	0.1081 (-8.64%)	**0.5925** (+0.47%)	0.7014 (-0.96%)	3.9206 (-0.13%)
LONG-HORIZON (FULL STEPS)						
M^{2}Sim (Ours)	**0.3089**	**0.1773**	**0.1666**	**0.4805**	**0.6393**	**3.7655**
w/o LCT-State	0.8770 (-183.91%)	0.2123 (-19.74%)	0.2104 (-26.29%)	0.4725 (-1.66%)	0.6187 (-3.22%)	3.9414 (-4.67%)
w/o LCT-Policy	0.5872 (-90.09%)	0.2028 (-14.38%)	0.1730 (-3.84%)	0.4314 (-10.22%)	0.5765 (-9.82%)	3.9738 (-5.53%)
w/o Sampling	0.3536 (-14.47%)	0.1885 (-6.32%)	0.1710 (-2.64%)	0.4482 (-6.72%)	0.5774 (-9.68%)	3.7782 (-0.34%)
REVERSAL (FULL STEPS)						
M^{2}Sim (Ours)	**0.5434**	**0.2127**	0.1986	0.4533	0.6065	**3.8384**
w/o LCT-State	1.1748 (-116.19%)	0.2536 (-19.23%)	0.2319 (-16.77%)	**0.4649** (+2.56%)	**0.6232** (+2.75%)	4.0610 (-5.80%)
w/o LCT-Policy	0.9359 (-72.23%)	0.2396 (-12.65%)	**0.1808** (+8.96%)	0.4309 (-4.94%)	0.5592 (-7.80%)	4.2301 (-10.20%)
w/o Sampling	0.7634 (-40.49%)	0.2281 (-7.24%)	0.1897 (+4.48%)	0.4214 (-7.04%)	0.5721 (-5.67%)	4.0404 (-5.26%)

4.3 ABLATION STUDY

We ablate **LCT-State**, **LCT-Policy**, and **Sampling** to assess their respective contributions across SHORT-HORIZON, LONG-HORIZON, and REVERSAL simulations (Table 2).

(1) LCT-State is critical for long-horizon distribution tracking. Removing LCT-State yields the largest degradation on distributional metrics as the horizon extends. In LONG-HORIZON simulation, **w/o LCT-State** sharply worsens KL ($0.3089 \rightarrow 0.8770$) and DTW ($0.1666 \rightarrow 0.2104$), indicating compounding errors in multi-step macro rollouts. Similarly, in REVERSAL simulation KL rises markedly ($0.5434 \rightarrow 1.1748$), showing that state-consistent mean-field evolution is crucial for non-monotonic dynamics.

(2) Sampling in LCT-Policy is the main driver of F1. Ablating LCT-Policy degrades classification accuracy, especially under long horizons. In LONG-HORIZON simulation, **w/o LCT-Policy** reduces Macro-F1 ($0.4805 \rightarrow 0.4314$) and Micro-F1 ($0.6393 \rightarrow 0.5765$), and in REVERSAL it further lowers Micro-F1 ($0.6065 \rightarrow 0.5592$). Notably, removing **Sampling** alone causes a comparable F1 drop (Micro-F1 $0.6393 \rightarrow 0.5774$ in LONG-HORIZON, $0.6065 \rightarrow 0.5721$ in REVERSAL), indicating that sampling-based long-horizon selection drives most of the F1 gains.

(3) Complementarity and a short-horizon trade-off. Overall, M^{2}Sim is the most balanced across settings, combining state-level rollout consistency with sampling-based long-horizon action selection. Under SHORT-HORIZON, **w/o LCT-State** and **w/o LCT-Policy** can slightly improve distance metrics (e.g., DTW), since short rollouts favor local curve fitting, while LCT constraints mainly benefit long-horizon stability and reversals. Notably, **w/o LCT-State** slightly improves REVERSAL F1 (Macro-F1 $0.4533 \rightarrow 0.4649$, Micro-F1 $0.6065 \rightarrow 0.6232$) despite worse tracking, indicating per-step label fitting at the expense of coherent multi-step macro evolution.

4.4 ANALYSIS AND DISCUSSION

We analyze why M^{2}Sim improves long-horizon stability and reversal fidelity through a key limitation of *MF-LLM*: its macro signal is *implicit and narrative-smooth*, preserving coherence but failing to reliably control the *amount* and *timing* of state mass that drives micro actions. M^{2}Sim addresses this by representing the macro state as an *explicit distribution* and feeding it into agent inputs, translating macro shifts into consistent micro reallocation. Additional analyses are provided in Appendix D.2.

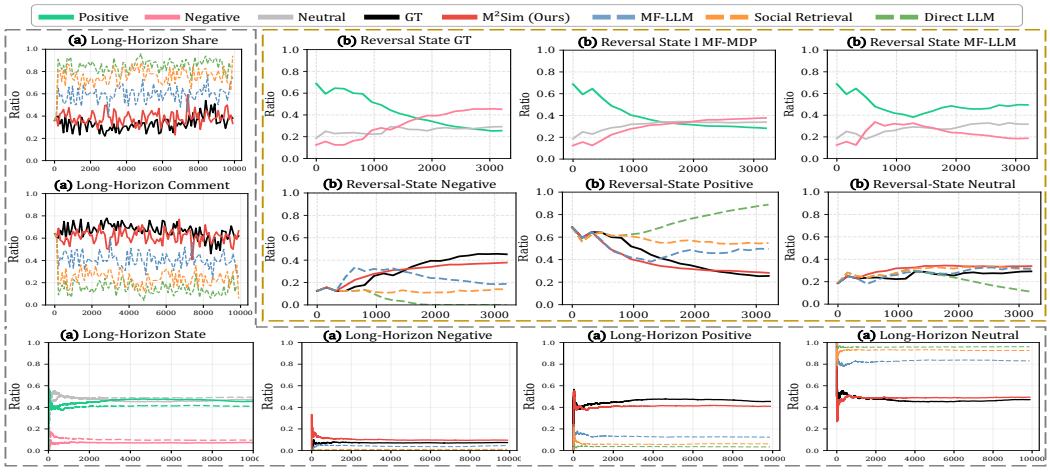

Figure 2: **Trajectory alignment of collective actions and state distributions.** We compare **Real data** with **M²Sim (Ours)**, *MF-LLM*, *Social Retrieval*, and *Direct LLM*. (a) In long-horizon simulation, **M²Sim** better matches action selection (e.g., comment and repost shares) and yields a closer fit to the state-distribution trajectory, while baselines drift more. (b) In a reversal event, **M²Sim** captures the macro-level state reversal, whereas *MF-LLM* fails to reproduce the turning point.

Long-horizon simulation. As shown in Fig. 2a, *MF-LLM*'s textual synopsis may describe "the crowd is cooling down" or "debate is intensifying" without specifying the mass of *Neutral* versus *Positive/Negative* states. Different interpretations yield different action mixtures, with deviations compounding over long horizons. Baselines tend toward a conservative regime with high neutrality, increasing repost share and distorting comment share. Although *MF-LLM* reduces drift relative to *Direct LLM* and *Social Retrieval*, it lacks precise control over neutrality. M²Sim predicts an explicit macro distribution and injects it into each agent's conditioning state: lower Neutral mass reduces neutral tendency and repost propensity, reallocating probability toward commenting and bringing action shares closer to real trajectories. Training against future distributions keeps the macro trajectory aligned, limiting drift that a plausible textual synopsis alone may conceal.

Reversal simulation. As shown in Fig. 2b, reversal requires the macro signal to *change direction* and propagate to micro decisions with the correct delay. *MF-LLM*'s temporally smooth, lexically consistent synopsis often remains "moderate" despite fluctuations, failing to express a decisive shift across state boundaries. The micro policy consequently follows the earlier storyline, producing inertia and missing the turning point. M²Sim's distributional state can represent sharp redistribution, such as Neutral collapsing and Negative rising, without a gradual narrative bridge; agents directly observe this distribution, allowing the action mixture to pivot accordingly. Long-horizon consistency training further downweights candidates whose rollouts remain "middle-of-the-road" when the real future requires reversal and upweights those producing the correct non-monotonic trajectory. Together, these mechanisms reproduce the macro-level *direction change* and subsequent micro-level *behavioral switch* with correct temporal alignment.

5 Conclusion

We study large-scale social network simulation through the coupling of micro-level behaviors and macro-level collective dynamics. Prior two-LLM mean-field simulators rely on macro-only updates, obscuring switching readiness and favoring self-reinforcing trajectories that miss opinion reversals. We propose **M²Sim**, a stateful mean-field simulator that couples latent opinion states of current participants with an evolving macro distribution through a learned state transition model. Long-horizon trajectory agreement guides action reselection, linking current decisions to future collective dynamics. Across real-world events, M²Sim improves short-horizon fidelity, long-horizon stability, and reversal tracking while reducing simulation drift.

AI USE STATEMENT

We used generative AI tools to assist with language editing, translation, and checking the clarity and consistency of mathematical notation in the manuscript. We also used LLMs as components of the simulation framework and for semantic annotation in evaluation, as described in the methodology and experimental settings. The authors reviewed and revised the AI-assisted text and mathematical expressions to ensure that they accurately reflected the intended methodology. We take responsibility for the final content of this work, including its methods, claims, results, and AI-assisted content.

ETHICS STATEMENT

This work studies opinion dynamics in social networks to support understanding of collective behavior and evaluation of interventions in simulated settings. These techniques could also be misused for targeted persuasion, misinformation amplification, or manipulation at scale. We frame M^2Sim as a research and analytical tool and encourage applications with appropriate ethical review, transparency, and privacy safeguards. Simulated opinion states should not be interpreted as verified psychological attributes of real individuals. Moreover, LLM-generated behaviors and semantic annotations may contain biases or errors that affect the simulated trajectories and their evaluation. Although we evaluate predictions using explicit semantic dimensions and quantitative trajectory metrics, these measures do not eliminate annotation uncertainty or establish the effectiveness of interventions in real-world settings.

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.

Robert Axtell. *Why agents?: on the varied motivations for agent computing in the social sciences*, volume 17. Center on Social and Economic Dynamics Washington, DC, 2000.

Dario Bauso, Hamidou Tembine, and Tamer Basar. Opinion dynamics in social networks through mean-field games. *SIAM Journal on Control and Optimization*, 54(6):3225–3257, 2016.

Youssef Bouanan, Gregory Zacharewicz, Judicael Ribault, and Bruno Vallespir. Discrete Event System Specification-Based Framework for Modeling and Simulation of Propagation Phenomena in Social Networks: Application to the Information Spreading in a Multi-Layer Social Network. *Simulation*, 95(5):411–427, 2019.

Romain Couillet, Samir M Perlaza, Hamidou Tembine, and Mérouane Debbah. Electrical vehicles in the smart grid: A mean field game analysis. *IEEE Journal on Selected Areas in Communications*, 30(6):1086–1096, 2012.

Paul Davidsson. Agent-Based Social Simulation: A Computer Science View. *Journal of Artificial Societies and Social Simulation*, 5(1), 2002.

Joshua M Epstein and Robert Axtell. *Growing Artificial Societies: Social Science from the Bottom Up*. Brookings Institution Press, 1996.

Linton Freeman. The development of social network analysis. *A Study in the Sociology of Science*, 1(687):159–167, 2004.

Chen Gao, Xiaochong Lan, Zhihong Lu, Jinzhu Mao, Jinghua Piao, Huandong Wang, Depeng Jin, and Yong Li. $S\hat{3}$: Social-network Simulation System with Large Language Model-Empowered Agents. *arXiv preprint arXiv:2307.14984*, 2023.

Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanities and Social Sciences Communications*, 11(1):1–24, 2024.

Nicolas Gast and Benny Van Houdt. A refined mean field approximation. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 1(2):1–28, 2017.

Volker Grimm, Eloy Revilla, Uta Berger, Florian Jeltsch, Wolf M Mooij, Steven F Railsback, Hans-Hermann Thulke, Jacob Weiner, Thorsten Wiegand, and Donald L DeAngelis. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *science*, 310(5750):987–991, 2005.

Davide Grossi, Frank Dignum, Mehdi Dastani, and Lamber Royakkers. Foundations of Organizational Structures in Multiagent Systems. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 690–697, 2005.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645(8081):633–638, 2025.

Dirk Helbing, Illés Farkas, and Tamas Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000.

Qingchao Kong, Wenji Mao, Guandan Chen, and Daniel Zeng. Exploring trends and patterns of popularity stage evolution in social media. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(10):3817–3827, 2018.

Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese journal of mathematics*, 2 (1):229–260, 2007.

Mathieu Lauriere, Sarah Perrin, Sertan Girgin, Paul Muller, Ayush Jain, Theophile Cabannes, Georgios Piliouras, Julien Pérolat, Romuald Elie, Olivier Pietquin, et al. Scalable deep reinforcement learning algorithms for mean field games. In *International conference on machine learning*, pp. 12078–12095. PMLR, 2022.

Mei Li, Xiang Wang, Kai Gao, and Shanshan Zhang. A survey on information diffusion in online social networks: Models and methods. *Information*, 8(4):118, 2017.

Yijun Liu, Wu Liu, Xiaoyan Gu, Allen He, Weiping Wang, and Yongdong Zhang. Popsim: Social network simulation for social media popularity prediction. *arXiv preprint arXiv:2512.02533*, 2025.

Peng Lu and Shizhao Nie. The strength distribution and combined duration prediction of online collective actions: Big data analysis and bp neural networks. *Physica A: Statistical Mechanics and its Applications*, 535:121031, 2019.

Peng Lu, Shizhao Nie, Zheng Wang, Ziwei Jing, Jianwu Yang, Zhongxiang Qi, and Wangmo Pujia. Predicting durations of online collective actions based on peaks' heights. *Communications in Nonlinear Science and Numerical Simulation*, 55:338–354, 2018a.

Peng Lu, Zheng Wang, Shizhao Nie, Wangmo Pujia, Pengfei Lu, and Baosheng Chen. Exploring the participate propensity in cyberspace collective actions: The 5‰ rule. *Physica A: Statistical Mechanics and Its Applications*, 503:582–590, 2018b.

Peng Lu, Hou Yang, Hao Li, Mengdi Li, and Zhuo Zhang. Swarm intelligence, social force and multi-agent modeling of heroic altruism behaviors under collective risks. *Knowledge-Based Systems*, 214:106725, 2021.

Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pp. 3818–3824. AAAI Press, 2016. ISBN 9781577357704.

Dennis L Meadows, William W Behrens, Donella H Meadows, Roger F Naill, Jørgen Randers, and Erich Zahn. *Dynamics of Growth in a Finite World*, volume 360. Wright-Allen Press, Cambridge, MA, 1974.

Qirui Mi, Siyu Xia, Yan Song, Haifeng Zhang, Shenghao Zhu, and Jun Wang. Taxai: A dynamic economic simulator and benchmark for multi-agent reinforcement learning. *arXiv preprint arXiv:2309.16307*, 2023.

Qirui Mi, Zhiyu Zhao, Siyu Xia, Yan Song, Jun Wang, and Haifeng Zhang. Learning macroeconomic policies based on microfoundations: A stackelberg mean field game approach. Technical report, 2024.

Qirui Mi, Mengyue Yang, Xiangning Yu, Zhiyu Zhao, Cheng Deng, Bo An, Haifeng Zhang, Xu Chen, and Jun Wang. Mf-llm: Simulating population decision dynamics via a mean-field large language model framework. *arXiv preprint arXiv:2504.21582*, 2025.

Shizhao Nie, Zheng Wang, Wangmo Pujia, Yuan Nie, and Peng Lu. Big data prediction of durations for online collective actions based on peak's timing. *Physica A: Statistical Mechanics and its Applications*, 492:138–154, 2018.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–22, 2023.

Joon Sung Park, Carolyn Q. Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S. Bernstein. Generative agent simulations of 1,000 people, 2024. URL https://arxiv.org/abs/2411.10109.

Jinghua Piao, Yuwei Yan, Jun Zhang, Nian Li, Junbo Yan, Xiaochong Lan, Zhihong Lu, Zhiheng Zheng, Jing Yi Wang, Di Zhou, et al. Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society. *arXiv preprint arXiv:2502.08691*, 2025.

Thomas C Schelling. Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1(2): 143–186, 1971.

Thomas C Schelling. *Micromotives and Macrobehavior*. WW Norton & Company, 2006.

Flaminio Squazzoni. The Micro-Macro Link in Social Simulation. *Sociologica*, 2(1):3–5, 2008.

Flaminio Squazzoni, Wander Jager, and Bruce Edmonds. Social Simulation in the Social Sciences: A Brief Overview. *Social Science Computer Review*, 32(3):279–294, 2014.

Toshiyuki Tanaka. A theory of mean field approximation. *Advances in Neural Information Processing Systems*, 11, 1998.

Leigh Tesfatsion. Agent-based computational economics: A constructive approach to economic theory. *Handbook of computational economics*, 2:831–880, 2006.

Adelinde M Uhrmacher and Danny Weyns. *Multi-Agent Systems: Simulation and Applications*. CRC Press, 2018.

Lei Wang, Heyang Gao, Xiaohe Bo, Xu Chen, and Ji-Rong Wen. Yulan-onesim: Towards the next generation of social simulator with large language models. In *Workshop on Scaling Environments for Agents*, 2025.

Yao Wang, Chungang Yang, Tong Li, Xinru Mi, Lixin Li, and Zhu Han. A survey on mean-field game for dynamic management and control in space-air-ground network. *IEEE Communications Surveys & Tutorials*, 26(4):2798–2835, 2024.

Haoran Wei, Yaofeng Sun, and Yukun Li. Deepseek-ocr: Contexts optical compression. *arXiv preprint arXiv:2510.18234*, 2025.

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.

An Yang, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoyan Huang, Jiandong Jiang, Jianhong Tu, Jianwei Zhang, Jingren Zhou, et al. Qwen2. 5-1m technical report. *arXiv preprint arXiv:2501.15383*, 2025a.

Jiachen Yang, Xiaojing Ye, Rakshit Trivedi, Huan Xu, and Hongyuan Zha. Learning deep mean field games for modeling large population behavior. *arXiv preprint arXiv:1711.03156*, 2017.

Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, and Jun Wang. Mean field multi-agent reinforcement learning. In *International conference on machine learning*, pp. 5571–5580. PMLR, 2018.

Ziyi Yang, Zaibin Zhang, Zirui Zheng, Yuxian Jiang, Ziyue Gan, Zhiyu Wang, Zijian Ling, Jinsong Chen, Martz Ma, Bowen Dong, Prateek Gupta, Shuyue Hu, Zhenfei Yin, Guohao Li, Xu Jia, Lijun Wang, Bernard Ghanem, Huchuan Lu, Chaochao Lu, Wanli Ouyang, Yu Qiao, Philip Torr, and Jing Shao. Oasis: Open agent social interaction simulations with one million agents, 2025b. URL https://arxiv.org/abs/2411.11581.

Xiang Zhang, Juntai Cao, Jiaqi Wei, Chenyu You, and Dujian Ding. Why prompt design matters and works: A complexity analysis of prompt search space in llms. *arXiv preprint arXiv:2503.10084*, 2025a.

Xinglang Zhang, Yunyao Zhang, ZeLiang Chen, Junqing Yu, Wei Yang, and Zikai Song. Logical phase transitions: Understanding collapse in LLM logical reasoning. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 18836–18860, San Diego, California, United States, July 2026a. Association for Computational Linguistics. ISBN 979-8-89176-390-6. doi: 10.18653/v1/2026.acl-long.858. URL https://aclanthology.org/2026.acl-long.858/.

Xinnong Zhang, Jiayu Lin, Xinyi Mou, Shiyue Yang, Xiawei Liu, Libo Sun, Hanjia Lyu, Yihang Yang, Weihong Qi, Yue Chen, et al. Socioverse: A world model for social simulation powered by llm agents and a pool of 10 million real-world users. *arXiv preprint arXiv:2504.10157*, 2025b.

Yunyao Zhang, Zikai Song, Hang Zhou, Wenfeng Ren, Yi-Ping Phoebe Chen, Junqing Yu, and Wei Yang. $ga - s^3$: Comprehensive social network simulation with group agents. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 8950–8970, Vienna, Austria, July 2025c. Association for Computational Linguistics. ISBN 979-8-89176-256-5. doi: 10.18653/v1/2025.findings-acl.468. URL https://aclanthology.org/2025.findings-acl.468/.

Yunyao Zhang, Xinglang Zhang, Junxi Sheng, Wenbing Li, Junqing Yu, Yi-Ping Phoebe Chen, Wei Yang, and Zikai Song. Semantic-aware logical reasoning via a semiotic framework. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 18349–18374, San Diego, California, United States, July 2026b. Association for Computational Linguistics. ISBN 979-8-89176-390-6. doi: 10.18653/v1/2026.acl-long.835. URL https://aclanthology.org/2026.acl-long.835/.

Appendix

A. Related Work

B. Theoretical Proofs

C. Detailed Experimental Setup

D. Additional Experimental Results and Analysis

A Related Work

Social Simulation Systems: Foundations and Limitations

Traditional social simulation systems can be broadly classified into three paradigms. *(1) Mechanistic models* describe collective behavior through explicit equations or procedural dynamics such as discrete-event and system-dynamics (Lu et al., 2021; Meadows et al., 1974; Bouanan et al., 2019). *(2) Empirical and statistical models* identify diffusion regularities from data, including PSP (Kong et al., 2018) and peak-based participation dynamics (Lu et al., 2018a; Lu & Nie, 2019). *(3) Agent-based models* capture emergent phenomena from local interactions among heterogeneous agents (Davidsson, 2002; Epstein & Axtell, 1996; Squazzoni, 2008; Uhrmacher & Weyns, 2018), with applications to collective dynamics (Helbing et al., 2000), market simulations (Tesfatsion, 2006), ecosystems (Grimm et al., 2005), and public policy (Axtell, 2000; Mi et al., 2023). Despite their success, these paradigms often rely on handcrafted rules, simplified assumptions, and fixed parameters, which limits adaptability and makes robust long-horizon simulation difficult in open-ended, evolving settings. Recent advances in large language models (LLMs) (Zhang et al., 2026a) offer a complementary direction: replacing brittle rule templates with contextual, generative decision-making to better generalize while preserving rich semantic interactions.

LLM-Based Agent Social Simulation: Progress and Gaps

Recent advances in LLMs (Achiam et al., 2023; Guo et al., 2025; Wei et al., 2025; Yang et al., 2025a) have enabled cognitively enriched social simulations where agents communicate, reason, and adapt through natural language (Wei et al., 2022; Zhang et al., 2026b; 2025a). At a high level, existing LLM-driven agent simulators fall into two paradigms. *(1) Prompted role-play ABM.* These systems drive open-ended interactions mainly through role specifications and prompting (often with lightweight memory), and are typically used in small-scale or qualitative settings; examples include *Generative Agents* (Park et al., 2023; 2024) and S^3 (Gao et al., 2023). *(2) Memory / retrieval-augmented ABM.* These systems inject social context via heuristic retrieval or summarization (e.g., recent/popular responses) to improve scalability while retaining language-based decision making; examples include *SocioVerse* (Zhang et al., 2025b), *AgentSociety* (Piao et al., 2025), and *GA-*S^3 (Zhang et al., 2025c). Despite steady progress, most approaches remain *micro-centric* and depend on prompting or heuristic memory/retrieval to carry evolving context, which is often brittle and struggles to retain decision-critical information over long horizons, resulting in unstable temporal dynamics and weak quantitative alignment with real-world collective trends.

Mean Field Approximation: Motivation and Challenges

Mean field approximation (Tanaka, 1998; Gast & Van Houdt, 2017) scales large multi-agent systems by replacing expensive pairwise interactions with interactions between each agent and a shared macro signal. This idea is formalized in mean-field game (MFG) theory (Lasry & Lions, 2007; Yang et al., 2018), where individual decisions and aggregate dynamics are coupled through a compact mean-field representation, and has been applied to domains such as social influence (Yang et al., 2017; Bauso et al., 2016), traffic control (Wang et al., 2024), energy optimization (Couillet et al., 2012), and economic policy (Grimm et al., 2005; Mi et al., 2024). Despite this scalability, classical MFGs assume stylized behaviors and environment models, making it difficult to capture contextual, language-mediated decisions in realistic social settings; neural mean-field variants improve expressiveness (Lauriere et al., 2022) but remain limited when interactions are open-ended and semantics-rich. Recent LLM-based simulators introduce a shared macro signal (Mi et al., 2025; Liu et al.,

Algorithm 1 M^{2}Sim Simulation Procedure

Require: Active agent pool size $n_t \in N$, horizon T, warm-up length t_{warm}; state transition model f_ψ, *MF-LLM* summariser μ_θ, agent policy π_ϕ; exogenous sianals $\{c_t\}_{t=1}^{T}$

Initialize: (r_0, m_0), active pool descriptors $\vec{p}_1$
for $t = 1$ **to** T **do**
  // Update distributional channel (and latent states)
  $(\vec{s}_t, m_t) \leftarrow f_\psi(m_{t-1}, r_{t-1}, \vec{p}_t, c_t)$ ▷ Eq. equation 3
  if $t \le t_{\text{warm}}$ **then**
    // Warm-up: use background actions and states
    $\vec{a}_t \leftarrow \vec{a}_t^*; \quad m_t \leftarrow m_t^*$
  else
    // Simulation: sample actions by policy
    for all $i \in \{1, \dots, n_t\}$ **do** ▷ active subset at step t
      $z_{i,t} \leftarrow (s_{i,t}, r_{t-1}, m_t, c_t, p_{i,t})$
      $a_{i,t} \sim \pi_\phi(\cdot \mid z_{i,t})$ ▷ Eq. equation 5
    end for
    $\vec{a}_t \leftarrow [a_{1,t}, \dots, a_{n_t,t}]$
  // Update textual channel
  $r_t \leftarrow \mu_\theta(r_{t-1}, m_t, \vec{a}_t, c_t)$ ▷ Eq. equation 4
  // Environment transition (agent pool refresh)
  $\vec{p}_{t+1} \leftarrow \textsc{Refresh}(\vec{p}_t;\, t)$ ▷ induces Eq. equation 6
end for
Output: Sequence of simulated actions $\{\vec{a}_t\}_{t=0}^{T}$

2025), yet macro-micro coupling remains weak: *MF-LLM* lacks explicit latent micro states, while *PopSim* relies on prompt-only design. The key challenge is to learn a compact, decision-critical mean field that supports stateful decisions and long-horizon macro evolution without drifting from real trajectories.

B Theoretical Proofs

B.1 Derivation of Prop.2.1

Proof. Fix a mean-field distribution m and let $s^\star \in \arg\max_{s \in \mathcal{S}} m(s)$. By the law of iterated expectation and Assumption (1),

$$\mathbb{E}[m_{t+1}(s^\star) \mid m_t = m] = \mathbb{E}_{\mathbf{a}_t \mid m_t = m}\left[\mathbb{E}[m_{t+1}(s^\star) \mid m_t = m, \mathbf{a}_t]\right] \tag{16}$$

$$\geq \mathbb{E}_{\mathbf{a}_t \mid m_t = m}\left[\widehat{q}(\mathbf{a}_t)(s^\star) - \epsilon\right] \tag{17}$$

$$= \mathbb{E}[\widehat{q}(\mathbf{a}_t)(s^\star) \mid m_t = m] - \epsilon. \tag{18}$$

Applying Assumption (2) yields

$$\mathbb{E}[m_{t+1}(s^\star) \mid m_t = m] \geq m(s^\star) + \eta - \epsilon.$$

Therefore,

$$\delta_m(s^\star) = \mathbb{E}[m_{t+1}(s^\star) \mid m_t = m] - m(s^\star) \geq \eta - \epsilon.$$

In particular, if $\eta \geq \epsilon$, then $\delta_m(s^\star) \geq 0$. □

B.2 Derivation of the Policy Model

The policy model π_ϕ governs microscopic decision-making. We use a factorized policy over the active set:

$$\pi_\phi(\vec{a}_t \mid \vec{z}_t) \;=\; \prod_{i=1}^{n_t} \pi_\phi(a_{i,t} \mid z_{i,t}). \tag{19}$$

Lookahead re-selection as latent-policy optimization. Starting from

$$\max_\phi \mathbb{E}_{\vec{a} \sim \pi_\phi}[R(\vec{a})]\,, \tag{20}$$

introduce a latent variable $\lambda \sim q(\lambda)$ and the conditional policy $\pi_\phi(\vec{a} \mid \lambda)$:

$$\begin{aligned}\max_\phi \mathbb{E}_{\vec{a}\sim\pi_\phi}[R(\vec{a})] &= \int R(\vec{a})\pi_\phi(\vec{a})\, d\vec{a} \\ &= \int_{\vec{a}} R(\vec{a}) \int_\lambda \pi_\phi(\vec{a} \mid \lambda) q(\lambda)\, d\lambda\, d\vec{a} \\ &= \int_\lambda \Big(\int_{\vec{a}} R(\vec{a})\pi_\phi(\vec{a} \mid \lambda)\, d\vec{a} \Big) q(\lambda)\, d\lambda \\ &= \mathbb{E}_{\lambda\sim q(\lambda)} \Big[\mathbb{E}_{\vec{a}\sim\pi_\phi(\cdot|\lambda)}[R(\vec{a})] \Big] \\ &= \mathbb{E}_{\lambda\sim q(\lambda)} \big[\hat{R}_\phi(\lambda) \big],\end{aligned} \tag{21}$$

where

$$\hat{R}_\phi(\lambda) := \mathbb{E}_{\vec{a}\sim\pi_\phi(\cdot|\lambda)}[R(\vec{a}, \lambda)]. \tag{22}$$

To keep $q(\lambda)$ tractable while still preferring high-reward latent instances, we regularize it toward a trivial reference p_{trivial} (e.g., uniform over dropout masks), yielding a bi-level objective:

$$\max_\phi \max_q \mathbb{E}_{\lambda\sim q(\lambda)}[\hat{R}_\phi(\vec{a}, \lambda)] - \frac{1}{\beta}\mathrm{KL}(\mathrm{q}||\mathrm{p}_{\mathrm{trivial}}). \tag{23}$$

The inner maximization has the textbook solution

$$q^* \propto p_{\text{trivial}}(\lambda) \exp(\beta \hat{R}_\phi(\vec{a}, \lambda)). \tag{24}$$

Plugging q^* back produces a log-sum-exp form (we keep the original notation; $Z(\phi)$ is the corresponding normalizer):

$$\begin{aligned}&\max_\phi \mathbb{E}_{\lambda\sim q^*(\lambda)}[\hat{R}_\phi(\vec{a}, \lambda)] - \frac{1}{\beta}\mathrm{KL}(\mathrm{q}^*||\mathrm{p}_{\mathrm{trivial}}) \\ =&\max_\phi \mathbb{E}_{q^*}[\hat{R}_\phi(\vec{a}, \lambda)] - \frac{1}{\beta}(\beta \mathbb{E}_{q^*}[\hat{R}_\phi(\vec{a}, \lambda)] - \log^{Z(\phi)}) \\ =&\max_\phi \frac{1}{\beta} \log^{\mathbb{E}_{p_{\text{trivial}}}[\exp(\beta \hat{R}_\phi(\vec{a}, \lambda))]}.\end{aligned} \tag{25}$$

$$Z(\phi) = \int p_{\text{trivial}}(\lambda) \exp(\beta \hat{R}) d\lambda = \mathbb{E}_{p_{\text{trivial}}}[(\exp \beta \hat{R})]. \tag{26}$$

From maximization to a weighted cost. Rewriting the maximization as minimization with $V = -\hat{R}$ gives

$$\min_\phi \frac{1}{\beta} \log^{\mathbb{E}_{p_{\text{trivial}}}[\exp(\beta V_\phi(\vec{a}, \lambda))]}.$$

Taking gradients reveals that the objective induces a *softmax weighting* over λ:

$$\begin{aligned}\nabla_\phi \frac{1}{\beta} \log^{\mathbb{E}_{p_{\text{trivial}}}[\exp(\beta V_\phi(\vec{a}^*, \lambda))]} &= \frac{1}{\beta} \frac{1}{Z(\phi)} \nabla_\phi Z(\phi) \\ &= \frac{1}{\beta} \frac{1}{Z(\phi)} \mathbb{E}_p[\beta \exp(\beta V_\phi) \nabla_\phi V_\phi] \\ &= \frac{1}{Z(\phi)} \mathbb{E}_p[\exp(\beta V_\phi) \nabla_\phi V_\phi] \\ &= \mathbb{E}_p \left[\frac{\exp(\beta V_\phi)}{Z(\phi)} \nabla_\phi V_\phi \right].\end{aligned}$$

and integrating yields

$$\frac{1}{\beta} \log^{\mathbb{E}_{p_{\text{trivial}}}[\exp(\beta V_\phi(\vec{a}^*, \lambda))]} = \mathbb{E}_{p_{\text{trivial}}}[w V_\phi], \tag{27}$$

where

$$w(\vec{a}, \lambda) = \frac{\exp(\beta \mathrm{V}_\phi)}{\mathrm{Z}(\phi)}$$

is exactly a $\mathrm{softmax}$ over λ with temperature β. Therefore the final objective becomes

$$\min_{\phi} \mathbb{E}_{\lambda \sim p_{\mathrm{trivial}}}[w V_\phi(\vec{a}, \lambda)]. \tag{28}$$

Instantiating the long-horizon cost. We define the long-horizon cost as a discounted divergence:

$$V_\phi(\vec{a}, \lambda) = \sum_{k=1}^{K} \gamma^{k-1} \mathrm{KL}\big(m^\star_{t+k} \,\|\, \hat{m}_{t+k}(\pi_\phi(\vec{a}), \lambda)\big)\,. \tag{29}$$

Aggregating over simulation time t and active agents n_t yields the prediction loss

$$\mathcal{L}_{\mathrm{pred}} = \sum_{t=1}^{T} \sum_{i=1}^{n_t} \mathbb{E}_p[w_{i,t} \sum_{k=1}^{K} \gamma^{k-1} \mathrm{KL}\big(m^\star_{t+k} \,\|\, \hat{m}_{t+k}(\pi_\phi)\big)]. \tag{30}$$

To stabilize optimization, we include an auxiliary text-supervision term (ground-truth actions) to reduce variance:

$$\mathcal{L}_{\mathrm{text}}(\phi) = -\mathbb{E}\left[\sum_{t=1}^{T} \sum_{i=1}^{n_t} \log \pi_\phi\big(\vec{a}^\star_{i,t} \mid \lambda\big)\right]. \tag{31}$$

The following bound shows that increasing $\pi_\phi(\vec{a}^*|\lambda)$ tightens the variance of the induced cost.

Theorem B.1. *If $\vec{a}$ is a discrete random vector and $|V_\phi| \le M$, then* $\mathrm{Var}(V) \le 4M^2(1-\pi_\phi(\vec{a}^*|\lambda))^2$.

$$\begin{aligned}
\left|\hat{R}_\phi(\vec{a}, \lambda) - R(\vec{a}^*, \lambda)\right| &= \left|\sum_{\vec{a}} R_\phi(\vec{a}, \lambda)\pi_\phi(\vec{a} \mid \lambda) - R(\vec{a}^*, \lambda)\right| \\
&\le \sum_{\vec{a}} \pi_\phi(\vec{a} \mid \lambda)\, |R_\phi(\vec{a}, \lambda) - R(\vec{a}^*, \lambda)| \\
&\le 2M \sum_{\vec{a} \neq \vec{a}^*} \pi_\phi(\vec{a} \mid \lambda) \\
&= 2M\,(1 - \pi_\phi(\vec{a}^* \mid \lambda))\,.
\end{aligned}$$

$$\mathrm{Var}(\hat{R}) = \mathbb{E}\Big[(\hat{R} - \mathbb{E}[\hat{R}])^2\Big] \le \mathbb{E}\Big[(\hat{R} - R(\vec{a}^*, \lambda))^2\Big] \le 4M^2\big(1 - \pi_\phi(\vec{a}^* \mid \lambda)\big)^2. \tag{32}$$

Finally, we optimize π_ϕ via the two-term objective

$$\mathcal{L}_{\mathrm{total}}(\phi) = \mathcal{L}_{\mathrm{pred}}(\phi) + \alpha\, \mathcal{L}_{\mathrm{text}}(\phi). \tag{33}$$

In practice, λ is instantiated by in-place dropout sampling within the forward pass, so drawing $\lambda \sim p_{\mathrm{trivial}}$ corresponds to sampling dropout subnetworks, while the weights w implement a soft selection over these dropout policy instances.

C Detailed Experimental Setup

C.1 Training Curves and Hyperparameters

Summary. All experiments are run on two Tesla V100S-PCIE-32GB GPUs. Table 3 reports the hyperparameters for both the state transition model (Event Transformer) and all LLM components (Mean Field IB-Tune, Policy IB-Tune, and Policy LCT-Tune), with LLMs fine-tuned from `Qwen2-1.5B-Instruct` using LoRA on WEIBO under a unified data format and sequence length.

Training curves. Figure 3 shows the optimization dynamics of our two core modules. The *state transition model* is trained with the KL-based transition loss $\mathcal{L}_{\mathrm{trans}}$ against the empirical mean-field distribution, while the *policy model* is trained with $\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{pred}} + \alpha\, \mathcal{L}_{\mathrm{text}}$, combining discounted K-step rollout divergence and NLL supervision on ground-truth actions.

Table 3: Model and training hyperparameters. The left block reports the state transition model (Event Transformer) used in M^{2}Sim; the right block reports LLM-based components fine-tuned from Qwen2-1.5B-Instruct.

State Transition (Event Transformer)		**Hyperparameter**	**Mean Field**	**IB-Tune**	**LCT-Tune**
Hidden size d_{model}	256	Base model	Qwen2-1.5B	Qwen2-1.5B	Qwen2-1.5B
#layers / #heads	3 / 8	Max sequence length	2048	2048	2048
Max sequence length	4096	Training dataset	WEIBO	WEIBO	WEIBO
Dropout	0.1	Training batch size	256	256	256
FFN dimension d_{ff}	1024	Micro batch size	8	8	8
Text Encoder	BERT	Max epochs	1	1	1
Optimizer	AdamW	Random Seed	46	46	46
Learning rate	2×10^{-5}	Learning rate	5×10^{-7}	5×10^{-7}	5×10^{-7}
Batch size (events)	4	LoRA rank/alpha	64/64	64/64	64/64
Max epochs	20	Prediction weight α_{coeff}	–	–	0.5
Weight decay	1×10^{-5}	Lookahead horizon K	–	–	30
Gradient clip	1.0	#candidates J	–	–	4
Loss function	$\mathcal{L}_{\text{trans}}$	Loss function	$\mathcal{L}_{\text{mean-field}}$	$\mathcal{L}_{\text{policy}}$	$\mathcal{L}_{\text{LCT}}$

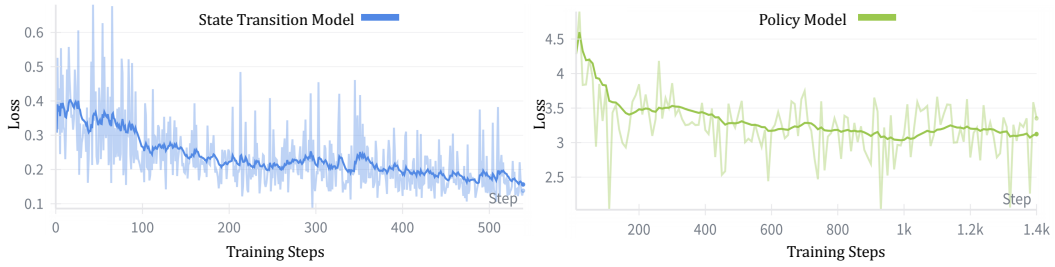

Figure 3: **Training loss curves for M^{2}Sim. Left:** Event Transformer (state transition model) optimized by $\mathcal{L}_{\text{trans}}$ (KL divergence between predicted and empirical mean-field distributions). **Right:** Policy model optimized by $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pred}} + \alpha\, \mathcal{L}_{\text{text}}$, combining discounted K-step rollout discrepancy with action NLL supervision.

C.2 Detailed Settings

Dataset. We follow *MF-LLM* (Mi et al., 2025) and adopt the WEIBO corpus (Ma et al., 2016) as the primary benchmark, which contains 5,000+ real-world events with temporally ordered individual responses and rich individual profiles, covering categories such as *Crime, Culture, Health, News, Politics, Sports,* and *Technology*. To better stress-test *reversal* and delayed-commitment dynamics beyond the original benchmark, we additionally curate a complementary **Reversal** collection by crawling public discussions from Weibo and Douyin, spanning domains including *Education, Economy, Society, Environment,* and *Campus*. We convert all newly collected data into the same *MF-LLM* event-centric format (event timeline, per-timestep active individual set, responses, and profiles), enabling plug-and-play training/evaluation under an identical simulation interface. The Reversal set contains much longer trajectories than WEIBO (up to 40,000+ timesteps), so we use full-length rollouts, as short horizons rarely exhibit reversals. For WEIBO, we evaluate long events (length > 1000) under both the default 300-step rollout (*MF-LLM*) and full-trajectory settings.

Evaluation Metrics. We adopt a micro-to-macro evaluation protocol:

(i) Micro-level (individual actions). We focus on *semantic-related* evaluation by annotating each generated (and real) action into one of 8 semantic dimensions using an LLM-based annotator (GPT-4o-mini). This LLM-based evaluation protocol follows prior mean-field simulators (*MF-LLM*) and has been empirically validated therein.

- **Rumor.** Whether the action *spreads* the discussed claim (believes/amplifies) or *counters* it (questions/refutes/clarifies).

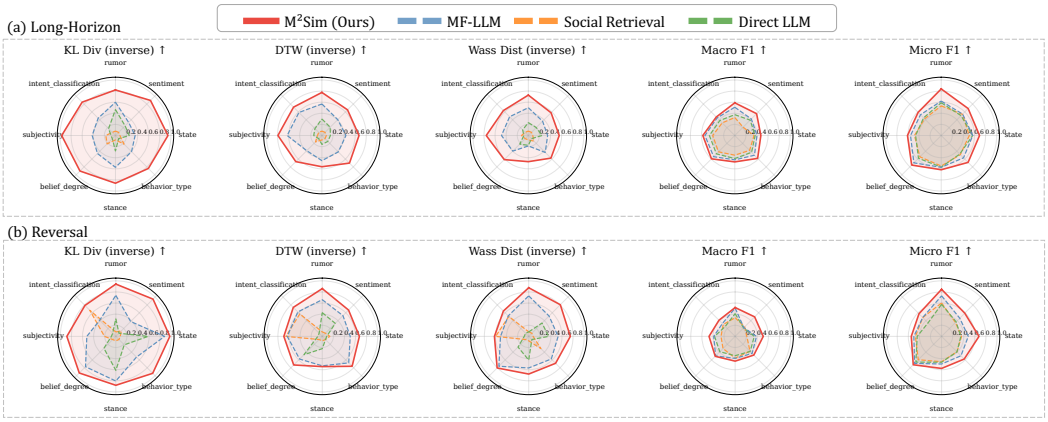

Figure 4: **Radar plot** comparing **M²Sim (Ours)** and three baselines (**MF-LLM**, **Social Retrieval**, **Direct LLM**)— on 5 distributional metrics and 8 semantic dimensions of actions. KL Divergence (Inverse), Wasserstein Distance (Inverse), and DTW Distance (Inverse) are inversely normalized so larger values indicate better performance, alongside Macro F1 and Micro F1. Larger areas denote superior performance, with **M²Sim** showing the highest semantic fidelity.

- **Sentiment.** The expressed emotional tone (including sarcasm/irony), e.g., angry, calm, happy, sad, fear, surprise.
- **State.** Overall polarity toward the topic: positive, negative, or neutral, including subtle negativity.
- **Behavior.** Interaction type: `share` (repost/forward) vs. `comment` (textual response).
- **Stance.** Position toward the topic: support, oppose, or neutral, including implicit opposition.
- **Belief.** Perceived truthfulness: believe vs. doubt (skepticism, requests for evidence, denial).
- **Subjectivity.** Subjective personal opinion vs. objective factual description.
- **Intent.** Communicative goal: question (seeking clarification), promotion (disseminating), or opinion (expressing viewpoint).

(ii) Macro-level (distributional dynamics). Building on the micro-level annotations, at each timestep t we map every action into a discrete label (under a chosen dimension) and aggregate the n_t actions into an empirical categorical distribution p_t; we compute the same distribution $\hat{p}_t$ for generated actions. We then compare the real trajectory $\{p_t\}_{t=1}^{T}$ and the generated trajectory $\{\hat{p}_t\}_{t=1}^{T}$ using:

1. **KL Divergence.** $\mathrm{KL}(p_t \,\|\, \hat{p}_t)$, averaged over timesteps, penalizing mismatched probability mass and being sensitive to mode dropping.
2. **Wasserstein Distance.** $W(p_t, \hat{p}_t)$, averaged over timesteps, measuring the cost of transporting probability mass and being more robust to small support shifts.
3. **Dynamic Time Warping (DTW).** DTW between the two time series $\{p_t\}$ and $\{\hat{p}_t\}$ (or scalar projections per label), evaluating temporal alignment by allowing elastic matching across timesteps and penalizing phase shifts.
4. **Negative Log-Likelihood (NLL).** The average $-\log \pi_\phi(a^\star_{i,t} \mid z_{i,t})$ over all ground-truth actions, measuring how well the learned policy assigns probability to real behaviors.
5. **Macro-F1.** F1 computed from predicted vs. real labels and averaged across classes (treating each class equally), highlighting performance on minority labels.
6. **Micro-F1.** F1 computed by aggregating true/false positives across all actions before forming F1, emphasizing overall accuracy dominated by frequent labels.

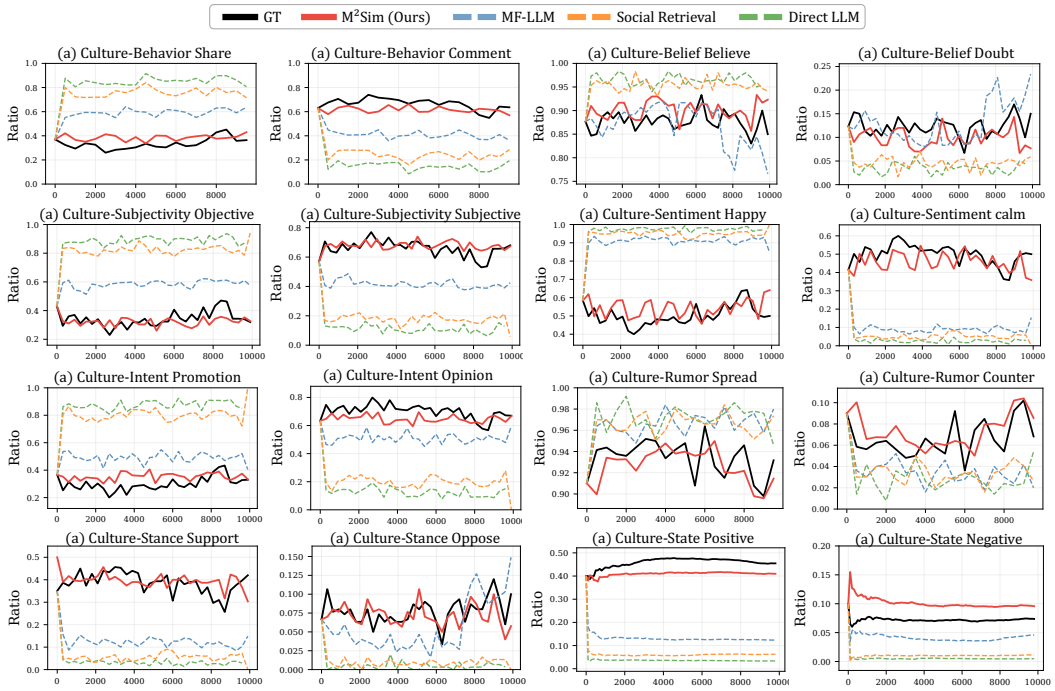

Figure 5: **Comparison on 8 semantic dimensions.** We compare **M²Sim (Ours)**, *MF-LLM*, *Social Retrieval*, and *Direct LLM*.

D Additional Experimental Results and Analysis

D.1 Full Results

Semantic fidelity under evolving states. Fig. 4 evaluates semantic fidelity over eight semantic-related dimensions (excluding **NLL Loss**); for **KL/Wasserstein/DTW** we use inverse normalization so higher is better. Across both *long-horizon* and *reversal* settings, M²Sim consistently achieves the largest radar area. Most notably, M²Sim leads by a wide margin on the **State** axis, showing that it best preserves state-consistent action meaning—enabled by explicitly modeling and coupling the macro sentiment distribution (positive/neutral/negative) with each agent's latent state during decision making. This state grounding propagates to closely related dimensions. Compared with all baselines, M²Sim shows clear gains on **Sentiment** and **Stance**, indicating that expressed attitudes track the underlying state rather than drifting with surface text patterns. We also observe a substantial improvement on **Behavior**, which is tightly linked to sentiment states (e.g., supportive vs. opposing engagement patterns under positive vs. negative shifts), consistent with M²Sim's better distributional alignment (**KL/Wass/DTW**) and higher **Macro F1/Micro F1**. Overall, the radar shapes suggest that jointly modeling macro and micro states is key to producing semantically faithful collective actions, especially when trajectories must adapt over time.

D.2 Full Analysis

Case Events. The case events used in Fig. 2 (Events a and b) and Fig. 7 (Events A-D) are summarized in Table 4.

Long-horizon semantic alignment in full dynamic simulations. Fig. 5 plots the 10,000-step trajectories of eight semantic dimensions for event (a) in the Great-Wall long-horizon simulation. M²Sim (red) is consistently the closest to GT (black) in both *level* and *trend*: across **Behavior** (Share/Comment), **Subjectivity** (Objective/Subjective), **Intent** (Promotion/Opinion), **Belief** (Believe/Doubt), and **Rumor** (Spread/Counter), it reproduces GT-like plateaus and fluctuations

Table 4: Representative events with opinion reversal across domains.

ID	Title	Domain	Description	Distinctive Features
a	Celebrity Coordinated Posting Controversy	Culture	A public figure released a critical social media post containing an unintended scheduling artifact, revealing coordinated narrative behavior on online platforms.	Observable evidence of organized opinion coordination; ineffective denial; persistent credibility erosion.
b	Global Mathematics Competition Eligibility Dispute	Education	An unexpected finalist from a non-traditional background initially triggered widespread admiration, later reversed by official findings of rule violations.	Rapid shift from emotional endorsement to scrutiny of procedural fairness and integrity.
A	Full Registration-Based IPO Reform	Economy	The implementation of a registration-based IPO system, alongside market-stabilization measures, initially generated optimism but later faced skepticism as outcomes diverged from expectations.	Transition from policy-driven enthusiasm to institutional performance reassessment.
B	Gradual Retirement Age Adjustment Policy	Society	A phased retirement age adjustment policy initially sparked strong resistance, later moderated by clarifications emphasizing flexibility and voluntariness.	Shift from collective anxiety to pragmatic individual adaptation.
C	Nuclear Wastewater Discharge and Public Response	Environment	A cross-border environmental discharge plan provoked intense public anxiety, followed by a gradual shift toward long-term scientific monitoring frameworks.	Opinion evolution from acute panic to evidence-based risk oversight.
D	University Library Harassment Allegation	Campus	An online harassment allegation prompted strong initial support for perceived victims, later complicated by additional evidence and procedural disclosures.	Rebalancing between moral advocacy and procedural objectivity.

rather than drifting toward saturated, nearly constant curves. *MF-LLM* is generally second-best but shows noticeable offsets and occasional instability, while *Social Retrieval* and *Direct LLM* frequently collapse into extreme, unbalanced patterns. The dominant gap appears on **State**: M^2Sim closely matches GT on both **State Positive** and **State Negative**, whereas all baselines underestimate them (curves near zero), effectively collapsing toward an overly neutral distribution. This state mismatch propagates to state-adjacent semantics—**Sentiment** (Happy/Calm) and **Stance** (Support/Oppose)—where baselines tend to produce one-sided polarity that does not follow GT's balance. With explicit conditioning on the macro state distribution and each agent's latent state, M^2Sim maintains coherent sentiment and stance dynamics, which further translates into more realistic **Behavior** trajectories over long horizons.

State reversals under exogenous signals in reversal events. Beyond the reversal case in Fig. 2(b), we further evaluate four additional reversal events with horizons ranging from 6,000 to 40,000 steps. These events feature long-run dynamics where the macro state distribution (positive/neutral/negative) can *reversal* under exogenous signals. As shown in Fig. 7, M^2Sim consistently captures both the turning points and the post-reversal trends, staying close to GT across events rather than converging to a stationary trajectory. This advantage comes from M^2Sim's explicit *state coupling*: it conditions decisions jointly on the macro state distribution and each agent's latent state, while injecting the exogenous signal as a direct driver of state evolution. When the exogenous signal shifts the macro distribution, M^2Sim propagates the change through agent states and back into the aggregate trajectory, producing the correct reversal dynamics. In contrast, *MF-LLM* relies on a text-based, coarse macro summary that becomes increasingly blurry over long horizons; after a short warm-up, it loses discriminative state information, so the predicted state curves remain nearly unchanged even when GT reverses. This is further reflected in the fine-grained components: M^2Sim aligns best with GT on **Positive**, **Negative**, and **Neutral** simultaneously, whereas baselines typically under-react and collapse toward an overly neutral or biased mixture.

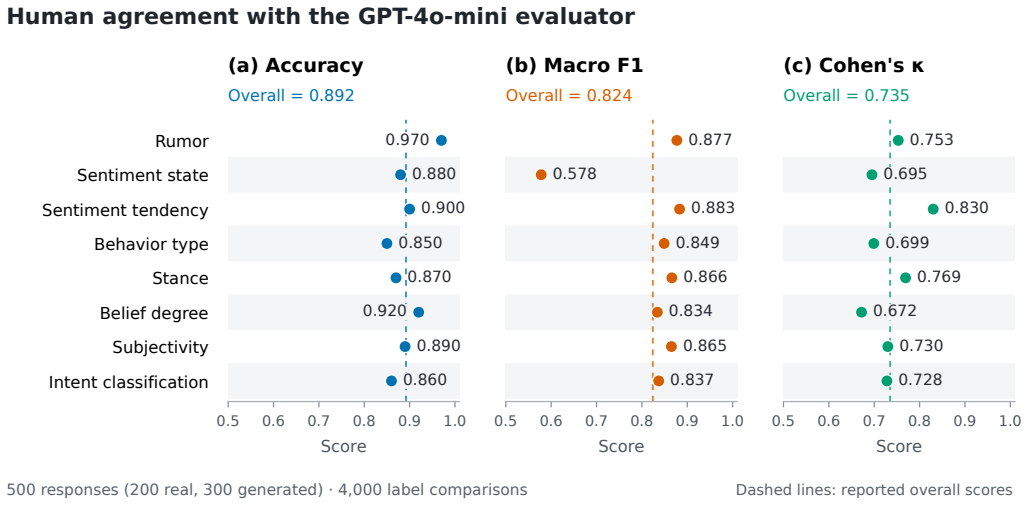

Figure 6: **Human validation of the GPT-4o-mini evaluator.** Human–evaluator agreement across eight semantic dimensions, measured by (a) Accuracy, (b) Macro F1, and (c) Cohen's κ. Dashed lines indicate the reported overall scores.

D.3 Human Validation of LLM-Based Evaluation

To assess the reliability of GPT-4o-mini as a semantic evaluator, we randomly sampled 500 responses, comprising 200 real and 300 generated responses. The samples were evenly assigned to five human annotators, each labeling 100 responses without access to the evaluator's labels. Each response was annotated along the same eight semantic dimensions used in our evaluation, yielding 4,000 human–evaluator comparisons.

Figure 6 shows overall Accuracy, Macro F1, and Cohen's κ of 0.892, 0.824, and 0.735, respectively, supporting the use of GPT-4o-mini for semantic evaluation. Agreement varies across dimensions, with sentiment state exhibiting a lower Macro F1 of 0.578, indicating that annotation uncertainty remains.

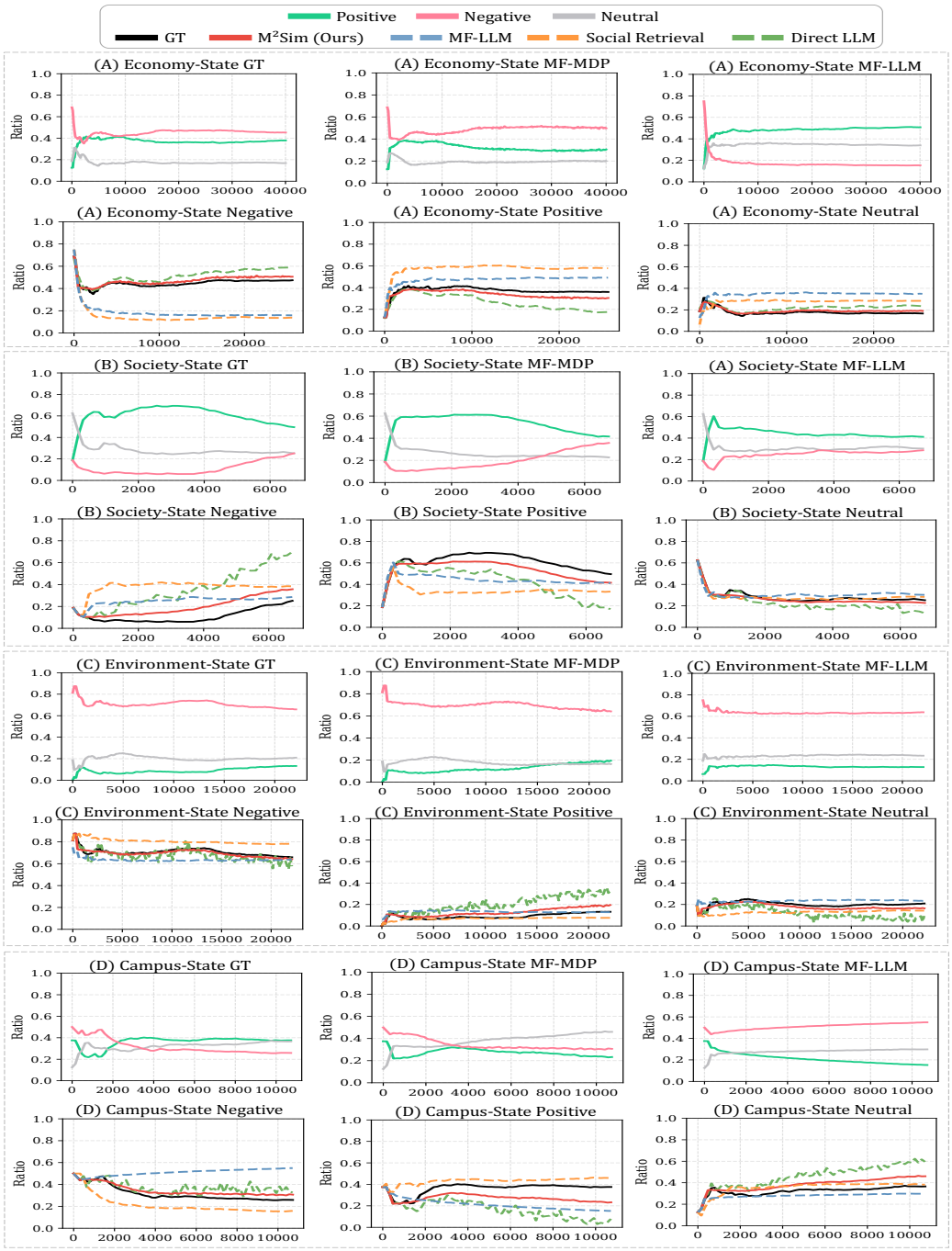

Figure 7: **State-distribution trajectories on four reversal events.** We compare **Real data** with **M²Sim (Ours)**, *MF-LLM*, *Social Retrieval*, and *Direct LLM*.